

Qualitätsbereinigung im Mietpreisindex: Schlussbericht

Zürcher Kantonalbank
Financial Engineering Immobilien



**Zürcher
Kantonalbank**

Impressum

Herausgeberin Zürcher Kantonalbank

Autoren Adrian Lüscher
Marco Salvi
Andreas Bröhl
Andrea Horehájová

Zürcher Kantonalbank

Titelbild © Zürcher Kantonalbank

© Copyright 2010 by Zürcher Kantonalbank

Inhaltsverzeichnis

1	MANAGEMENT SUMMARY	4
2	AUSGANGSLAGE	6
3	METHODISCHE GRUNDLAGEN UND VORGEHEN	7
3.1	Theoretische Grundlagen	7
3.2	Hedonic Repricing	8
3.3	Die Frage der Gewichtung	9
3.4	Umfang der Qualitätsbereinigung	10
3.5	Modellierungsstrategie	11
4	DATENGRUNDLAGEN UND DATENBEREINIGUNG.....	12
4.1	Datenbasis Mietpreis-Strukturerhebung 2003	12
4.2	Geodaten	18
4.3	Datenbereinigung	21
5	DATENIMPUTATIONEN UND –TRANSFORMATIONEN	30
5.1	Imputation fehlender Werte der Eigentumsform.....	30
5.2	Datentransformationen.....	31
6	MODELLEVALUATION UND ERGEBNISSE	35
6.1	Modellschätzung und Modellergebnisse	35
6.2	Ausreisserbehandlung und Residuenanalyse.....	39
6.3	Modellvergleiche	40
6.4	In- und Out-of-Sample Analyse.....	43
6.5	Modellanwendung für Qualitätsbereinigung im Mietpreisindex	44
6.6	Fazit und Empfehlungen.....	45
	LITERATUR.....	46
7	ANHANG	48
7.1	Verteilung der MSE-Beobachtungen pro Postleitzahlgebiet	48
7.2	Geodaten	49
7.3	Datenbereinigung MSE 2003 – Nettomiete.....	57
7.4	Erweiterte Modellergebnisse.....	58
7.5	Tabellenverzeichnis	61
7.6	Abbildungsverzeichnis.....	62
7.7	SAS-Daten und Programme	62

1 Management Summary

Auftrag und Zielsetzung

Die Abteilung Financial Engineering Immobilien der Zürcher Kantonalbank wurde vom Bundesamt für Statistik (Sektion Preise) mit der Entwicklung eines Systems für die Qualitätsbereinigung im Mietpreisindex (MPI) beauftragt. Ziel des Mandats ist es, ein Modell zu entwickeln, das Qualitätsunterschiede zwischen verschiedenen Wohnungen zuverlässig schätzt. Die Methodik soll auf die Rahmenbedingungen des MPI abgestimmt sein und sich für eine quartalsweise Produktion eignen. Mit Hilfe des Modells soll der MPI um die Preiskomponente bereinigt werden können, die aus der rotationsbedingten Qualitätsänderung der Stichprobe resultiert. Der vorliegende Schlussbericht dokumentiert das methodische Vorgehen, die Datengrundlagen sowie Ergebnisse und Empfehlungen.

MPI-Stichprobenrotation ändert Qualität

Der Mietpreisindex misst die Entwicklung der Nettomieten auf Niveau Schweiz im Zeitverlauf. Die Stichprobe umfasste bisher rund 5'000 Mietwohnungen mit quartalsweiser Rotation vom jeweils ältesten Achtel, um die Repräsentativität bezüglich der Grundgesamtheit zu wahren. Mit dieser Rotation sind jedoch auch Qualitätsveränderungen der Wohnungen in der Stichprobe verbunden. Die Stichprobe ist zwar geschichtet nach Kategorien des Gebäudealters und der Zimmerzahl, wobei die Stichprobenanteile dieser Zellen über die Zeit stabil bleiben. Für weitere Qualitätsunterschiede zwischen den quartalsweise in einer Zelle neu gezogenen Wohnungen und solchen, die aus der Stichprobe hinausrotiert werden, wurde bislang jedoch nicht kontrolliert. Will der MPI die "reine" Preisänderung der Nettomieten messen, so muss er um diese Preiskomponente bereinigt werden.

Vergleich von Modellpreisen kontrolliert für Qualitätsänderung

Zu diesem Zweck wird ein hedonisches Mietpreismodell entwickelt, das eine Quantifizierung der Mietpreisbeiträge einzelner Merkmale der Wohnung, der kleinräumigen sowie der grossräumigen Lage erlaubt. Mit dem Modell kann für jede Wohnung ein theoretischer Mietpreis berechnet werden, basierend auf vorgängig definierten und objektiv messbaren Qualitätsmerkmalen. Aus dem direkten Vergleich dieser geschätzten Mietpreise für die in einer Zelle neu gezogenen bzw. die aus der Stichprobe ausscheidenden Wohnungen, lässt sich die Preiskomponente berechnen, die auf die unterschiedliche Stichprobenzusammensetzung zurückzuführen ist.

Methodisches Vorgehen und Datengrundlagen

Unser Vorgehen orientiert sich am Ansatz des so genannten "hedonic repricing". Theoretische und methodische Grundlagen dazu sind im Detailkonzept zur Revision 2010 des Landesindex der Konsumentenpreise (LIK) beschrieben und an anderer Stelle in diesem Bericht kommentiert. Datengrundlage bildet die Mietpreis-Strukturerhebung 2003 (MSE), die mit Hilfe eines leistungsstarken geografischen Informationssystems (GIS) mit einer Vielzahl lagespezifi-

scher Variablen angereichert wurde. Damit konnten auch sehr kleinräumige Lageeigenschaften (z.B. Immissionen) abgebildet und deren Einflüsse auf den Mietpreis quantifiziert werden. Die vom spezifischen Mietobjekt unabhängigen preisrelevanten Lageeigenschaften wurden gesamtschweizerisch für jeden Siedlungshektar berechnet. Das entwickelte Mietpreismodell lässt sich quartalsweise auf die MPI-Stichprobe anwenden und ermöglicht eine Qualitätsbereinigung im Sinne des "hedonic repricing".

Ergebnisse des hedonischen Mietpreismodells

Das finale Modell weist eine gute Erklärungskraft auf, für zwei Drittel der Beobachtungen beträgt die Differenz zwischen geschätzter und effektiver Nettomiete maximal 22 Prozent. Die finale Modellspezifikation zur Schätzung der Nettomieten umfasst insgesamt 58 Variablen. Dabei unterscheiden wir Objekteigenschaften (Gebäudealter, Zimmerzahl, Wohnfläche usw.), Mikrolageeigenschaften (Distanz zum See, Strassenlärm, usw.) sowie Makrolageeigenschaften (Steuerbelastung, Erreichbarkeitspotenzial von Arbeitsplätzen, usw.). Gemessen an ihrer relativen Einflussstärke (standardisierte Koeffizienten) auf die geschätzte Nettomiete kommt den Objekteigenschaften "Gebäudealter" sowie "Zimmerfläche" die grösste Bedeutung zu, d.h. diesbezügliche Unterschiede der ausgetauschten Wohnungen führen zu den grössten Qualitätskorrekturen im Mietpreisindex. Danach folgen die Merkmale der Makrolage – Erreichbarkeitspotenzial und Steuern – und schliesslich die Merkmale der Mikrolage.

Fazit und Empfehlungen

Die gewählte, auch von Eurostat empfohlene Methode des "hedonic repricing" ist gut geeignet, um eine Qualitätsbereinigung der MPI-Berechnung durchzuführen. Die Umsetzung ist auf die Rahmenbedingungen des MPI abgestimmt und lässt sich in die quartalsweise Produktion integrieren. Die statistische Qualität des hedonischen Mietpreismodells kann als gut bezeichnet werden. Für die Modellanwendung auf die MPI-Daten wären vertiefte Analysen und grössere Fallzahlen in den einzelnen Zellen – besonders bei den 1- bzw. 6-Zimmerwohnungen jüngeren Baujahrs – notwendig. Die im Rahmen der MPI-Revision 2010 erfolgten zusätzlichen Massnahmen im Zusammenhang mit der Stichprobenaufstockung sind deshalb zu begrüßen.

Eine Modellkalibrierung bzw. Neuschätzung drängt sich kurzfristig nicht auf, zumal dafür aktuellere Mietstrukturdaten als die MSE 2003 notwendig wären. Alternativ ist eine Modellierung auf Basis gepoolter MPI-Daten zu überlegen. Die topografischen Lagefaktoren sind invariant über die Zeit. Aktualisierungen wären bei den Steuerdaten und den Erreichbarkeitspotenzialen (Betriebszählungsdaten) möglich, sofern die entsprechenden Daten vorliegen.

2 Ausgangslage

Revision 2010 des Landesindexes der Konsumentenpreise

Im Rahmen der Revision 2010 des Landesindexes der Konsumentenpreise (LIK) wird neben anderen Warenkorbbereichen auch der Bereich der Mieten revidiert. Nebst anderen Revisionspunkten kommt der Behandlung von Qualitätsunterschieden beim Ersatz von Wohnungen in der Stichprobe des Mietpreisindex (MPI) eine zentrale Bedeutung zu. Die MPI-Stichprobe umfasst bisher rund 5000 Mietwohnungen, deren Mietpreisentwicklung jeweils über eine Periode von acht Quartalen beobachtet wird. Danach werden diese Wohnungen im Rahmen des Stichprobenplans durch neu gezogene Wohnungen ersetzt. Jedes Quartal erfolgt der Austausch des jeweiligen ältesten Achtels der Stichprobe. Die mit diesem Wohnungsaustausch verbundenen Qualitätsunterschiede in der Stichprobe sollen mit einem geeigneten System bereinigt werden. Die Sektion Preise des Bundesamt für Statistik hat zu diesem Zweck im November 2009 ein entsprechendes Forschungsmandat ausgeschrieben.

Auftrag und Zielsetzung

Die Abteilung Financial Engineering Immobilien der Zürcher Kantonalbank wurde im Rahmen des ordentlichen Ausschreibungsverfahrens im Januar 2010 mit diesem Mandat beauftragt. Ziel des Mandats ist die Entwicklung eines Modells, das die Qualitätsunterschiede zwischen den verschiedenen Wohnungen zuverlässig schätzt. Das Modell soll eine quartalsweise Produktion ermöglichen und auf die methodischen Rahmenbedingungen des Mietpreisindex abgestimmt sein, wie sie im Detailkonzept¹ zur Revision des LIK beschrieben sind.

Aufbau der Studie

Der Aufbau des vorliegenden Berichts gliedert sich folgendermassen: Das folgende Kapitel 3 gibt eine Übersicht über die methodischen Grundlagen und das gewählte Vorgehen. In Kapitel 4 werden die für das Schätzmodell verwendeten Daten sowie die Prozesse der Geokodierung und Datenbereinigung besprochen. Vorgenommene Datenimputationen (Ersatz fehlender Werte) und Datentransformationen folgen in Kapitel 5. Die Modellergebnisse sowie Vergleiche mit weiteren Modellen werden in Kapitel 6 vorgestellt. Im Kapitel 7 (Anhang) finden sich eine umfangreiche kartografische Dokumentation der verwendeten Lagevariablen sowie weitere detaillierte Modellinformationen.

¹ "Landesindex der Konsumentenpreise: Revision 2010"; BFS 2009, Version 4.0 D

3 Methodische Grundlagen und Vorgehen

3.1 Theoretische Grundlagen

Die hedonische Bewertungsmethode geht davon aus, dass sich der Preis eines Gutes auf dessen Eigenschaften zurückführen lässt. Dabei wird mit geeigneten statistischen Methoden der Preis (Mietpreis) in einzelne Preiskomponenten für die Merkmale des Objekts (Wohnung) bzw. der Lage zerlegt.

Die Anwendung von hedonischen Methoden kann auf eine lange Tradition zurückblicken: Bereits in den 20er Jahren findet sich eine Anwendung auf dem Gebiet der Landwirtschaft (Waugh, 1928). Mittlerweile wurde die Methode für die Preisbildung von verschiedensten Gütern (Autos, Liegenschaften, Computer, Mikrowellenöfen) verwendet. Wichtige Anwendungsgebiete sind die Bildung von qualitätsbereinigten Preisindizes und die Auswirkung von Veränderungen in den Qualitätsmerkmalen auf die Preise. In den letzten Jahren hat sich die Methode für die standardisierte Bewertung von Liegenschaften in der Praxis etabliert.

Paradoxerweise geht die Anwendung der Methode ihrer theoretischen Fundierung voraus. Die Beziehung zwischen dem Preis einer Wohnung und ihrer Merkmale wurde von Rosen (1974) im Rahmen eines mikroökonomischen Modells erstmals theoretisch hergeleitet. Rosen zeigte, dass die Preise der Wohnungsmerkmale (auch implizite bzw. hedonische Preise genannt) auf die Nachfrage und das Angebot nach diesen Merkmalen zurückzuführen sind, auch wenn die Merkmale nicht einzeln sondern nur als Bündel (in Form einer Wohnung) erworben werden können.² Aus Rosens Analyse geht zudem hervor, dass die mikroökonomische Theorie der funktionellen Form der hedonischen Gleichung kaum Restriktionen auferlegt. Jede funktionelle Form, welche die Daten fittet, ist grundsätzlich mit dieser Theorie konsistent. Formal kann die hedonische Preisfunktion für eine Mietwohnung i ausgedrückt werden als Funktion:

$$x_i = f(WG_i, ST_i, UW_i)$$

Dabei bezeichnet x die (MPI: Netto-)Wohnungsmiete (auch in log. Form), WG die Wohnung- und Gebäudeeigenschaften, ST die Merkmale des Standortes (Makrolage) und UW die Merkmale der Mikrolage, allen voran Umweltfaktoren wie Ruhe, Exposition, Aussicht usw. Die Frage der Wahl der geeigneten Variablen wird eingehend in den folgenden Kapiteln behandelt.

² Für eine Übersicht über die hedonische Methode vgl. Palmquist (2004). Alleine auf dem Gebiet der Umweltökonomie umfasst eine nicht abschliessende Bibliographie über 200 Referenzen (<http://www.sscnet.ucla.edu/ssc/labs/cameron/nrs98/hedoninv.htm>).

3.2 Hedonic Repricing

Das Hauptziel jedes Preisindexes ist die Messung der Preisveränderung bei konstanter Qualität. In der amtlichen Statistik wird der hedonische Ansatz vermehrt als Methode für die Qualitätsbereinigung eingesetzt. Verschiedene Abwandlungen der hedonischen Qualitätsbereinigung sind denkbar [vgl. Triplett, 2004]. Die im Detailkonzept (S. 56ff.) vorgeschlagene Methode ist eine Variante des "Hedonic Quality Adjustment Approach", auch "Hedonic Re-pricing" genannt [De Haan, 2007]. Sie ist mit der Art und Weise verwandt, wie die amtliche Statistik traditionell mit der Frage der Qualitätsbereinigung umgegangen ist.

Hedonic Re-pricing funktioniert vereinfachend wie folgt: Angenommen die Wohnung i werde von der Stichprobe eliminiert und mit einer Wohnung j ersetzt.³ Die entsprechenden Mieten bezeichnen wir mit x_i^0 und x_j^1 , wobei die Zahl im Exponent die Basisperiode (0) bzw. die laufende Periode (1) darstellt. Um die Qualitätsunterschiede zwischen diesen Objekten zu bereinigen, wird x_j^1 mit dem Verhältnis der geschätzten Mieten $\hat{x}_i^0 / \hat{x}_j^0$ multipliziert. Die geschätzten Mieten beruhen auf einer hedonischen Regression, welche mit Daten aus der Basisperiode bestimmt wurde.⁴

Das Detailkonzept (S. 59ff.) sieht vor, dass die Qualitätsbereinigung der neu in die Stichprobe des MPI aufgenommenen Mieten mit dem hedonischen Modell quartalsweise durchgeführt wird. Es wird also deutlich, dass die beschriebene Methode hohe Anforderungen an die hedonische Regression voraussetzt.

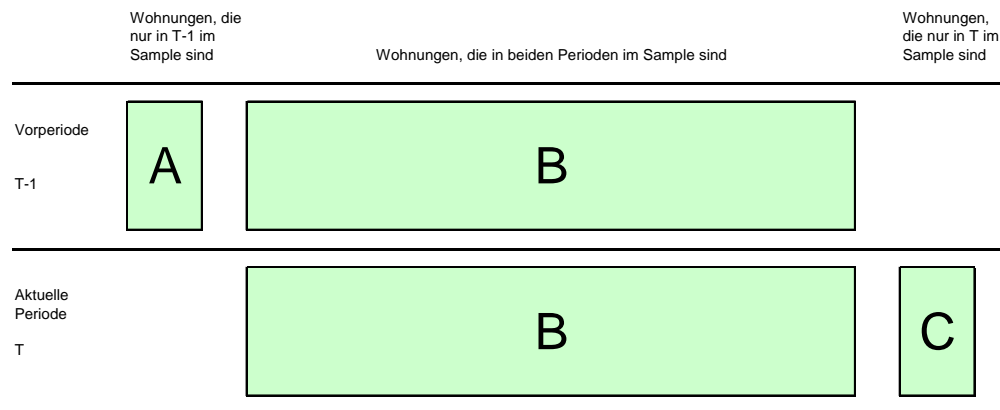
3.2.1 Anwendung des Hedonic Repricing auf die MPI-Berechnung

Die Anwendung der Qualitätsbereinigung nach der Hedonic Repricing Methode erfolgt folgendermassen. Jede Zelle der MPI-Stichprobe wird zum Zeitpunkt T (aktuelles Quartal) in zwei Bereiche aufgeteilt (vgl. Abbildung 1). Der Teil B umfasst diejenigen Beobachtungen, die bereits im Vorquartal in der Stichprobe waren (Folgebefragung). Der Teil C beinhaltet die neu gezogenen Wohnungen (Erstbefragung).

³ Wir abstrahieren hier von der Tatsache, dass im Rahmen des MPI die Stichprobe ex post in Zellen geschichtet wird. Für die Details der Stichprobenziehung verweisen wir auf das Detailkonzept (S. 35-47).

⁴ Es ist in Wirklichkeit nicht notwendig, jede aus der Stichprobe eliminierte Wohnung mit ihrer NaFr.olgerin direkt zu verlinken. Entscheidend für die Qualitätsbereinigung ist die *durchschnittliche* Differenz der Attribute der neuen und der eliminierten Wohnungen (siehe auch Detailkonzept, S. 61).

Abbildung 1: Unterteilung einer MPI-Stichprobenzelle für die Qualitätsbereinigung



Zum Zeitpunkt T-1 (Vorquartal) erfolgt die Unterteilung in Wohnungen, die das letzte Mal teilnehmen (Block A) und solche, die noch mindestens ein Quartal in der Stichprobe bleiben (Block B).

Die Indexberechnung für die Wohnungen im Teil B erfolgt wie bisher. Da es sich in beiden Quartalen um die gleichen Wohnungen handelt, kann auf eine Qualitätsbereinigung verzichtet werden, die gemessene Preisänderung entspricht der "echten" Mietpreisteuerung.

Eine Qualitätsbereinigung ist dagegen bei den ausgetauschten Wohnungen (Block A vs. C) notwendig. Für diese Wohnungen wird mit dem hedonischen Modell der geschätzte Mietpreis berechnet. Anschliessend werden die geschätzten Mieten im Teil A (\hat{x}_A^{t-1}) bzw. Teil C (\hat{x}_C^t) gemittelt. Das Verhältnis (\hat{g}^t) dieser gemittelten Modellmieten gibt an, wie sich die Mieten ohne "echte" Preisänderung entwickelt hätten, d.h. es misst die Preiskomponente, die sich allein aus der Qualitätsänderung ergibt.

$$\hat{g}^t = \frac{\hat{x}_C^t}{\hat{x}_A^{t-1}} = \frac{\left(\prod_{i=1}^{n_C^t} (\hat{x}_{iC}^t)^{p_i^t} \right)^{1/\sum_i p_i^t}}{\left(\prod_{i=1}^{n_A^{t-1}} (\hat{x}_{iA}^{t-1})^{p_i^{t-1}} \right)^{1/\sum_i p_i^{t-1}}}$$

3.3 Die Frage der Gewichtung

Die wichtigsten Herausforderungen in der Bildung eines hedonischen Regressionsmodells der Mietpreise sind primär empirischer Art. Eine Ausnahme dazu bildet die Frage der Gewichtung der Regressionsstichprobe, welche oft zu Missverständnissen führt. Sollen die Beobachtungen in einer hedonischen Regression besonders gewichtet werden? Sollen beispielsweise die Gewichte, die bei der Berechnung des MPI-Indexes verwendet werden, auch bei der Schätzung des hedonischen Modells berücksichtigt werden?

An dieser Stelle müssen wir klar zwischen *Gewichtung des Mietpreisindexes* und *Gewichtung in der hedonischen Regression* unterscheiden. Eine Gewichtung im MPI ist offensichtlich sinnvoll, weil die Repräsentativität des Indexes direkt von der Stichprobenziehung abhängt. Eine Gewichtung der hedonischen Regression bzw. ihrer Stichprobe ist jedoch a priori nicht zwingend. Das Ziel des hedonischen Modells liegt in der Lieferung einer möglichst präzisen Schätzung bzw. einer möglichst genauen Qualitätsbereinigung. Die Präzision der Schätzung lässt sich – unter anderem – auf die Variabilität der Stichprobe zurückführen. Die Varianz des gewöhnlichen Kleinste-Quadrate-Schätzers ist direkt proportional zur Inverse der Kovarianzmatrix der Regressoren. Die Schätzung ist umso präziser, je grösser diese Kovarianz bzw. je "heterogener" die Stichprobe ist. Eine Gewichtung der Stichprobe (nach Ausgaben oder Anzahl) würde hier die Effizienz des Schätzers reduzieren, weil sie die häufig vorhandenen Objekte stärker als die selteneren gewichten würde.

3.4 Umfang der Qualitätsbereinigung

Eine weitere Methodenfrage stellt sich bei der Bestimmung der Merkmale, die in der Qualitätsbereinigung berücksichtigt werden soll. Die hedonische Methode führt den Mietpreis einer Wohnung auf ihre (Objekt- und Lage-) Eigenschaften zurück. Inwiefern sollen aber die Eigenschaften des *Mietvertrages*, wie Eigentübertyp oder Vertragsdauer, bei der Modellierung einbezogen werden?

Das Beispiel der Vertragsdauer hilft uns diese Frage zu schärfen. Auf dem Schweizer Mietmarkt beeinflussen zahlreiche Vorschriften zum Mietrecht die Preisbildung. So sind bei laufenden Mietverträgen und bei Mieterwechseln der freien Mietpreisanpassung Grenzen gesetzt. Es ist zu erwarten, dass die einschränkende Wirkung der Vorschriften umso stärker ist, je stabiler die Mieterschaft und je urbaner die Lage. Verschiedene hedonische Studien haben in der Tat einen ökonomischen und statistischen Zusammenhang zwischen der Höhe der Miete und der Verweildauer nachgewiesen (Baranzini et al., 2005; Statistik Stadt Zürich, 2009).

Es wäre demnach verfehlt, die Vertragsdauer von der hedonischen Regression auszuschliessen. Es ist insbesondere zu erwarten, dass die Vertragsdauer mit weiteren Eigenschaften des Hauses, beispielsweise dem Gebäudealter, korreliert – die Vertragsdauer in Neubauten ist zwangsläufig tief. Mit dem Ausschluss dieses Merkmals würde sich ein "omitted variable bias" in die Regression einschleichen; die Schätzungen der impliziten hedonischen Preise wären verzerrt.

Aus dieser Erkenntnis folgt allerdings nicht, dass die neuen Beobachtungen der MPI-Stichprobe um unterschiedliche Vertragsdauern bereinigt werden sollen. Ziel des MPI ist die Berechnung der Mietpreisentwicklung bei konstanter Qualität *der Wohnungen*. Die Vertragsdauer beeinflusst zwar die Miete, sie stellt jedoch keine Eigenschaft der Wohnung dar. Würde die Mietregulierung "über Nacht" abgeschafft, ist zu erwarten, dass die Mieten im Durchschnitt steigen würden. Diese Erhöhung liesse sich jedoch nicht auf die Veränderung

der Eigenschaften der Wohnungen zurückführen. Es wäre also sinnvoll, dass sie als allgemeine Preiserhöhung vom MPI registriert werden würde.

Zusammenfassend stellen wir fest, dass die Schätzung eines hedonischen Modells von der hedonischen Qualitätsbereinigung der MPI-Stichprobe methodologisch klar zu trennen ist. Bei der Schätzung des Modells – wo Fragen der Effizienz und Konsistenz der Schätzung im Vordergrund stehen – sind sämtliche mietpreisbestimmenden Merkmale zu berücksichtigen. Bei der Bereinigung der Qualität sollen hingegen nur jene Merkmale berücksichtigt werden, welche die Qualität der Wohnung (inkl. Lagemerkmale) direkt beeinflussen.

3.5 Modellierungsstrategie

Das hedonische Regressionsmodell wurde auf der Grundlage der MSE-2003 Daten spezifiziert. Mehrere Modellvarianten mit unterschiedlichen Zielgrössen (Bruttomieten, Nettomieten mit/ohne imputierte Werte) wurden getestet. Dabei haben wir eine Spezifikation gewählt, die ohne wesentliche Änderungen auf die MPI-Stichprobe übertragen werden kann. Um ein möglichst effizientes Modell zu generieren, wurde auf eine Gewichtung der Stichprobendaten⁵ verzichtet. Aus dem gleichen Grund haben wir die Dauer des Mietvertrages als Proxygrösse (welche keinen direkten Nutzen stiftet, jedoch das Modell verbessert) in die Schätzgleichung einbezogen. In der Modellanwendung auf die MPI-Stichprobe wird der Effekt der Vertragsdauer jedoch nicht berücksichtigt, da dieser keine Qualitätseigenschaft der Wohnung darstellt.

Ein weiteres Modell wurde mit den MPI-Daten geschätzt. Mit geeigneten statistischen Methoden haben wir überprüft, ob sich die Eigenschaftspreise signifikant unterscheiden und ob sich die Annahme konstanter Eigenschaftspreise rechtfertigen lässt.

⁵ Gewichtete Varianten wurden überprüft, die Koeffizienten und die Modellgüte haben sich dabei nur unwesentlich verändert.

4 Datengrundlagen und Datenbereinigung

In diesem Abschnitt beschreiben wir die der Modellierung zu Grunde liegende Datenbasis sowie die Prozesse und statistischen Verfahren der Datenbereinigung. Mit der Mietpreis-Strukturerhebung 2003 steht uns eine insgesamt hervorragende Datenbasis zur Verfügung, die eine umfassende, repräsentative Abdeckung des schweizerischen Mietwohnungsbestandes darstellt.

4.1 Datenbasis Mietpreis-Strukturerhebung 2003

Mit Stichtag 1. November 2003 führte die Sektion PREIS des Bundesamts für Statistik (BFS) zum zweiten Mal nach 1996 eine gross angelegte Mietpreis-Strukturerhebung (MSE03) durch. Die MSE stellt die umfassendste und aktuellste Erhebung der schweizerischen Mietlandschaft dar. Wir verwenden diese Datenbasis als grundlegenden Baustein für die im Rahmen des vorliegenden Mandats durchzuführende Bereinigung des Mietpreisindex. In einem Methodenbericht des BfS (Potterat 2005) wurden die Analyse des Rücklaufs, die verwendeten Einsetzungsverfahren, die Gewichtung, sowie die diversen Schätzverfahren beschrieben. Wir lehnen uns im Folgenden weitgehend an dieses Vorgehen an, abweichende Regelungen werden kommentiert.

4.1.1 Geokodierung der Mietpreis-Strukturerhebung

Insgesamt umfasst die MSE03 234'153 Beobachtungen auf. Davon weisen 48'378 Beobachtungen im Originaldatensatz keine Geokoordinaten auf, d.h. sie würden für die Modellierung verloren gehen, da sie nicht mit den Lagevariablen verknüpft werden können. Dies entspricht 20.7%. Verwendbar wären somit lediglich 185'775 Beobachtungen (79.3%). Daher wurde basierend auf dem Adressdatensatz der MSE eine erneute Geokodierung mittels des ZKB Algorithmus durchgeführt.

Dabei werden zwei Zielsetzungen verfolgt:

1. Möglichst viele weitere (bisher nicht-geokodierte) Beobachtungen für die Modellierung zu geokodieren.
2. Bisher „ungenau“ geokodierte Beobachtungen in dem Datensatz korrigieren.

Anschliessend werden die Ergebnisse der ZKB Geokodierung mit denjenigen des Originaldatensatzes abgeglichen. Für den Abgleich mit den Original-Koordinaten der MSE werden aus dem ZKB Algorithmus nur die Zuordnungen der Kategorie „exakter Match“ inklusive der Kategorie „guter Match“ verwendet (Stufen 1 bis 7). Das sind die Beobachtungen die adressgenau verortet werden können. Eine Übersicht über die möglichen Stufen der Treffergenauigkeit ist in Tabelle 1 ersichtlich.

Tabelle 1: Übersicht über die Stufen der Treffergenauigkeit des ZKB Algorithmus

Stufe	Beschreibung	Technische Beschreibung
1	exakter Match	PLZ/STR/NR/ORT
2	exakter Match	PLZ/STR/NR/ORT (mit String-Abstand bei Strasse)
3	exakter Match mit Korrektur	PLZ/STR/NR/ORT (stand.)
4	exakter Match mit Korrektur	PLZ/STR/NR/ORT (stand., mit String-Abstand)
5	guter Match	PLZ/STR/NR (stand., mit String-Abstand)
6	guter Match	STR/NR/ORT (stand., mit String-Abstand)
7	guter Match	PLZ/STR/NR/ORT ohne NR-Zusatz (stand., mit String-Abstand)
8	strassengenauer Match	PLZ/STR/NR ohne NR-Zusatz (stand., mit String-Abstand)
9	strassengenauer Match	STR/ORT ohne NR-Zusatz (stand., mit String-Abstand)
10	strassengenauer Match	PLZ/STR/ORT (stand.)
11	Ortgenauer Match	PLZ/ ORT
12	Ortgenauer Match	PLZ
99	kein Match: PLZ im Ausland oder leer	PLZ im Ausland oder leer
999	Ausland	Ausland

4.1.1.1 Ergebnisse der erneuten ZKB Geokodierung

Mittels des ZKB-Algorithmus konnten 83.44% aller Beobachtungen adressgenau geokodiert werden. Die Tabelle 2 zeigt die detaillierte Ergebnisse. Diejenigen Adressen, die lediglich auf Stufe Strasse oder Ort übereinstimmen, werden für den Abgleich mit den originalen Geokoordinaten der MSE nicht berücksichtigt.

Tabelle 2: Ergebnis des ZKB Algorithmus

Geokodierungsstufe	Anzahl	Anteil (%)	Kumulative Anzahl	Kumulativer Anteil (%)
exakter Match	189'338	80.86	189'218	80.86
exakter Match mit Korrektur	2'807	1.20	192'025	82.06
guter Match	3'236	1.38	195'261	83.44
strassengenauer Match	12'080	5.16	207'341	88.60
Ortgenauer Match	26'311	11.24	233'652	99.84
kein Match: PLZ im Ausland	380	0.16	234'152	100
Ausland	1	0	234'153	100

4.1.1.2 Abgleich von ZKB und BfS Geokodierung

Insgesamt kommt es bei 11'393 Beobachtungen zu den exakt identischen Koordinaten (metergenau). Bei der grossen Mehrheit (165'380) stimmen die zugeordneten Koordinaten nicht exakt überein. 9'002 Beobachtungen weisen nur die Originalkoordinaten aus dem GWR auf. 18'608 Beobachtungen konnten neu über den ZKB-Algorithmus geokodiert werden. 29'770 Beobachtungen können auch nach der zusätzlichen Geokodierung nicht verortet werden. Diese gehen endgültig für die Modellierung verloren (Tabelle 3).

Tabelle 3: Vergleich der Geokodierungen

Geokodierungsstufe	Anzahl	Anteil (%)	Kumulative Anzahl	Kumulativer Anteil (%)
identische Treffer	11'393	4.87	11'393	4.87
unterschiedliche Treffer	165'380	70.63	176'773	75.49
alter Treffer	9'002	3.84	185'775	79.34
neuer Treffer	18'608	7.9	204'383	87.23
kein Treffer	29'770	12.77	234'153	100

4.1.1.3 Validierung der Ergebnisse

Die hohe Anzahl an unterschiedlichen Treffern macht weitere Auswertungen und gegebenenfalls Korrekturen erforderlich, auf die im Folgenden eingegangen wird.

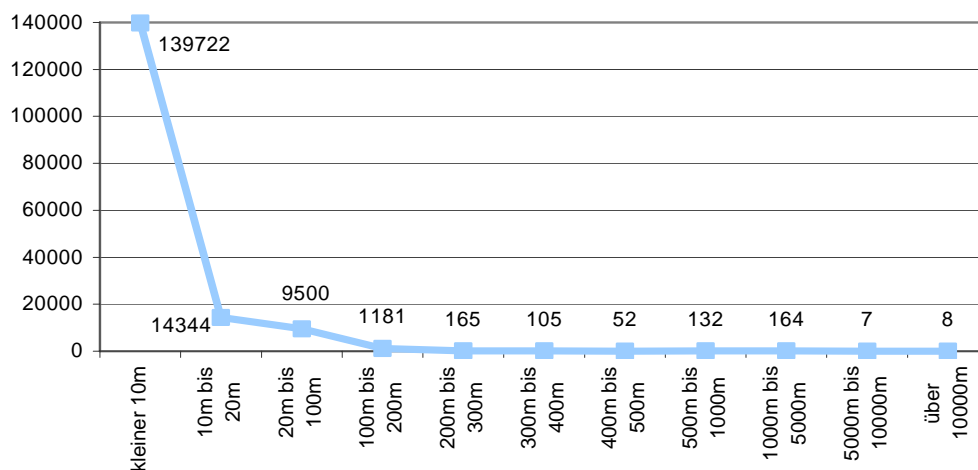
Grundsätzliche Problematik der unterschiedlichen Treffer

Die Geokodierung via „Geopost“ setzt die Koordinaten auf den Hauseingang. Die Koordinaten der originalen MSE-Geokodierung liegen in der Mitte der Gebäude. Zur genaueren Analyse der Unterschiede werden die Distanzen zwischen den Koordinaten der MSE- und der ZKB-Geokodierung berechnet und anschliessend klassifiziert. Die Tabelle 4 und Abbildung 2 zeigen die Verteilung der Beobachtungen über 11 Distanzklassen, wobei rund 60% der Beobachtungen in die erste Klasse (0 bis 9m) fallen.

Tabelle 4: Verteilung der Beobachtungen über die Entfernungsklassen (N=234'153)

Entfernungsklasse	Anzahl	Anteil (%)	Kumulative Anzahl	Kumulativer Anteil (%)
kleiner 10 m	139'722	59.67	139'722	59.67
10 bis 20 m	14'344	6.13	154'066	65.80
20 bis 100 m	9'500	4.06	163'566	69.85
100 bis 200 m	1'181	0.50	164'747	70.36
200 bis 300 m	165	0.07	164'912	70.43
300 bis 400 m	105	0.04	165'017	70.47
400 bis 500 m	52	0.02	165'069	70.50
500 bis 1'000 m	132	0.06	165'201	70.55
1'000 bis 5'000 m	164	0.07	165'365	70.62
5'000 bis 10'000 m	7	0.00	165'372	70.63
über 10'000 m	8	0.00	165'380	70.63
identische Treffer	11'393	4.87	176'773	75.49
alter oder neuer Treffer	27'610	11.74	204'383	87.23
kein Treffer	29'770	12.77	234'153	100

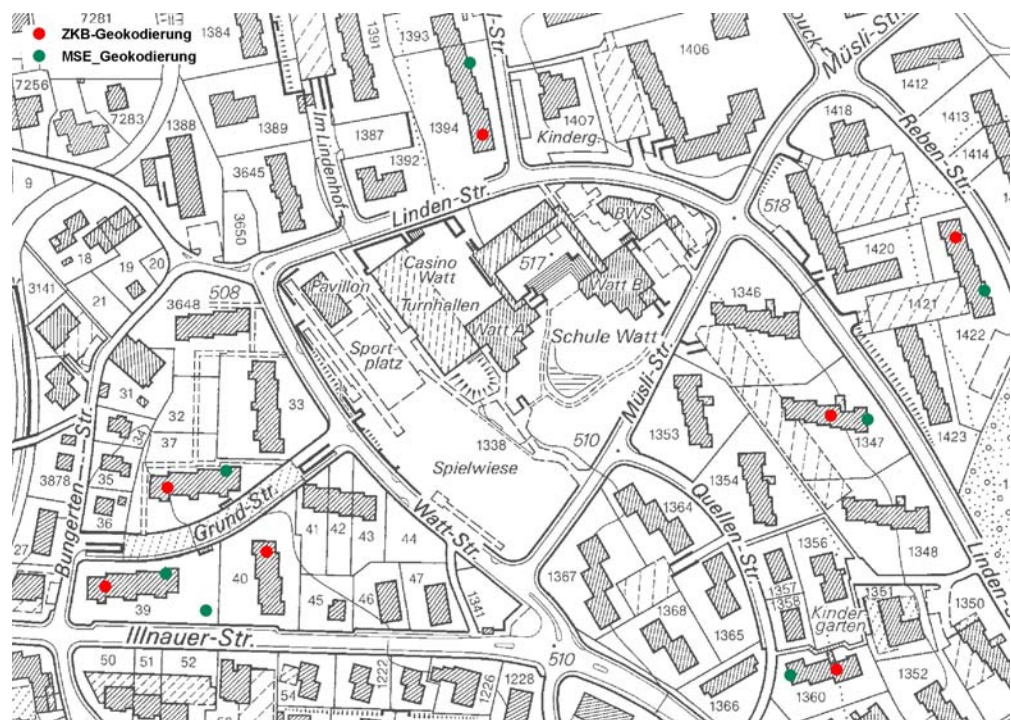
Abbildung 2: Verteilung der Beobachtungen über die Entfernungsklassen (N=165'380)



Im Folgenden werden die Abweichungen in den Entfernungsklassen genauer betrachtet, um die Relevanz der Unterschiede in der Geokodierung aufzuzeigen. Die Unterschiede in den ersten beiden Klassen rühren allein von dem Unterschied in den Geokodierungen von GWR und Geopost her. Die Geokoordinate des GWR liegt in der Gebäudemitte und die Geokoordinate von Geopost liegt am Gebäudeeingang. Durch die geringen Abweichungen sind die Auswirkungen auf die Zuordnung der Variablen der Mirkolage gering.

Beobachtungen mit Entfernungen in der Klasse "20m bis 100m" treten hauptsächlich in städtischen und in Agglomerationsgemeinden auf. Wie am folgenden Detailausschnitt (Abbildung) zu erkennen ist, hängt dies von der Grösse der Häuser und den dementsprechend grösseren Abständen zwischen Gebäudeeingang (MSE, GWR) und Gebäudemittelpunkt (Geopost) ab.

Abbildung 3: Beispiel Entfernungsklasse 20m bis 100m



Um eine einheitliche Datengrundlage zu gewährleisten, werden bei allen Beobachtungen der Entfernungsklassen bis 100m die Geokoordinaten des ZKB-Algorithmus verwendet. Die Beobachtungen mit grösseren Abständen weisen keine räumliche Systematik auf. Aus diesem Grund wurden die Beobachtungen der Klassen ab "100m bis 200m" einzeln überprüft und entsprechend korrigiert.

Tabelle 5: Verteilung über die Entfernungsklassen

Entfernungsstufe	Anzahl	Anzahl Fehler ZKB-Algorithmus	Anteil Fehler ZKB-Algorithmus (%)
100 bis 200 m	1'181	8	0.68
200 bis 300 m	165	2	1.21
300 bis 400 m	105	11	10.48
400 bis 500 m	52	5	9.62
500 bis 1'000 m	132	21	15.91
1'000 bis 5'000 m	164	54	32.93
5'000 bis 10'000 m	7	2	28.57
über 10'000 m	8	0	0.00

Von den 1'814 Beobachtungen, bei denen die Koordinaten um mehr als 100m voneinander abweichen, werden bei 103 Objekten die Geokoordinaten der MSE übernommen. Bei den restlichen 1'711 Beobachtungen wird die Geokodierung des ZKB-Algorithmus verwendet.

4.1.1.4 Finale Ergebnisse der nachträglichen Geokodierung

Die Qualität der originalen Geokodierung der MSE lässt sich als gut bezeichnen. Um allerdings eine grösstmögliche Anzahl an Beobachtungen für die Modellierung nutzen zu können, wurde zusätzlich der in der ZKB entwickelte Geokodierungsalgorithmus eingesetzt. Ein weiterer Vorteil der Nachbearbeitung ist, dass eine einheitliche Grundlage für die Zuweisung von Mirkolagefaktoren am Standort verwendet werden kann.

Insgesamt können 204'383 Beobachtungen adressgenau geokodiert werden und stehen somit der Modellierung zur Verfügung. Das entspricht 87.3% der gesamten MSE-Erhebung. Dagegen bleiben 29'770 Beobachtungen (12.7%) auch nach der zusätzlichen Geokodierung für die Modellierung unbrauchbar.

Von den ursprünglichen Koordinaten der MSE werden 20'498 übernommen, welche sich wie folgt aufteilen. Bei insgesamt 11'393 Beobachtungen kam es zu identischen Treffern (metergenau). Bei 9'002 Beobachtungen konnte der ZKB-Algorithmus keine Adresse zuordnen. Bei der händischen Korrektur der Entfernungsstufen über 100m (BfS vs ZKB) wurden bei 103 Beobachtungen die originalen Geokoordinaten der MSE übernommen.

Von der Geokodierung mittels des ZKB-Algorithmus werden insgesamt 183'885 Geokoordinaten übernommen. 18'608 Beobachtungen konnten im Vergleich zum BfS zusätzlich geokodiert werden. Bei 165'277 Beobachtungen wurde die Geokordinate des ZKB-Algorithmus übernommen. Davon stellen 163'566 Umkodierungen die Verschiebung der Koordinate von der Gebäudemitte zum Hauseingang dar, entsprechen also einer Standardisierung. Bei der händischen Korrektur der Entfernungsstufen über 100m wurden bei 1'711 Beobachtungen die ZKB-Geokoordinaten übernommen.

Tabelle 6: Übersicht über die verwendeten Geokoordinaten (Herkunft versus Differenzen)

Kategorie	Gesamt	Herkunft	
		MSE-Koordinaten	ZKB-Koordinaten
alter Treffer (MSE)	9'002	9'002	0
neuer Treffer (ZKB)	18'608	0	18'608
identische Treffer	11'393	11'393	0
Distanz:			
kleiner 10 m	139'722	0	139'722
10 bis 20 m	14'344	0	14'344
20 bis 100 m	9'500	0	9'500
100 bis 200 m	1'181	8	1'173
200 bis 300 m	165	2	163
300 bis 400 m	105	11	94
400 bis 500 m	52	5	47
500 bis 1'000 m	132	21	111
1'000 bis 5'000 m	164	54	110
5'000 bis 10'000 m	7	2	5
über 10'000 m	8	0	8
kein Treffer	29'770	0	0
Gesamtzahl	234'153	20'498	183'885

Im Anhang befinden sich zwei Karten (Abbildung 14 bzw. Abbildung 15) mit der Verteilung der geokodierten bzw. der nicht-geokodierten MSE-Beobachtungen pro Postleitzahlgebiet.

4.2 Geodaten

In diesem Kapitel werden die zur Verfügung stehenden Geodaten beschrieben, welche für die Modellierung zur Verfügung standen. Die Übersichtskarten zu den jeweiligen Datensätzen befinden sich im Anhang.

4.2.1 Lärmdaten

Mit dem Projekt SonBase hat das Bundesamt für Umwelt (BAFU) erstmalig eine Datenbasis geschaffen, die flächendeckend die Lärmbelastungen für die Schweiz darstellt. Diese liegt für die unterschiedlichen Lärmquellen aus Strassen-, Bahn- und Flugverkehr in einem Raster von 10m x 10m vor. Die Daten liegen für den Tagesverlauf (06.00 bis 22.00) und für die Nacht (22.00 bis 06.00) vor.

4.2.2 Topographische und klimatische Faktoren der Mikrolage

Flächendeckende Faktoren zur Beschreibung lokaler topographischer und klimatischer Verhältnisse wurden von der Firma Ernst Basler + Partner AG seitens der ZKB eingekauft. Insgesamt stehen 7 Variablen zur Verfügung. Alle Variablen mit Ausnahme der Wolkenbedeckung stehen in einer Auflösung von 25 Metern zur Verfügung.

Aussicht

Die Variable Aussicht stellt eine Quantifizierung der theoretisch sichtbaren Landschaftsfläche dar. Die theoretische Sicht wird in einem Radius von 30km um den Standort berechnet, wobei eine Aussichtshöhe von 2m über Grund angenommen wird. Meteorologische Einschränkungen sowie Gebäude und Vegetation werden in der Berechnung nicht berücksichtigt.

Seesicht

Mit der Variable Seesicht wird die theoretische Sicht auf stehende Gewässer mit einer Fläche grösser als 0.02qkm abgebildet. Es werden Gewässer mit einer maximalen Distanz bis zu 10km zum Standort in Betracht gezogen, wobei wiederum eine Aussichtshöhe von 2m über Grund angenommen wird. Meteorologische Einschränkungen sowie Gebäude und Vegetation werden nicht berücksichtigt.

Bergsicht

Mit der Bergsicht werden die Anzahl sichtbarer Berggipfel gemessen. Berücksichtigt werden die höchsten oder dominantesten Gipfel der Schweiz gemäss dem Schweizerischen Gipfelverzeichnis (www.gipfelverzeichnis.ch). Es wird eine Aussichtshöhe von 2 m über Grund und eine maximale Sichtweite von 120km angenommen. Meteorologische Einschränkungen sowie Gebäude und Vegetation werden nicht berücksichtigt.

Hangneigung in Prozent und Grad

Die Steilheit des Geländes hat einerseits einen Einfluss auf die Baukosten und somit auf Eigentumspreise bzw. Mieterträge. Andererseits können über die Hangneigung und die Exposition besonders gute Hanglagen wie Süd-/Südwestlagen definiert werden. Die Hangneigung wird gemessen in Grad oder Prozent. Der prozentuale Höhenunterschied entspricht dem Verhältnis aus Höhenunterschied und waagerechter Strecke. Als Beispiel: 10m Höhenunterschied auf 100m = 10% oder 5.7 Grad.

Exposition

Die Lage eines Hanges in Bezug auf die Flächennormale wird als Exposition (Ausrichtung) bezeichnet. Die Expositionsrichtung ist in Grad bezogen auf die Nordrichtung (0-360 Grad) angegeben:

- Exposition = 0°: Hang, der nach Norden ausgerichtet ist ("Schattenseite")
- Exposition = 90°: Osthang, Hang mit viel Morgensonne
- Exposition = 180°: Südhang, "Sonnseite"
- Exposition = 270°: Westhang, Hang mit viel Abendsonne (aber in Mitteleuropa am meisten den Witterungseinflüssen ausgesetzt)
- Exposition = -1°: flaches Gelände

Sonnenscheindauer Sommer und Winter

Theoretische Zeitdauer in Stunden, in der eine wolkenlose und ungetrübte klare Atmosphäre an einem bestimmten Ort Sonnenschein ermöglicht. Neben der Sonnenscheindauer im Sommer wird auch die Dauer im Winter berechnet. Die Sonnenscheindauer hängt neben dem Breitengrad und dem Landschaftshorizont auch von der Jahreszeit des betreffenden Standorts ab. Als Sommerwert ist der 21. Juni 2008 (Sommersonnenwende) und als Winterwert der 21. Dezember 2008 (Wintersonnenwende) dargestellt.

Wolkenbedeckung

Durchschnittliche und über 6 Jahre (August 2002 bis September 2008) gemittelte Bewölkung über der Schweiz (in Prozent). Die Datengrundlage "Bewölkungsgrad" stammt vom Satelliten AVHRR (Advanced Very High Resolution Radiometer) und wurde durch die Gruppe Fernerkundung (RSGB) des Geographischen Instituts der Universität Bern bezogen. Die Auflösung beträgt 1000m.

Seelagen

Um Lagen am See besser abbilden zu können, wurde in einem Radius von 10 km für alle Rasterpunkte die Distanz zum See in Metern gemessen. Daraus resultieren zwei Variablen. Zum Einen die direkte Seelage, welche für alle Objekte gilt, die weniger als 200 m vom See entfernt liegen. Zum Anderen wird die Distanz zum See in Metern gemessen.

4.2.3 Geodaten zur Beschreibung der Makrolage

Reisezeiten

Grundlage zur Berechnung der Reisezeiten bildet die Reisezeitmatrix vom Bundesamt für Raumentwicklung (ARE) von 2009. Die Reisezeiten stehen dabei für alle Schweizer Gemeinden zur Verfügung. Die 10 grössten Stadt-

Gemeinden (Zürich, Bern, Genf, Winterthur, Thun, Luzern, Biel, St. Gallen, Lausanne, Basel) werden zudem in Kreise aufgeteilt.

Unterschieden wird nach der Reisezeit mit dem öffentlichen Verkehr, dem unbelasteten Individualverkehr sowie dem belasteten Individualverkehr. Im Modell wird mit der Reisezeit die jeweils minimale Reisezeit aus öffentlichem Verkehr bzw. unbelastetem Individualverkehr gemeint. Daraus wird für jede Gemeinde die Reisezeit zum nächsten Grosszentrum oder zum nächsten Mittelzentrum berechnet.

Erreichbarkeitspotenzial

Als weiterer Indikator zur Beschreibung der Zentralität wird das Erreichbarkeitspotenzial für jede Hektare berechnet. Dabei wird die Summe der Arbeitsplätze in einem Umkreis von 20km, gewichtet mit $1/\text{Distanz}^2$ (Gravitationsmodell), ermittelt. Die Anzahl Arbeitsplätze pro Hektare stammen aus der Betriebszählung 2008 (Quelle: BFS).

Steuerbelastung

Die Steuerbelastung eines Haushalts in verschiedenen Gemeinden⁶ wurde von der Firma Braingroup für drei Haushaltstypen und drei Nettoeinkommen berechnet. Relevant ist die gesamte Steuerbelastung, d.h. die Summe aus den Gemeindesteuern und den Steuern von Kanton und Eidgenossenschaft. Kirchensteuern werden nicht berücksichtigt.

Für jede Gemeinde gibt es somit 9 Kennzahlen für die Steuerbelastung. Unterschieden werden die Haushaltstypen „allein stehend“, „verheiratet ohne Kinder“ und „verheiratet mit 2 Kindern“ und die drei Nettoeinkommen 50'000, 100'000 und 200'000 Franken.

Diese Kennzahlen werden zur besseren Vergleichbarkeit auf 1 Franken Nettoeinkommen normiert.

Tabelle 7: Übersicht Steuerklassen

Steuerbelastung (%) /Haushaltstyp	Nettoeinkommen FR. 50'000	Nettoeinkommen FR. 100'000	Nettoeinkommen FR. 200'000
Alleinstehend	Mittel = 12%	Mittel = 20%	Mittel = 28%
Verheiratet ohne Kinder	Mittel = 7%	Mittel = 14%	Mittel = 24%
Verheiratet mit 2 Kindern	Mittel = 3%	Mittel = 10%	Mittel = 21%

Für ein Haushaltseinkommen von 100'000 bezahlen Alleinstehende im Durchschnitt über alle Gemeinden 20% Steuern. Die Belastung schwankt dabei von 9% in der günstigsten Gemeinde bis zu 27% in der teuersten.

Steuerbares Einkommen pro Gemeinde

Der Median-Ertrag der steuerbares Einkommen ist gesamtschweizerisch vergleichbar und liefert einen Hinweis auf die wirtschaftliche Leistungsfähigkeit der in einer Gemeinde wohnhaften Personen. Die im Modell verwendeten Da-

⁶ BFS-Nummern gemäss 2009

ten stammen aus dem Jahre 2003 und wurden von der Eidgenössischen Steuerverwaltung (ESTV) zur Verfügung gestellt.

4.3 Datenbereinigung

In diesem Abschnitt werden die MSE-Daten im Detail beschrieben. Die Datenbereinigung sowie die Variablenselektion werden vertieft dargestellt.

Der MSE-Datensatz besteht insgesamt aus mehr als 100 Variablen und 234'153 Beobachtungen für die ganze Schweiz. Damit stellt die MSE die umfassendste Beschreibung des schweizerischen Mietmarktes dar. Für die weitergehenden Analysen wurden 112'957 Mieter⁷ berücksichtigt. Die Tabelle 8 zeigt die Beobachtungsauswahl (wie im MSE2003 Methodenbericht: Gewichtung und Schätzverfahren).

Tabelle 8: Mengengerüst zu den Selektionsschritten der MSE-Daten

Variable	Ausprägungen	Anzahl
workflow_status	7=Bereit für Auswertung	194'673
	8=inaktivierte Fragebogen (Heime etc.)	38'441
	9=inaktiviert da unplausible Werte	1'039
adresse_typ	1=Hauptwohnsitz	207'500
	2=Ferien-/ Zweitwohnung	6'720
	3=Geschäftsadresse	7'114
	Missing	12'819
wohnot_ok	1=ausgedruckte Adresse ist Wohnadresse	194'717
	2=Person ist umgezogen	12'188
	Missing	27'248
wohn_status	1=Mieter	110'878
	2=Untermieter	1'078
	3=Mitglied Wohnbaugenossenschaft	3'108
	4=Alleineigentümer	68'800
	5 und 6=Miteigentümer	12'836
	Missing	37'453
adresse_subtyp	1=Eigentümer	82'437
	2=Mieter	116'741
	3=Untermieter	1'024
	4=Möblierte Wohnung	1'596
	5=Untermieter ein Teil der Wohnung	1'044
	6=Pachtvertrag	496
	7=Dienstwohnung	488
	8=Gratiswohnung	4'148
	Missing	26'179
Total für die Auswertungen		112'957
Selektionskriterien		

Nach Berücksichtigung der ZKB-Geokodierung und dem Geokodierungs-Abgleich verbleiben von diesen 112'957 Objekten noch 106'691 Objekte, das sind 6'221 Objekte mehr als mit der ursprünglichen BfS-Geokodierung allein. Die Datenverluste aufgrund fehlender Geokodierbarkeit konnten somit

⁷ Für die Analyse wurden nur diejenigen MSE-Beobachtungen berücksichtigt, die mit den Selektionskriterien der MPI-Stichprobe übereinstimmen. D.h. Mietwohnungen, die dauerhaft zu Wohnzwecken gemietet sind ohne: Eigentümerwohnungen, Geschäftsadressen, Zweitwohnungen, Kollektivhaushalte und Spezialfälle (Gratiswohnungen usw.).

halbiert werden und betragen lediglich rund 5.5% (6'266) des Ausgangsdatensatzes.

4.3.1 Ausschlusskriterien

Für die Modellierung müssen die verwendeten Variablen aller Beobachtungen Werte aufweisen. Daher sind nicht alle Datensätze für die Analyse verwendbar. Ein Teil der Daten ist unvollständig oder wurde fehlerhaft von den damaligen Mietern und Mieterinnen ausgefüllt, wodurch sich das Mengengerüst für die Modellierung weiter reduziert. Grundsätzlich wollen wir so wenige Beobachtungen aus der Analyse ausschliessen wie möglich. Die Datenbereinigungskriterien mit den entsprechenden Ausschlussmengen sind in der folgenden Tabelle 9 dargestellt.

Tabelle 9: Ausschlusskriterien und Mengengerüst (Zwischensample, N=106'691)

Variable	Ausschlusskriterium	Wert	Anzahl	In Prozent Zwischensample
Nettomiete	Unplausible Werte	0, 99999	494	0.5%
	Missing	.	477	0.5%
Wohnfläche	Unplausible Werte	999	3'283	3%
Anzahl Zimmer	Unplausible Werte	99	526	0.5%
	Zimmerzahl grösser 6	7 bis 14	1'011	1%
Bauperiode (Gebäudealter)	Missings (nach Abgleich mit GWR Daten)	.	2'847	2.7%
Verweildauer (jahr_einzug)	Unplausible Werte	<=1936	314	0.3%
Stockwerk	Unplausible Werte	99	667	0.6%
	Stockwerk kleiner 0	-9 bis -1	1009	0.9%
	Stockwerk grösser 20	21 bis 69	101	0.1%
Summe			9'162	8.6%

Die meisten Beobachtungsausschlüsse finden bei den Kriterien Wohnfläche (3%) und Bauperiode (2.7%) statt. Hier werden mit definierten Grenzwerten unplausible Werte ausgeschlossen, die das Modellergebnisse verfälschen würden. Bei der Variable Anzahl Zimmer wurden alle Beobachtungen mit mehr als 6 Zimmern aus der Analyse ausgeschlossen (1%). Diese Entscheidung wurde im Hinblick auf die Vergleichbarkeit mit der MPI-Berechnung getroffen, welche nur Wohnungen mit 1 bis 6 Zimmern berücksichtigt.

Insgesamt reduziert sich der Stichprobenumfang damit um weitere 9'162 Objekte (8.6%). Als Grundlage für die Modellierung wird somit ein Endstichprobenumfang von 97'529 Beobachtungen benutzt (Tabelle 10). Dies entspricht 86% der Anfangstichprobe.

Tabelle 10: Stichprobenumfang

Sample	Anzahl	In Prozent
MSE 2003 –Mietwohnungen	112'957	100
Endstichprobe –für Modellierung	97'529	86
Ausgeschlossen aus der Analyse	15'428	14

4.3.2 Nettomiete

Die Zielgrösse des hedonischen Mietpreismodells und damit die wichtigste Variable für die Analyse ist die monatliche Nettomiete. Diese wird im Folgenden untersucht. Die Tabelle 11 zeigt für welche Beobachtungen die Preisangaben (Bruttomiete, Nettomiete und Nebenkosten) vorhanden sind. Bei 29'187 Beobachtungen wurden fehlende Angaben zur Nettomiete vom BFS imputiert (d.h. nach statistischen Regeln vervollständigt). Für die weitergehenden Analysen und Modellierung übernehmen wir diese Imputation.

Tabelle 11: Analyse der Nettomiete (Häufigkeit)

Code	Bruttomiete (BM)	Nettomiete (NM)	Nebenkosten (NK)	N	In Prozent	Beschreibung
1	–	x	–	4'643	4.4	OK
2	x	–	–	28'545	26.8	Nettomiete imputation (von BFS)
3	x	x	–	634	0.6	BM=NM, Nettomiete imputation (von BFS)
4	–	x	x	735	0.7	OK
5	x	x	x	72'134	67.6	OK
6	x	–	x	0	0	Keine Beobachtungen
7	–	–	x	0	0	Keine Beobachtungen
8	–	–	–	0	0	Keine Beobachtungen
Total				106'691	100	

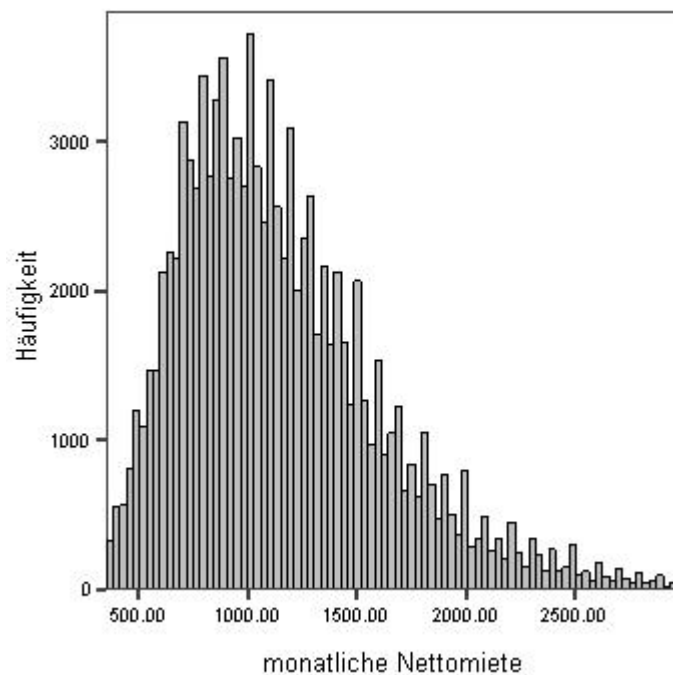
Teilweise übernommen aus Methodenbericht MSE2003, Gewichtung und Schätzverfahren

Im nächsten Schritt wurden die unplausiblen und fehlenden Werte aus der Analyse ausgeschlossen (vgl. Tabelle 9). Als Variable für die Modellierung errechnen wir danach die Nettomiete per Mietperiode. Die Mietperiode ist bei über 99% der Fälle „monatlich“, bei 0.2% „quartalweise“ und bei 0.5% ist es „jährlich“.

Die durchschnittliche Nettomiete in die Stichprobe beträgt 1'170 Fr, der Median ist 1'070 Fr. Die folgende Abbildung 3 zeigt die Verteilung der Nettomieten, begrenzt für Beobachtungen zwischen dem 1%-Quantil (350 Fr) und dem 99%-Quantil (3'000 Fr). Die Verteilung ist erwartungsgemäss rechtsschief.

Man kann unerwartete Sprünge bei runden Werten wie z.B. 1000, 1100, 1200, 2000 usw. aus dem Histogramm identifizieren. Dies hängt eng mit der Datenerhebungsmethode zusammen. Die Daten wurden mittels einer Umfrage bei Mietern und Mieterinnen gesammelt. Es ist wahrscheinlich, dass gewisse Mieter nicht genau die Aufteilung von Miete zwischen Nettomiete und Nebenkosten wissen und deswegen wurden die Werte teilweise gerundet. Bei den imputierten Nettomieten ist dieser Unterschied teilweise geglättet (Abbildung 32 im Anhang). Das gleiche Bild kann auch bei der Wohnfläche beobachtet werden.

Abbildung 3: Verteilung der Nettomieten (Histogramm)



4.3.3 Weitere Variablen

Einige Variablen wie Wohnfläche, Anzahl Zimmer (ohne Küche und Badezimmer), Gebäudealter, Verweildauer und weitere werden im Folgenden dargestellt.

4.3.3.1 Gebäudealter – Abgleich mit GWR-Daten

Die Tabelle 12 zeigt den Abgleich von Bauperiode (und Baujahr) von MSE2003 Daten und dem Gebäude- und Wohnungsregister (GWR)⁸. Wie aus der Tabelle ersichtlich ist, wurde der grösste Anteil (20 Prozent) der Gebäude zwischen 1971 und 1980 gebaut. Es resultierten 2'847 fehlende Beobachtungen, was rund 3 Prozent der Gesamtstichprobe entspricht. Diese Beobachtungen lagen in der MSE2003 in der Kategorie „vor 1971 gebaut“, wiesen aber im GWR keine Altersangaben auf.

Tabelle 12: Bauperiode, Abgleich von MSE2003 und GWR Daten

	MSE	N	In Prozent	2003-2001	2000-1999	1998-1996	1995-1994	1993-1991	1990-1986	1985-1981	1980-1971	vor 1971
GWR												
2001-2003	327	0.3		275	31	6	2	3	1	2	1	6
1996-2000	5'172	5.0		22	659	2'316	188	84	135	188	410	1'170
1991-1995	4'293	4.2		2	4	117	1'320	2'557	164	16	21	92
1986-1990	5'533	5.4		3	5	12	8	47	4'971	251	108	128
1981-1985	4'976	4.9		3	0	2	3	4	100	4'666	159	39
1971-1980	14'977	14.6		6	7	18	12	8	26	142	14'024	734
1961-1970	19'620	19.1		9	5	17	20	9	20	28	338	19'174
1946-1960	16'591	16.2		18	9	13	5	9	11	14	66	16'446
1919-1945	10'812	10.5		13	10	12	6	5	13	11	22	10'720
vor 1919	14'149	13.8		41	44	46	30	43	38	46	94	13'767
Missing	6'147	6.0		449	234	303	311	613	552	380	735	2'570
Total	102'597	100		841	1'008	2'862	1'905	3'382	6'031	5'744	15'978	64'846

⁸ BFS

4.3.3.2 Wohnfläche und Zimmerzahl

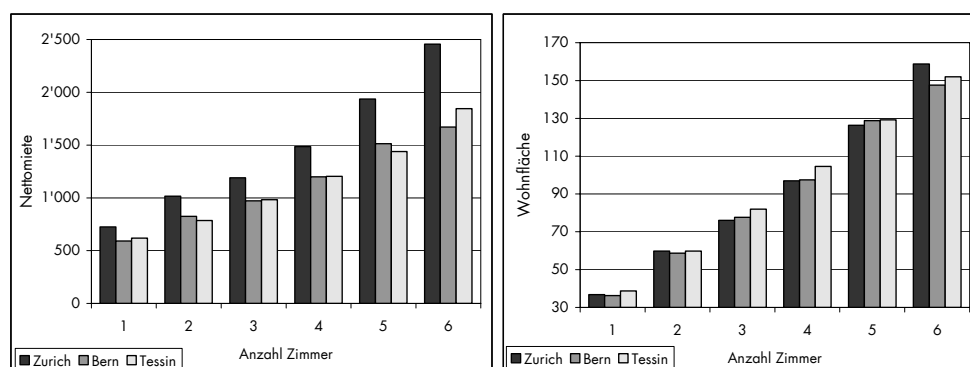
In der folgenden Tabelle 13 ist die durchschnittliche Nettomiete und Wohnfläche nach Zimmerzahl ersichtlich. Wie erwartet steigt die Nettomiete und Wohnfläche mit steigender Zimmerzahl. Diese Steigerung ist aber keine kontinuierliche. Eine 1 – Zimmerwohnung mit 37 Quadratmetern kostet 640 Fr und eine 2-Zimmerwohnung mit rund 60 Quadratmetern kostet 870 Fr. Eine Wohnung mit 6-Zimmern kostet dreimal so viel wie eine 1-Zimmerwohnung und ist dabei viermal grösser.

Tabelle 13: Durchschnittliche Nettomiete und Wohnfläche pro Zimmer

Zimmerzahl	Anzahl	In Prozent	Durchschnittliche Nettomiete	Durchschnittliche Wohnfläche
1	3'736	3.7	640	37
2	16'162	15.9	867	58
3	37'823	37.2	1'059	78
4	32'987	32.5	1'307	99
5	8'587	8.5	1'646	126
6	2'291	2.3	1'919	155
Total	101'586	100		

Fast 40 Prozent der Stichprobe sind 3-Zimmerwohnungen. Zusammen mit den 4-Zimmerwohnungen entspricht das rund 70 Prozent der Stichprobe. Die 6- und 1-Zimmerwohnungen stellen nur einen kleinen Teil der Stichprobe dar, 2 bzw. 4 Prozent. Abbildung 4 dient als Beispiel für kantonale Unterschiede: verglichen werden die Kantone Zürich, Bern und Tessin.

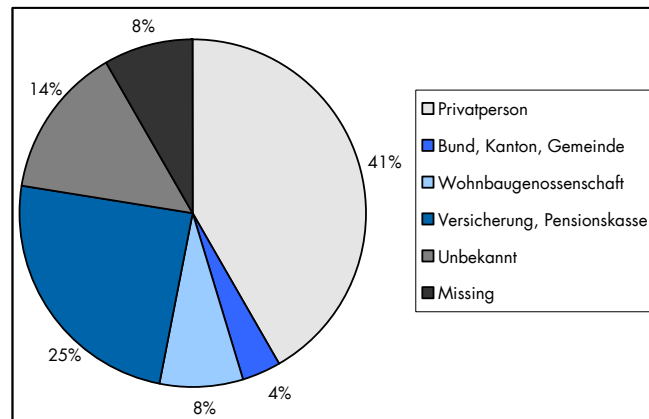
Abbildung 4: Durchschnittliche Nettomiete pro Zimmer, Vergleich für Kanton Zürich, Zug und Tessin



4.3.3.3 Eigentübertyp

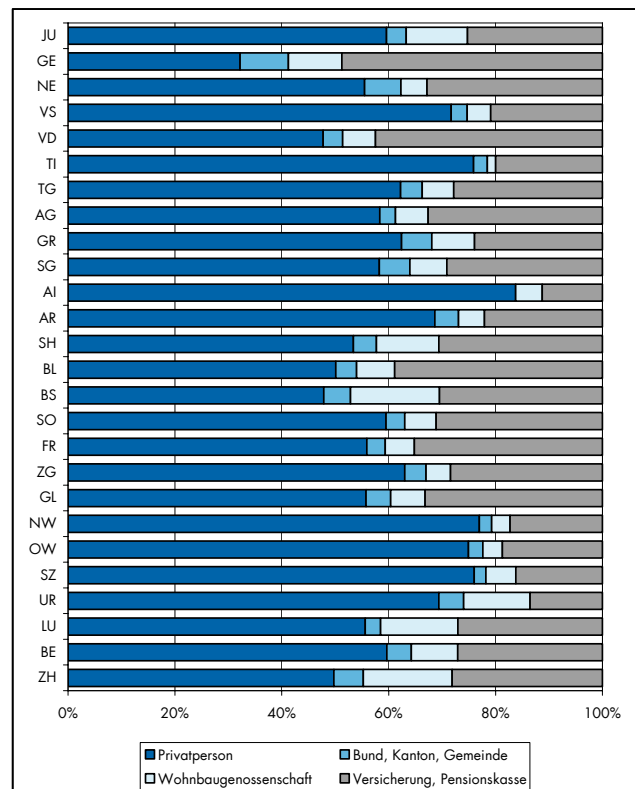
Diese Variable beschreibt, ob die Wohnung im Besitz einer Privatperson, des Bundes, eines Kantons, einer Pensionskasse oder einer Versicherungsgesellschaft ist. Zwischen öffentlichen/gemeinnützigen und privaten Besitzern bestehen Unterschiede im Preissetzungsverhalten, was zu markanten Mietpreisdifferenzen führt. Dies geht aus den empirischen Analysen der MSE03 sowie aus anderen Studien (Statistik Stadt Zürich, 2009) hervor. Das Merkmal ist deshalb im Mietpreismodell zu berücksichtigen (vgl. 3.5). Die Verteilung nach Eigentübertyp ist in Abbildung 5 dargestellt. Die Abbildung 6 zeigt die Unterschiede nach Kantonen.

Abbildung 5: Eigentübertypverteilung in der Stichprobe



Bei 22 Prozent (Unbekannt und Missing) der Beobachtungen ist der Eigentübertyp nicht bekannt. Um diese Beobachtungen dennoch für die Modellierung nutzen zu können, werden fehlende Angaben mit einem Imputationverfahren nachgefüllt. Die Methode und die Resultate sind in Abschnitt 5.1 beschrieben.

Abbildung 6: Verteilung Eigentübertyp pro Kanton



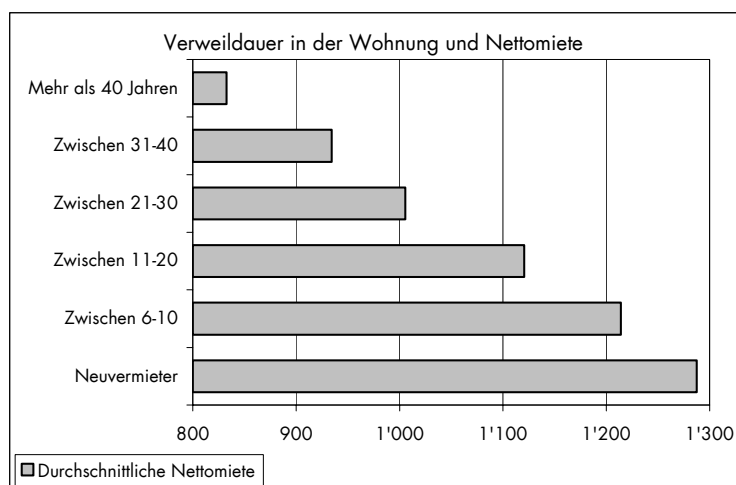
Die grössten Anteile von Wohnungen mit öffentlichen/gemeinnützigen Eigentümern (Wohnbaugenossenschaft, Bund, Kanton) weisen grossstädtische Kantone wie Zürich (22%), Basel-Stadt (21%) oder Genf (19%) auf. Den tiefsten kantonalen Anteil weist das Tessin mit 4% auf.

4.3.3.4 Verweildauer

Die Verweildauer (Dauer des Mietverhältnisses) bildet zwar kein Wohnungs- oder Gebäudemerkmal, wirkt sich infolge der Mietregulierung jedoch auf die Mietpreisentwicklung in laufenden Verträgen aus. Wie wir in Abschnitt 3.4 ausgeführt haben, wird dieses Merkmal für die Modellierung verwendet, nicht aber für die Qualitätsbereinigung des MPI.

Die durchschnittliche Verweildauer beträgt 13 Jahre (Median: 8 Jahre). Abbildung 7 zeigt die unterschiedlichen Durchschnittsmieten in Abhängigkeit von der Verweildauer. Erwartungsgemäss fällt die Nettomiete bei längerer Verweildauer. Bei neuen Mietverhältnissen kostet eine Wohnung durchschnittlich rund 1'300 Fr., bei Altmietern rund 850 Fr.

Abbildung 7: Verweildauer in der Stichprobe



4.3.3.5 Weitere Wohnungsmerkmale

Im Folgenden werden weitere Wohnungsmerkmale wie Wohnungstyp, Cheminée usw. ausgewertet. Die folgende Abbildung 8 zeigt die Häufigkeit dieser Merkmale in der Stichprobe. Der Anteil an Beobachtungen mit „Missings“ ist insgesamt sehr klein. Für die variable Einbauküche beträgt er z. B. gerade einmal 0.2%, dies entspricht 244 Beobachtungen. Insgesamt (listwise) weisen nur 1% der Beobachtungen in der Stichprobe fehlende Werte auf.

Abbildung 8: Weitere Wohnungsmerkmale

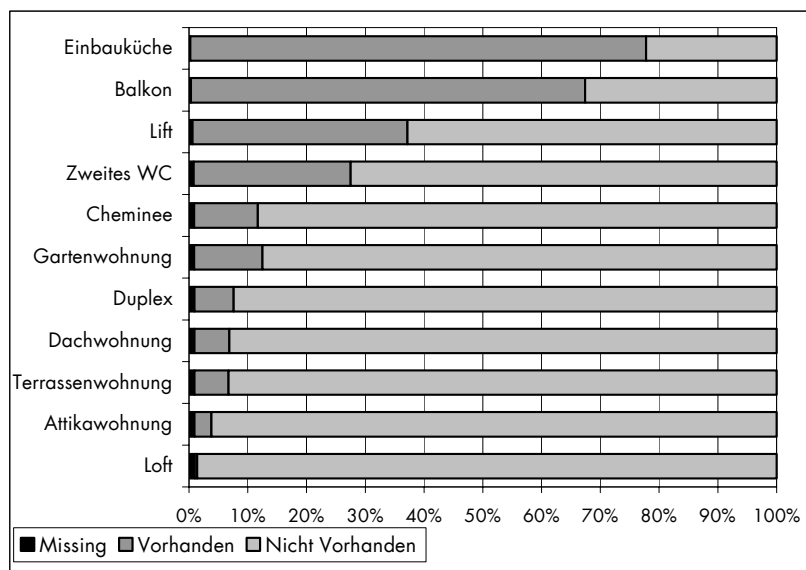
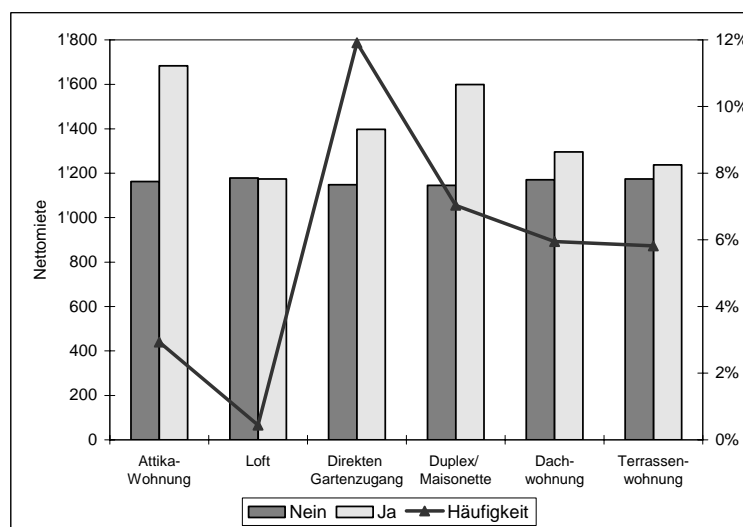


Abbildung 9 zeigt die Häufigkeiten und durchschnittliche Nettomieten für verschiedene Wohnungstypen. Die Stichprobe umfasst z.B. 2 Prozent Attikawohnungen mit einer durchschnittlichen Nettomiete von 1'680 Fr.

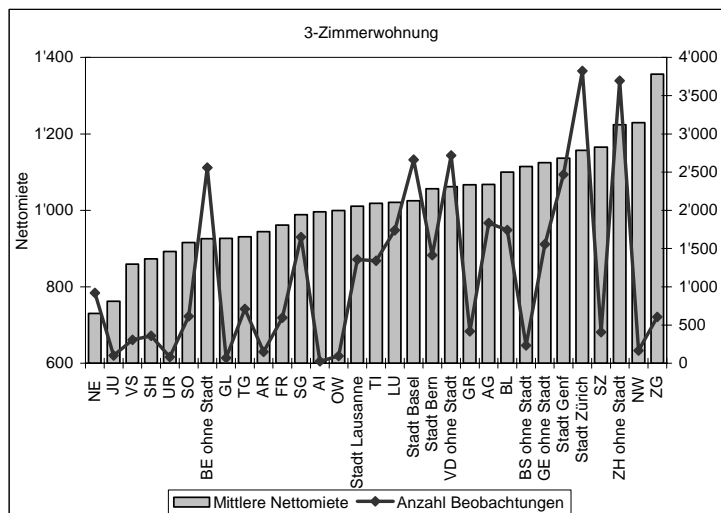
Abbildung 9: Häufigkeiten und durchschnittliche Nettomieten verschiedener Wohnungstypen



4.3.4 Regionale Unterschiede und Verteilung der Stichprobe

Abbildung 10 zeigt das Mengengerüst sowie durchschnittliche Nettomieten nach Kantonen und Grossstädten für eine 3-Zimmerwohnung.

Abbildung 10: Durchschnittliche Nettomieten und Häufigkeiten der 3-Zimmerwohnungen nach Kanton



FAZIT: Datengrundlage und Datenbereinigung

Mit der MSE 2003 steht uns eine qualitativ und quantitativ hervorragende Datenbasis zur Verfügung. Nach der beschriebenen Datenbereinigung umfasst die für die Modellierung verfügbare Endstichprobe 97'529 Beobachtungen. Einzig bei der Variable Eigentübertyp wurden fehlende Werte mit statistischen Verfahren ersetzt. Das Vorgehen und die Ergebnisse werden im nächsten Kapitel beschrieben.

5 Datenimputationen und -transformationen

5.1 Imputation fehlender Werte der Eigentumsform

Die Variable Eigentübertyp stellt für die Schätzung eine sehr wichtige Variable dar. Insgesamt fehlen bei 21'753 Beobachtungen (22%) entsprechende Informationen. Diese lassen sich mit statistischen Methoden - sogenannten Imputationsverfahren – anhand vorhandener Informationen ersetzen.

Tabelle 14: Missingsanalyse Eigentübertyp

Eigentübertyp	Anzahl	Anteil	Kumulative Anzahl	Kumulativer Anteil	Kumulativer Anteil
Privatperson	41'363	41.4	41'363	41.4	53
Bund, Kanton, Gemeinde	3'769	3.8	45'132	45.2	58
Wohnbaugenossenschaft	7'911	7.9	53'043	53	68
Versicherung, Pensionskasse	25'000	25	78'043	78	100
Unbekannt	13'740	13.8	91'783	91.8	-
keine Angabe	8'013	8.2	99'796	100	-

Um die Qualität verschiedener Imputationsverfahren zu testen wurden zunächst alle Beobachtungen mit bekanntem Eigentübertyp selektiert (Insample). Bei 10% dieser Beobachtungen (Zufallsstichprobe) wird der bekannte Eigentübertyp auf „Missing“ gesetzt. Die statistischen Imputationsregeln werden mit dem ganzen Insample modelliert und dann auf diese 10% (Testsample) angewendet. Aus dem Vergleich der imputierten Werte mit den bekannten Werten lässt sich die Trefferquote und damit die Qualität der Imputation beurteilen. Für die Imputation des Eigentübertyps als kategoriale Variable wurde ein Regressionsverfahren gewählt, welches die Möglichkeit mehrfacher Imputationen erlaubt. Eine Beschreibung über die Vorgehensweise im Statistikprogramm SAS findet sich in Yang, C. 2002.

Die Imputation fehlender Werte lässt sich in drei Teilprozesse gliedern.

- Im ersten Schritt werden die fehlenden Werte mehrfach imputiert (proc mi). In das Schätzmodell gingen folgende erklärende Variablen ein: Nettomiete, Wohnfläche, Bauperiode, Dauer des Mietverhältnisses sowie die regionale Verteilung (auf Gemeindeebene) zwischen den Eigentümerformen. Diese sind unterteilt in: Anteil private Vermieter, Anteil Wohnbaugenossenschaften und Anteil Versicherungen und Pensionskassen.
- Im zweiten Schritt werden die Ergebnisse der Imputationen mittels geeigneter Standardprozeduren analysiert (proc logistic, proc mianalyze). Dabei werden dieselben Modellvariablen eingesetzt wie im ersten Schritt.
- Im dritten Schritt werden die Ergebnisse zusammengeführt (proc score).

Nach Aufsetzen des Algorithmus wird anschliessend noch der Einfluss verschiedener multipler Imputationen getestet. Nachfolgend die Ergebnisse:

- 1-fache Imputation (logistische Regression): 45,6% Trefferquote
- 5-fache Imputation (logistische Regression): 53,7% Trefferquote
- 10-fache Imputation (logistische Regression): 53,7% Trefferquote

Bei 10facher Imputation wird trotz des höheren Aufwandes keine Verbesserung des Ergebnisses erzielt. Aus diesem Grund wird die 5-fache Imputation für den Eigentübertyp verwendet. Tabelle 15 zeigt die Verteilung des Eigentübertyps nach Imputation im Insample, Tabelle 16 diejenige in der Gesamtstichprobe.

Tabelle 15: Verteilung Insample Eigentübertyp

Eigentübertyp	Anzahl	Anteil	Kumulative Anzahl	Kumulativer Anteil
Privatperson	41'330	53	41'330	53
Bund, Kanton, Gemeinde	3'810	4.9	45'140	57.9
Wohnbaugenossenschaft	7900	10.1	53'040	68
Versicherung, Pensionskasse	25'003	32	78'043	100

Tabelle 16: Verteilung Imputation Eigentübertyp gesamte Stichprobe

Eigentübertyp	Anzahl	Anteil	Kumulative Anzahl	Kumulativer Anteil
Privatperson	50'533	50.6	50'533	50.6
Bund, Kanton, Gemeinde	4'814	4.8	55'347	55.4
Wohnbaugenossenschaft	10'061	10.1	65'408	65.5
Versicherung, Pensionskasse	34'388	34.5	99'796	100

5.2 Datentransformationen

Im Hinblick auf erwünschte Verteilungseigenschaften werden die Zielvariable (Nettomiete) sowie einige der erklärenden Variablen entsprechend der Literatur logarithmiert (z.B. Triplett, 2004). Dies betrifft insbesondere die zur Beschreibung der Makrolage verwendeten Variablen Steuerfuss und Steuerkraft. Des Weiteren werden für kategoriale Variablen (z.B. Zimmerzahl) so genannte Dummy- bzw. Binärvariablen (mit Ausprägung 0 bzw. 1) gebildet. Einige der GIS-Variablen (z. B. Seesicht, Bergsicht) werden ebenfalls mit Binärvariablen abgebildet. Tabelle 17 und Tabelle 19 geben einen Überblick über sämtliche in den Schätzmodellen verwendeten Variablen und die entsprechenden Transformationen. Zusätzlich beinhalten die zwei Tabellen auch andere getestete Variablentransformationen. Für die Variable Gebäudealter haben wir z. B. vier mögliche Spezifikationen getestet:

- Kategorie-Dummies für die Bauperiode (GWR Definition von Perioden)
- Kategorie-Dummies für 4 Altersklassen (anhand von MPI Altersklassendefinition)
- Kontinuierliche Variable (für die Berechnung war Baujahr als Mittelpunkt der Bauperiode definiert) mit Alter im Quadrat

- Kontinuierliche Variable mit Alter im Quadrat und kubisches Alter.

In die endgültige Schätzung geht die Variable Gebäudealter mit linearem und quadratischem Term ein (geb_alter; geb_alter2). Die Strukturvariablen Zimmerzahl, Eigentübertyp und Wohnungstyp bilden den Kern der Schätzfunktion und werden als Dummy-Variablen (DV) oder Klassen-Dummies verwendet.

Tabelle 17: Datentransformationen, Strukturvariablen

Variable	Nennung	Transformation
Nettomiete	Innmiete	Log(Nettomiete)
Zimmergrösse	lnzigroes	Log(Wohnfläche/Zimmerzahl)
Zimmerzahl	zimmer1 * bis zimmer6	Dummyvariable für Zimmerzahl 1 - 6, (zimmer3; 1 wenn es eine 3-Zimmerwohnung ist, 0 sonst)
Gebäudealter	gbaup8011 * bis gbaup8020	Dummyvariable für Bauperiode (vgl. GWR)
	geb_alter	Kontinuierliche Variable, berechnet als Differenz zwischen Umfragejahr (2003) und Mittelpunkt von Bauperiode: Bauperiode vor 1919 => 93 Jahre Bauperiode 1919 - 1945 => 71 Jahre Bauperiode 1946 - 1960 => 50 Jahre Bauperiode 1961 - 1970 => 38 Jahre Bauperiode 1971 - 1980 => 28 Jahre Bauperiode 1981 - 1985 => 20 Jahre Bauperiode 1986 - 1990 => 15 Jahre Bauperiode 1991 - 1995 => 10 Jahre Bauperiode 1996 - 2000 => 5 Jahre Bauperiode 2001 - 2003 => 1 Jahr2
	geb_alter2	Alter * Alter
	geb_alter3	Alter * Alter * Alter
	alterk1 * bis alterk4	4 Alterklassen: alterk1 - Gebäudealter zwischen 0-3 Jahren alterk2 - Gebäudealter zwischen 4-12 Jahren alterk3 - Gebäudealter zwischen 13-22 Jahren alterk4 - Gebäudealter mehr als 23 Jahre
Verweildauer	dauer	2003 - jahr_einzug
Garage	gar_in	DV, 1 wenn Garage in Nettomiete inbegriffen, 0 sonst
Eigentübertyp	etyp5_23	DV, 1 wenn Eigentümer Wohnbaugenossenschaft, Bund, Kanton oder Gemeinde ist, 0 sonst (Eigentümer ist Privatperson, Versicherung oder Pensionskasse)
	etyp5_1 * bis etyp5_4	DV für Eigentübertyp etyp5_1 = Privatperson etyp5_2 = Bund, Kanton, Gemeinde etyp5_3 = Wohnbaugenossenschaft etyp5_4 = Versicherung, Pensionskasse, etc.
Wohnungstyp	spez_attika	DV für Attikawohnung
	spez_duplex	DV für Duplex - Wohnung

*Referenzkategorie

Bei der Wahl der zahlreichen Bestimmungsfaktoren musste auch beachtet werden, dass diese nicht zu stark miteinander korrelieren. Hohe Korrelationen der erklärenden Variablen werden als Multikollinearität bezeichnet. Diese verletzt die Annahmen der angewandten Schätzmethode der kleinsten Quadrate (OLS) und kann zu Problemen führen (u.a. Instabilität der Ergebnisse).

Die folgende Tabelle 18 zeigt die Korrelationen zwischen der Nettomiete und den Lärmvariablen⁹ am Tag und in der Nacht. Erwartungsgemäss sind Stras-

⁹ Die Lärmvariablen sind immer definiert ab einem gewissen Grenzwert (50dB für Lärm während der Tag bzw. 45dB für Lärm während der Nacht).

sen- und Bahnlärm negativ mit der Nettomiete korreliert. Was überrascht ist ein positiver Zusammenhang zwischen Fluglärm und der Nettomiete. Der Anteil Beobachtungen mit Fluglärm ist gering (6% für Tag bzw. 2% für Nacht), weist allerdings eine höhere Durchschnittsmiete als der Rest der Beobachtungen auf. Was hingegen nicht überrascht, ist die hohe Korrelation zwischen den Lärmvariablen Tag und Nacht, bei Strassenlärm beträgt sie z.B. 0.91. Das beeinflusst die Variablenwahl für die Schätzung. Für die Modellierung haben wir verschiedene Transformationen der Lärmvariablen getestet, unter anderem:

- kontinuierliche Variable mit verschiedenen Grenzwerten
- Dummy-Variable, die den Wert eins annimmt, wenn der Lärm während der Nacht über 50dB (55dB, 60dB) liegt, 0 sonst
- Dummy-Klassen für Lärm zwischen 50-55, 56-60, 61-65 und über 66dB
- Day-Night Average Sound Level¹⁰ (Nelson, 2009)

Der Strassenlärm während der Nacht berücksichtigt die stärkere Belästigung von Bewohnern als der Tageslärm (vgl. Banfi et al.).

Tabelle 18: Korrelationen zwischen Lärmvariablen und Nettomiete

Variable	Netto- miete	Strassenlärm		Bahnlärm		Fluglärm
		Tag	Nacht	Tag	Nacht	Tag
Nettomiete	1					
Strassenlärm Tag	-0.048	1				
Strassenlärm Nacht	-0.024	0.794	1			
Bahnlärm Tag	-0.014	0.020	0.026	1		
Bahnlärm Nacht	-0.015	0.021	0.025	0.910	1	
Fluglärm Tag	0.018	-0.012	0.010	0.008	0.003	1
Fluglärm Nacht	0.013	-0.018	-0.002	0.003	-0.001	0.681

Tag: über 50dB, Nacht: über 45dB

Weiter haben wir verschiedene GIS-basierte Variablen getestet wie Bergsicht und Seesicht.

¹⁰ Day-Night Average Sound Level (L_{DN}) accounts for the fact that a given noise is usually more annoying at night, and thus a 10dB penalty is given to noise at night.

$$L_{DN} = 10 \log \left[\frac{15}{24} 10^{(L_D/10)} + \frac{9}{24} 10^{((L_N/10)-10)} \right]$$

Tabelle 19: Datentransformationen, Variablen der Makro- und Mikrolage

Variable	Nennung	Transformation
Erreichbarkeitspotenzial	lnerreichpot	Log(erreichpotenzial)
Seeufer	dsee200	DV für Seeufer, 1 wenn Distanz zum See kleiner 200 Meter ist, 0 sonst
Distanz zum See	distanz_see	Kontinuierliche Variable, Distanz zum See im km
Seesicht	Seesicht_e0 bis seesicht_e3	Dummyvariable (Klassen): Keine Seesicht* Kleine Seesicht, 7 bis 100 ha Mittlere Seesicht, 101 bis 300 ha Gute Seesicht, > 300 ha
Bergsicht	bergsicht	DV für Anzahl sichtbare Gipfel: 0 – 2, keine Bergsicht 3 – 12, mittlere Bergsicht mehr als 12, gute Bergsicht
Neigung	slope0, slope9 und slopemax	DV für Hangneigung (in Grad): 0 – 4*, keine Hangneigung 5 – 9, mittlere Hangneigung grosser 9, hohe Hangneigung
Exposition	NW_GE	DV für NW Exposition im Kanton GE, berechnet als Interaktion zwischen NW und DV für Kanton GE
Strassenlärm	Laerm_tag50	Kontinuierliche Variable, berechnet aus Taglärm über 50dB, 0 sonst
	Laerm_nacht45	Kontinuierliche Variable, berechnet aus Nachtlärm über 45dB, 0 sonst
Bahnlärm	Bahn_tag50	Kontinuierliche Variable, berechnet aus Bahnlärm während des Tages über 50dB, 0 sonst
	Bahn_nacht45	Kontinuierliche Variable, berechnet aus Bahnlärm während der Nacht über 45dB, 0 sonst
Fluglärm	Flug_tag50	Kontinuierliche Variable, berechnet aus Fluglärm während der Tag über 50dB, 0 sonst
Steuerbelastung	AL_pro1	Mittlere Steuerbelastung pro Nettoeinkommen Für Alleinstehende mit Nettoeinkommen 50'000, 100'000 oder 200'000 Fr.
Steuerkraft	Insteuersatz	Steuerbares Einkommen
Grossstadt		DV für Grossstädte Zürich, Bern, Basel, Lausanne und Genf
Kanton		DV für Kantone

*Referenzkategorie

Die standortbezogenen Strukturvariablen berücksichtigen sämtliche 26 Kantone und 5 Grossstädte (Zürich, Bern, Basel, Lausanne und Genf). Diese fliessen in die Schätzung als Dummy-Variablen ein. Die Stadt Zürich bildet dabei die Referenzkategorie.

6 Modellevaluation und Ergebnisse

In diesem Kapitel stellen wir das Schätzmodell und die Ergebnisse vor. Wir behandeln den Umgang mit "Ausreißern" sowie die Residuenanalyse. Es folgen ein Vergleich verschiedener getesteter Modellvarianten sowie In- und Out-of-Sample Analysen. Zum Schluss wird die Modellanwendung auf den MPI besprochen.

6.1 Modellschätzung und Modellergebnisse

Die theoretischen Grundlagen für die Schätzung mit dem hedonischen Ansatz haben wir im Kapitel 3 sowie im Zwischenbericht eingehender erläutert. Die Grundidee dabei: Die Wohnungsmieter bezahlen für nutzenstiftende Eigenschaften der Wohnung. Die Zahlungsbereitschaften für diese einzelnen Merkmale werden mit einer multiplen linearen Regression berechnet. Der Marktwert bzw. die Miete einer Wohnung setzt sich somit zusammen aus der Summe der Einzelpreise ihrer Eigenschaften.

Zu erklärende Variable

Die monatliche Nettomiete der Wohnung bildet in unserem Fall die zu erklärende bzw. abhängige Variable. Die Nebenkosten wie bspw. die Heizkosten sind darin nicht inbegriffen.

Unabhängige Variablen

In Tabelle 20 werden die Variablen aufgelistet, welche in die Schätzgleichung einfließen. Die Variablen können in drei Gruppen unterteilt werden: Strukturvariablen sowie Variablen der Mikro- und Makrolage. Einige Variablen wurden zusätzlich transformiert (Logarithmierung, Bildung von Kategorien, Quadrierung). Diese Transformationen sind im Schätzwert (Tabelle 26) ersichtlich.

Im Folgenden werden die Schätzergebnisse für die doppeltlogarithmische¹¹ funktionale Form dargestellt, wobei auch Schätzungen mit der linearen und semilogarithmischen Funktion durchgeführt wurden. Bei empirischen Untersuchungen wird die doppeltlogarithmische Funktion am häufigsten verwendet. Zudem konnte festgestellt werden, dass die Signifikanz der erklärenden Variablen sowie die Erklärungskraft bei diesen Modellen vergleichbar ist.

¹¹ D.h. das Modell ist ein lineares Regressionsmodell mit der logarithmierten Nettomiete als Zielvariable (zu erklärende Variable) und Prädiktorvariablen (erklärende Variablen) aus Struktur-, Makro- und Mikrovariablen. Somit wirken die Eigenschaften multiplikativ.

Tabelle 20: Definition der Variablen

Merkmale der Wohnung	
Alter	Alter des Gebäudes, d.h. die Differenz zwischen Stichjahr und Baujahr
Zimmer	Anzahl bewohnbare Zimmer ohne Küche und Bad
Wohnfläche	Fläche der Wohnung in m ² (mit Küche und Badezimmer)
Stockwerk	Stockwerk, in welchem sich die Wohnung befindet
Garage	Die Wohnung verfügt über eine Garage (die Miete für die Garage ist in der Nettomiete inbegriffen)
Attika	Attikawohnung (Wohnung im obersten Geschoss, mit Dachterrasse)
Duplex	Duplexwohnung (Wohnung auf zwei Etagen)
Verweildauer	Dauer des Mietverhältnisses in der Wohnung (in Jahren)
Eigentübertyp	Der Eigentümer ist eine Genossenschaft, Stadt, Kanton, Versicherung oder Privatperson
Merkmale der Makrolage	
Steuerkraft	Steuerkraft in der Gemeinde
Steuerbelastung	Steuerbelastung in der Gemeinde
Erreichbarkeit	Anzahl erreichbare Arbeitsplätze (anhand Gravitationsmodell)
Grossstadt	Grossstadtspezifische Konstanten (Zürich, Bern, Lausanne und Genf)
Kanton	Kantonsspezifische Konstanten
Merkmale der Mikrolage	
Seeufer	Minimale Luftentfernung zwischen Seeufer und einem bestimmten Hektar ist kleiner als 200 m
Distanz zum See	Minimale Luftentfernung zwischen See und einem bestimmten Hektar (in m)
Seesicht	Sichtbare Seefläche (in ha) von einem bestimmten Hektar
Bergsicht	Anzahl sichtbare Bergen von einem bestimmten Hektar
Strassenlärm	Durchschnittlicher Nachtwert der Strassenlärmimmissionen (in dB) für einen bestimmten Hektar
Bahnlärm	Durchschnittlicher Tageswert der Bahnlärmimmissionen (in dB) für einen bestimmten Hektar
Neigung	Hangneigung in %
Exposition	Himmelsrichtung eines Hektars: Ausprägungen: keine Exposition (flach), Nord bis Süd (E), Süd bis Nord (O).

Die finale Modellspezifikation sieht formalisiert folgendermassen aus:

$$\ln(\text{Nettomiete}) = \text{Konstante} + \beta_1 \cdot \text{zimmer2} + \beta_2 \cdot \text{zimmer3} + \beta_3 \cdot \text{zimmer4} + \beta_4 \cdot \text{zimmer5} + \beta_5 \cdot \text{zimmer6} + \beta_6 \cdot \text{lnzigroes} + \beta_7 \cdot \text{wohn_stock} + \beta_8 \cdot \text{geb_alter} + \beta_9 \cdot \text{geb_alter2} + \beta_{10} \cdot \text{gar_in} + \beta_{11} \cdot \text{spez_attika} + \beta_{12} \cdot \text{spez_duplex} + \beta_{13} \cdot \text{etyp5_23} + \beta_{14} \cdot \text{Inerreichpot} + \beta_{15} \cdot \text{dsee200} + \beta_{16} \cdot \text{distanz_see} + \beta_{17} \cdot \text{seesicht_e1} + \beta_{18} \cdot \text{seesicht_e2} + \beta_{19} \cdot \text{seesicht_e3} + \beta_{20} \cdot \text{slope9} + \beta_{21} \cdot \text{slopemax} + \beta_{22} \cdot \text{SW} + \beta_{23} \cdot \text{NW_GE} + \beta_{24} \cdot \text{laerm_nacht45} + \beta_{25} \cdot \text{bahn_tag50} + \beta_{26} \cdot \text{bergsicht_d2} + \beta_{27} \cdot \text{bergsicht_d3} + \beta_{28} \cdot \text{AL_pro1} + \beta_{29} \cdot \text{lnskraftmed} + \beta_{30} \cdot \text{Bern} + \beta_{31} \cdot \text{Lausanne} + \beta_{32} \cdot \text{Genf} + \beta_{33-58} \cdot \text{kantonspezifische Konstante} + \varepsilon$$

Tabelle 21 zeigt die Ergebnisse der linearen Regression der logarithmierten Nettomieten auf die Attribute der Wohnungen. Das Schätzverfahren basiert auf der Kleinst-Quadrate-Methode (OLS).¹² Die Schätzungen erfolgten mit dem Statistikprogramm SAS.

Der Erklärungsgehalt des Modells darf als hoch bezeichnet werden. Das Modell weist ein korrigiertes R² von rund 66 Prozent auf, d.h. 66 Prozent der Va-

¹² Z.B. Greene (1997).

rianz werden durch das Modell erklärt. Der Standardfehler der Regression liegt bei 22 Prozent, das bedeutet das rund zwei Drittel der Beobachtungen im Bereich von plus/minus 22 Prozent um den wahren Wert streuen. Die erweiterten Ergebnisse (VIF und Residuenplots) sind in Anhang 7.4.1 ersichtlich.

Tabelle 21: Schätzergebnisse

	Schätzwert	Std. Abw.	t-Wert	Pr > t		Schätzwert	Std. Abw.	t-Wert	Pr > t
Objekteigenschaften					Makrovariablen				
Alter	-0.010160	0.000	-85.1	<.0001	Erreichbarkeitspotenzial	0.0542	0.001	46.5	<.0001
Alter im Quadrat	0.000075	0.000	66.9	<.0001	Steuerkraft pro 1 CHF	-0.2774	0.047	-6.0	<.0001
In (Wohnfläche/Anzahl Zimmer)	0.3729	0.003	108.0	<.0001	In Steuersatz	0.5335	0.012	42.9	<.0001
2-Zimmerwohnung	0.4105	0.005	90.9	<.0001					
3-Zimmerwohnung	0.6833	0.004	155.8	<.0001	Bern	-0.0444	0.009	-5.2	<.0001
4-Zimmerwohnung	0.9002	0.005	199.8	<.0001	Lausanne	-0.1283	0.010	-13.2	<.0001
5-Zimmerwohnung	1.0846	0.005	212.0	<.0001	Genf	-0.0805	0.008	-10.0	<.0001
6-Zimmerwohnung	1.2122	0.007	169.7	<.0001	Kanton ZH ohne Stadt	-0.0480	0.004	-11.5	<.0001
Verweildauer (in Jahren)	-0.0077	0.000	-117.1	<.0001	Kanton BE ohne Stadt	-0.1131	0.008	-13.9	<.0001
Stockwerk	0.0045	0.000	11.6	<.0001	Kanton VD ohne Stadt	-0.1222	0.008	-14.4	<.0001
Garage inbegriffen	0.0657	0.003	21.8	<.0001	Kanton GE ohne Stadt	-0.0986	0.009	-11.4	<.0001
Attikawohnung	0.1659	0.005	35.7	<.0001	Kanton LU	-0.1111	0.005	-21.2	<.0001
Duplexwohnung	0.0670	0.003	20.5	<.0001	Kanton UR	-0.1888	0.017	-10.8	<.0001
Wohnbaugenossenschaft, Bund, Kanton, Gemeinde	-0.1667	0.002	-76.7	<.0001	Kanton SZ	-0.1553	0.011	-14.4	<.0001
Mikrovariablen					Kanton OW	-0.1362	0.015	-9.0	<.0001
Entfernung zu Seeufer < 200 m	0.0057	0.003	1.9	0.0628	Kanton NW	-0.1009	0.012	-8.4	<.0001
Distanz zum See (in km)	-0.0025	0.000	-5.6	<.0001	Kanton GL	-0.1491	0.019	-8.0	<.0001
Seesicht:					Kanton ZG	-0.1146	0.009	-12.8	<.0001
Kleine, 7 bis 100 ha (D)	0.0171	0.003	5.4	<.0001	Kanton FR	-0.1374	0.009	-14.9	<.0001
Mittlere, 101 bis 300 ha (D)	0.0474	0.004	13.4	<.0001	Kanton SO	-0.1505	0.009	-17.3	<.0001
Gute, > 300 ha (D)	0.0670	0.003	20.5	<.0001	Kanton BS	-0.0277	0.008	-3.6	3E-04
Bergsicht:					Kanton BL	-0.0610	0.007	-8.6	<.0001
Mittlere, 3 bis 12 Gipfel (D)	0.0043	0.002	2.0	0.044	Kanton SH	-0.1567	0.009	-16.8	<.0001
Gute, >12 Gipfel (D)	0.0161	0.003	6.1	<.0001	Kanton AR	-0.1857	0.012	-15.9	<.0001
Neigung:					Kanton AI	-0.1429	0.038	-3.7	2E-04
4% - 9% (D)	0.0150	0.002	7.0	<.0001	Kanton SG	-0.1355	0.006	-22.7	<.0001
> 9% (D)	0.0341	0.003	12.7	<.0001	Kanton GR	-0.0635	0.008	-8.4	<.0001
Exposition:					Kanton AG	-0.1034	0.005	-19.1	<.0001
SuedWesten (D)	0.0090	0.002	4.9	<.0001	Kanton TG	-0.1874	0.007	-28.5	<.0001
NordWesten, nur Kanton GE (D)	0.0182	0.008	2.2	0.0282	Kanton TI	-0.1461	0.006	-24.9	<.0001
Lärm:					Kanton VS	-0.2723	0.010	-26.5	<.0001
Strassenlärm Nacht (über 45dB)	-0.0020	0.000	-5.8	<.0001	Kanton NE	-0.3008	0.012	-25.1	<.0001
Bahnlärm Tag (über 50dB)	-0.0041	0.001	-4.9	<.0001	Kanton JU	-0.2388	0.018	-13.5	<.0001
					Konstante	2.6260	0.080	32.7	<.0001
Beobachtungen	91'239				D=Dummy Variable				
Korrigiertes R ²	0.663				ln = natürliches Logarithm				
Standardfehler der Schätzung (RMSE)	0.227								

Die Koeffizienten im doppeltlogarithmischen Modell können als prozentuale Veränderung des Mietpreises bei einer prozentualen Änderung eines bestimmten Merkmals interpretiert werden: Der Koeffizient für die Zimmerwohnfläche beträgt 0.373, d.h. eine Erhöhung der Zimmerwohnfläche um 1% erhöht die Nettomiete um 0.37%.

Die wichtigsten Ergebnisse aus der Modellschätzung können in wenigen Punkten zusammengefasst werden. Der Nettomietpreis einer Wohnung steigt mit:

- zunehmender Zimmeranzahl
- zunehmender Zimmerwohnfläche
- dem Vorhandensein von Qualitätsfaktoren wie Attikawohnung
- abnehmendem Gebäudealter (m.a.W. je jünger desto teurer)
- zunehmend zentraler Lage
- guter See- und Bergsicht
- Hangneigung und Süd(West)Lage
- zunehmender Steuerkraft

Umgekehrt fällt der Nettomietpreis einer Wohnung mit:

- zunehmendem Strassenlärm bei Nacht
- zunehmendem Bahnlärm bei Tag
- zunehmender Steuerbelastung

Eine weitere Variable berücksichtigt die Dauer des Mietverhältnisses. Für jedes Vertragsjahr reduziert sich der Mietzins um 0.8 Prozent. Dieser negative Zusammenhang zwischen Verweildauer und Nettomiete ist auf die Mietregulierung zurückzuführen. Diese Variable ist signifikant und wichtig für die Modellierung, fließt aber nicht in die Qualitätsbereinigung des MPI ein.

Schliesslich wurden die Auswirkungen von unterschiedlichen Eigentübertypen untersucht. Ist der Wohnungseigentümer eine Wohnbaugenossenschaft, der Bund, ein Kanton oder eine Gemeinde, dann kostet die Wohnung rund 17 Prozent weniger, als wenn der Eigentümer der Wohnung eine Privatperson, eine Versicherung oder eine Pensionskasse ist (vgl. auch Thalmann, 1987).

Im Modell werden kantons- und grosszentrenspezifische¹³ Konstanten (Fixed-Effects) gebildet, die systematische Unterschiede im Mietpreisniveau zwischen den Kantonen und den Grossstädten erfassen, welche nicht durch die übrigen Modellparameter erklärt werden können. Die Stadt Zürich bildet die Referenzgruppe für den Vergleich dieser Koeffizienten. Für Bern beträgt diese Differenz im Vergleich zu Zürich z.B. -4 Prozent und für Lausanne knapp -13 Prozent.

Wie kann die Einflussstärke der Merkmale verglichen werden? Dazu werden die Koeffizienten standardisiert und somit untereinander vergleichbar: Um wieviel Standardabweichungen steigt/sinkt (je nach Vorzeichen des Beta-Koeffizienten) die Nettomiete, wenn sich die unabhängige Variable (x) um eine Standardabweichung erhöht. Die folgende Tabelle 22 zeigt die standardisierten Koeffizienten der kontinuierlichen Struktur- und Lagevariablen. An der Interpretierbarkeit ändert sich durch die Standardisierung nichts. Wenn das Gebäudealter bspw. um eine Standardabweichung zunimmt, dann sinkt die Miete um 0.68 Standardabweichungen (*ceteris paribus*).

¹³ Als Grosszentren gelten Zürich, Bern, Basel, Lausanne und Genf.

Die Variable mit der grössten Auswirkung auf die Nettomiete ist das Gebäudealter, gefolgt von der Verweildauer und der Zimmergrösse. Generell lässt sich sagen, dass die wichtigsten Variablen im Modell die Strukturvariablen sind, gefolgt von den Variablen der Makrolage und der Mikrolage. Im Rahmen der Modellvergleiche (Abschnitt 6.3) werden diese Ergebnisse noch detailliert.

Tabelle 22: Standardisierte Koeffizienten, sortiert nach relativer Stärke

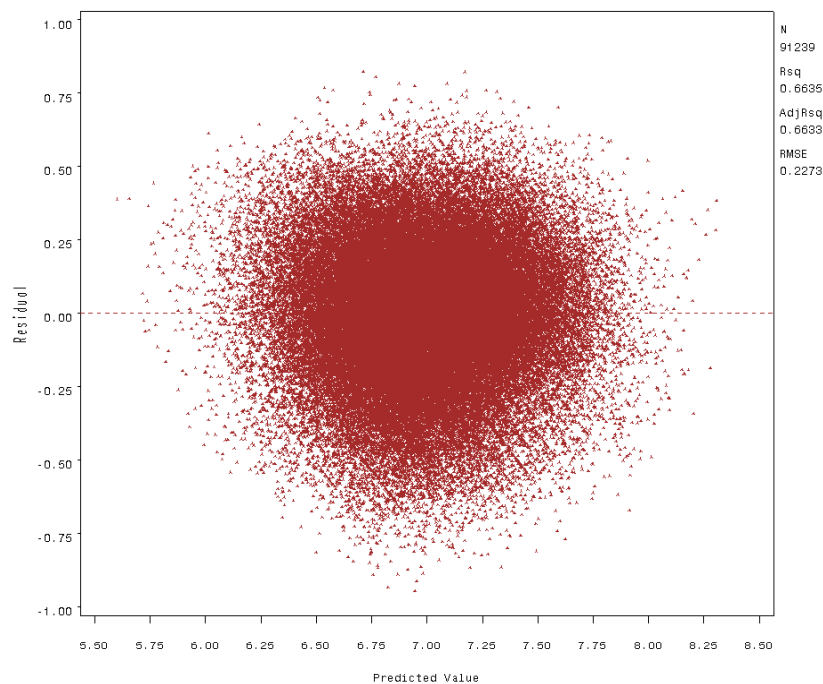
Nr.	Variable	Abs. St.Estimate
1	Gebäudealter (in Jahren)	-0.684
2	Gebäudealter im Quadrat	0.532
3	Verweildauer (in Jahren)	-0.239
4	Zimmergrösse (in m ² , log)	0.231
5	Erreichbarkeitspotenzial	0.158
6	ln Steuerkraft	0.144
7	ln Steuersatz	-0.062
8	Distanz zum See (in km)	-0.028
9	Stockwerk	0.024
10	Strassenlärm Nacht (über 45dB)	-0.011
11	Bahnlärm Tag (über 50dB)	-0.009

6.2 Ausreisserbehandlung und Residuenanalyse

Die Erkennung und korrekte Behandlung von „Ausreissern“ ist eine wichtige Aufgabe jeder empirischen Arbeit. Ausreisser in den Daten können zu verzerrten Ergebnissen oder Fehlspezifikationen führen. Je mehr Beobachtungen und Variablen kontrolliert werden müssen, desto schwieriger ist diese Aufgabe. Typischerweise werden als Ausreisser diejenigen Beobachtungen definiert, die sich stark von der Mehrheit der Beobachtungen in einer Stichprobe unterscheiden bzw. grosse Auswirkungen auf die Koeffizienten aufweisen. Um Ausreisser identifizieren und eliminieren zu können, benutzen wir die Cook's-Distanz¹⁴. Im vorliegenden Fall werden 4'640 Beobachtungen (ca. 5%) der Stichprobe ausgeschlossen. Eine vertiefte Analyse dieser Beobachtungen findet sich im Anhang (Abschnitt 7.4.2).

Im Rahmen der Residuenanalyse analysieren wir die Verteilung der Residuen anhand der jeweiligen Histogramme (Abbildung 33 Anhang) und vergleichen die Verteilung mit der (unterstellten) Normalverteilung. Es zeigen sich keine Auffälligkeiten, die auf Fehlspezifikationen des Modells hindeuten würden. Weiter analysieren wir den sogenannten Tukey-Anscombe Plot, ein Streudiagramm bei dem die Residuen gegen die "prognostizierten" Werte geplottet werden. Dabei sollten die Residuen unabhängig (zufällig) von den Prognosewerten streuen, d.h. wir sollten bspw. keine Trichterform sehen (grössere Residuen mit steigenden Preisen), was auf eine falsche oder unvollständige Modellspezifikation hinweisen würde. Auch hier stellen wir aufgrund der Analysen keine Auffälligkeiten fest (Abbildung 11).

¹⁴ Die Daumenregel für die Ausreisserelimination ist hier die Distanz grösser als $4/N$ oder 0.00004 (N = Stichprobenumfang).



6.3 Modellvergleiche

Welche Erklärungskraft haben Modelle, die nur mit den Strukturvariablen auskommen? Welche Modellverbesserung resultiert durch Hinzufügen der Lagevariablen? Wie ändert sich der Stichprobenumfang, wenn wir auf die (vom BfS) imputierten Nettomieten verzichten? Wie ändert sich das Modell, wenn wir anstelle der Nettomieten die Bruttomieten als Zielgrösse verwenden? Um diese und weitere Fragen beantworten zu können, haben wir verschiedene Modellspezifikationen getestet und verglichen.

Die Unterschiede bezüglich Samplegrösse und Modellfit (korrigiertes R^2) sind in Tabelle 23 dargestellt. Um die Einflüsse der drei verschiedenen Variablengruppen (Struktur-, Makro-, Mikrolagefaktoren) beurteilen zu können, haben wir das Modell stufenweise erweitert und für verschiedene Zielgrössen (Zeilen A bis G) berechnet. In Spalte A verfügt das Modell nur über die Strukturvariablen, in der Spalte B kommen die Variablen der Makrolage dazu und Spalte C entspricht dem Gesamt-Modell (Full-Model).

Tabelle 23: Modellvergleiche: Erklärungsgehalt (Korrigiertes R^2) und Samplegrösse

			X	Y	Z**
			Strukturvariablen	+Makrovariablen	+Mikrovariablen
A	Nettomieten	N	96'627	95'879	95'879
	Alle Beobachtungen	KorrR ²	0.428	0.526	0.539
B	Nettomieten; ohne imputierte Nettomieten (ON)	N	67'145	66'593	66'593
		KorrR ²	0.486	0.593	0.604
C	Nettomieten; ohne imputierten Eigentübertyp (OE)	N	75'680	75'074	75'074
		KorrR ²	0.437	0.550	0.556
D	ON + OE	N	54'241	53'786	53'786
		KorrR ²	0.491	0.604	0.616
E	Bruttomiete	N	92'385	91'674	91'674
		KorrR ²	0.476	0.573	0.586
F	Nettomieten; bereinigte Stichprobe*	N	-	-	91'239
		KorrR ²	-	-	0.663
G	Bruttomiete bereinigt*	N	-	-	86'551
		KorrR ²	-	-	0.693

*Cook's distance

N = Anzahl Beobachtungen, KorrR² = Gesamterklärungsgehalt des Modells

Modellbenennung, z.B. Modell EZ = Modell mit Bruttomieten als zu erklärende Variable und alle Modellvariablen (Struktur-, Makro- und Mikrovariablen), EZ = Zeile E, Spalte Z

ON = Sample ohne imputierten Nettomieten. Diese Beobachtungen fliessen nicht in die Schätzung.

OE = Sample ohne imputierten Eigentübertyp. Diese Beobachtungen fliessen nicht in die Schätzung.

**Modell nur mit Struktur- und Mikrovariablen: N = 96'627, KorrR² = 0.445

Die wichtigsten Ergebnisse aus dem Modellvergleich:

- Die Integration von Makrovariablen führt zu einer deutlichen Modellverbesserung. Das korrigierte R^2 steigt um mehr als 10 Prozent.
- Die Mikrovariablen sind hoch signifikant, führen aber nur zu einer marginalen Verbesserung des Gesamterklärungsgehaltes. Das korrigierte R^2 erhöht sich leicht um 0.6 bis 1.3 Prozent.
- Der Verzicht auf imputierte Variablen wie die Nettomiete oder den Eigentübertyp erhöht den Erklärungsgehalt. Gleichzeitig führt dies allerdings zu einem deutlich reduzierten Sampleumfang. Der Verzicht auf die imputierten Nettomieten bedeutet eine Reduktion des Sample um 30 Prozent (29'286 Beobachtungen).
- Das Modell mit der Bruttomiete als erklärende Variable erhöht den Erklärungsgehalt des Modells um 3 Prozent, führt aber zu einer Samplereduktion um 5 Prozent.

Wie stabil sind die Koeffizienten in den verschiedenen Modellen? In der folgenden Tabelle 24 werden die Koeffizienten der Strukturvariablen für vier verschiedene Modelle verglichen.

Tabelle 24: Koeffizientenvergleich der Strukturvariablen aus verschiedenen Modellen

Variable	Model AX**	Model BY	Model CZ	Model FZ
2-Zimmerwohnung	0.390	0.404	0.412	0.411
3-Zimmerwohnung	0.649	0.679	0.691	0.683
4-Zimmerwohnung	0.849	0.895	0.912	0.900
5-Zimmerwohnung	1.034	1.081	1.091	1.085
6-Zimmerwohnung	1.152	1.230	1.239	1.212
Zimmergrösse (m2, in log)	0.367	0.371	0.376	0.373
Verweildauer (in Jahren)	-0.0072	-0.0076	-0.0078	-0.0077
Stockwerk	0.008	0.006	0.003	0.005
Gebäudealter	-0.008	-0.010	-0.010	-0.010
Gebäudealter im Quadrat	0.000059	0.000074	0.000073	0.000075
Garage inbegriffen	0.00752*	0.054	0.061	0.066
Attikawohnung	0.1792	0.175	0.174	0.165
Duplexwohnung	0.0421	0.054	0.065	0.067
Wohnbaugenossenschaft, Bund, Kanton, Gemeinde	-0.1315	-0.171	-0.204	-0.167

Signifikanzniveau aller Variablen 1%

*Signifikanzniveau 10%

** Modellbezeichnung gem. Tabelle 23: AZ = Zeile A, Spalte Z

Die wichtigsten Ergebnisse aus dem Koeffizientenvergleich der verschiedenen Modelle sind:

- Die Koeffizienten der Strukturvariablen sind stabil.
- Die Signifikanz der Variablen und deren Vorzeichen bleiben unverändert.
- Dies trifft auch auf die Koeffizienten der Mikro- und Makrovariablen zu (Tabelle 27, Anhang).

FAZIT: Modellvergleiche

Im MPI werden Nettomieten erhoben und deren Preisveränderung gemessen. Daher ist es sinnvoll, das Modell zur Qualitätsbereinigung ebenfalls auf Nettomieten zu modellieren. Der Verzicht auf imputierte Nettomieten führt zu einer leicht besseren Modellanpassung, reduziert aber den Stichprobenumfang um 30 Prozent. Wie vorstehende Analysen und Vergleiche zeigen, bietet die Variante FC (vgl. Tabelle 23) ein Optimum. Wir modellieren somit auf die Nettomieten als Zielgrösse unter Verwendung imputierter Nettomieten (BfS) und Eigentübertypen (ZKB).

6.4 In- und Out-of-Sample Analyse

Die Qualität des hedonischen Modells haben wir mit In-Sample und Out-of-Sample¹⁵ Analysen getestet. Die Differenz zwischen effektiven Nettomieten und den geschätzten Werten ist in beiden Fällen gering, d.h. das Modell verfügt über eine gute Prognosegüte. Beim In-Sample Test lagen rund 75 Prozent der Beobachtungen innerhalb des Bereichs von +/- 25 Prozent. Bei der Out-of-Sample Analyse war dieser Anteil mit 76 Prozent noch leicht grösser.

Die folgende Tabelle 25 zeigt die Schätzqualität der In-Sample und Out-of-Sample Analysen nach Kanton. Verglichen werden die jeweiligen Stichprobenanteile, die in einem Intervall von ± 20 Prozent, ± 25 Prozent und ± 30 Prozent Abweichung des Prognosewerts von der effektiven Nettomiete liegen.

Die Ergebnisse auf kantonaler Ebene variieren erheblich. So liegen im Kanton Appenzell Innerrhoden 100 Prozent der Beobachtungen im Intervall von ± 20 (36 Beobachtungen), im Kanton Genf sind es unter 60 Prozent. Zusammenfassend lässt sich festhalten, dass die Testresultate für das hedonische Modell insgesamt als gut zu bewerten sind.

Tabelle 25: In-Sample und Out-of-Sample Auswertungen

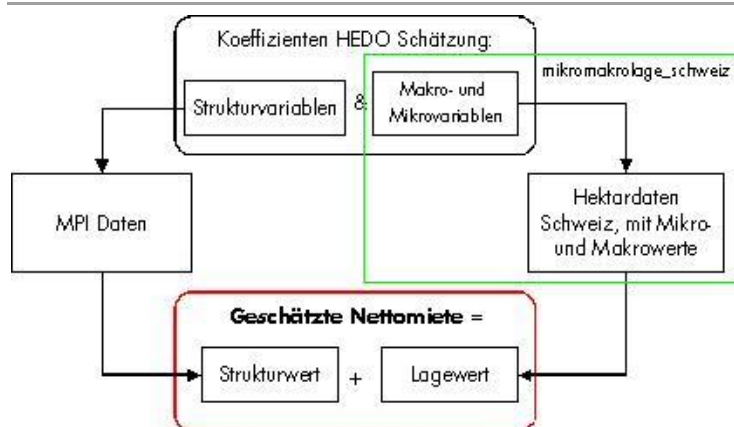
Kanton/ Stadt	In-Sample				Out-of-Sample			
	Anzahl	in +/- 20%	in +/- 25%	in +/- 30%	Anzahl	in +/- 20%	in +/- 25%	in +/- 30%
Kanton ZH	19'037	65.7	75.3	82.5	3'817	61.6	71.4	79.3
Kanton BE	9'157	65	75.4	82.8	1'796	61.0	72.1	80.5
Kanton LU	4'927	65.5	75.3	83.4	1'039	65.6	76.2	84.6
Kanton UR	190	85.3	94.7	100	45	88.9	97.8	100
Kanton SZ	1'248	69.7	80.7	87.8	247	66.4	76.9	87
Kanton OW	249	87.6	93.6	97.6	45	88.9	95.6	97.8
Kanton NW	450	80.9	90.7	94.2	106	75.5	87.7	93.4
Kanton GL	160	90	98.1	100	18	94.4	100	100
Kanton ZG	1'778	64.8	76.7	84.0	375	68.5	78.7	84
Kanton FR	1'429	68.6	78.7	86.5	274	67.2	75.9	86.9
Kanton SO	1'471	69.3	79.5	87.1	294	72.8	79.9	89.5
Kanton BS	6'099	71.5	86.8	92.1	1'208	65.1	75.2	82.9
Kanton BL	4'003	67.8	77.8	84.7	856	68.1	78	86.3
Kanton SH	800	70.4	81.1	90	168	69	81.5	86.3
Kanton AR	437	75.5	88.1	92.7	96	76	88.5	93.8
Kanton AI	36	100	100	100	5	100	100	100
Kanton SG	5'264	66	75.9	83.5	1'062	66.6	76.1	81.8
Kanton GR	1'290	68	79.1	86.4	259	63.7	77.6	85.3
Kanton AG	4'998	67.8	78.4	85	1'023	67.6	77.7	84
Kanton TG	2'246	68.5	78.9	85.7	436	69	80	85.6
Kanton TI	2'479	66.8	76.3	84.1	493	70.8	78.9	86
Kanton VD	9'434	66.3	76.6	83.8	1'823	65.1	76.3	82.5
Kanton VS	841	74.4	85.6	92	149	75.2	85.9	91.3
Kanton NE	1'996	61.9	73.8	82.5	378	60.1	72.8	80.4
Kanton GE	11'011	59.5	70.8	79	2'202	57.2	68.2	76.5
Kanton JU	213	87.3	95.3	99.5	35	85.7	94.3	100
Schweiz	91'243	64.2	74.6	82.2	18'249	64.2	74.6	82.1

¹⁵ Für die Out-of-Sample Analyse haben wir die Endstichprobe (N=91'243) proportional geteilt. Wir haben ein random sample von 20% selektiert. Die grössere (80%, N=72'994) haben wir für das Refitting des Modells benutzt und die kleinere (20%, N=18'249) Stichprobe für die Out-of-Sample Auswertungen.

6.5 Modellanwendung für Qualitätsbereinigung im Mietpreisindex

In diesem Abschnitt wird auf die Modellanwendung für die Qualitätsbereinigung im Mietpreisindex (MPI) eingegangen. Zunächst wird der Prozess zur Berechnung qualitätsbereinigter Nettomieten mittels der hedonischen Methode beschrieben. Die nachfolgende Abbildung 12 veranschaulicht den Datenprozess schematisch. Der eigentliche Berechnungsprozess setzt sich aus zwei Schritten zusammen, wobei in einem ersten Schritt (linker Pfad in der Abbildung) eine lageunabhängige "Strukturmiete" geschätzt wird. Diese berechnet sich aufgrund der Strukturkoeffizienten des hedonischen Modells und den spezifischen Wohnungs- und Gebäudemerkmalen (Zimmerzahl, Fläche, Alter usw.). In einem zweiten Schritt wird entsprechend der Geo-Koordinate jeder Wohnung noch der jeweilige Hektar-Lagewert (Abbildung 12) zugeordnet. Dieser Hektarlagewert enthält die Koeffizienten aller verwendeten Variablen der Makro- bzw. Mikrolage. Die Summe (Strukturmiete und Lagewert) ergibt die geschätzte lagespezifische Nettomiete (logarithmiert) bzw. der Exponent davon die geschätzte Nettomiete.

Abbildung 12: Schematische Darstellung der Berechnung



Die Basis für die Berechnung stellen die geschätzten Koeffizienten und die spezifischen Merkmale der Wohnungen dar. Diese können in zwei Gruppen unterteilt werden: Strukturgruppe und Lagegruppe (oder hektarspezifische Gruppe). In die Strukturgruppe gehören die Strukturkoeffizienten und die entsprechenden Wohnungseigenschaften. Für die Berechnung des Strukturwerts werden die benötigten Wohnungseigenschaften mit den entsprechenden Koeffizienten multipliziert und summiert.¹⁶

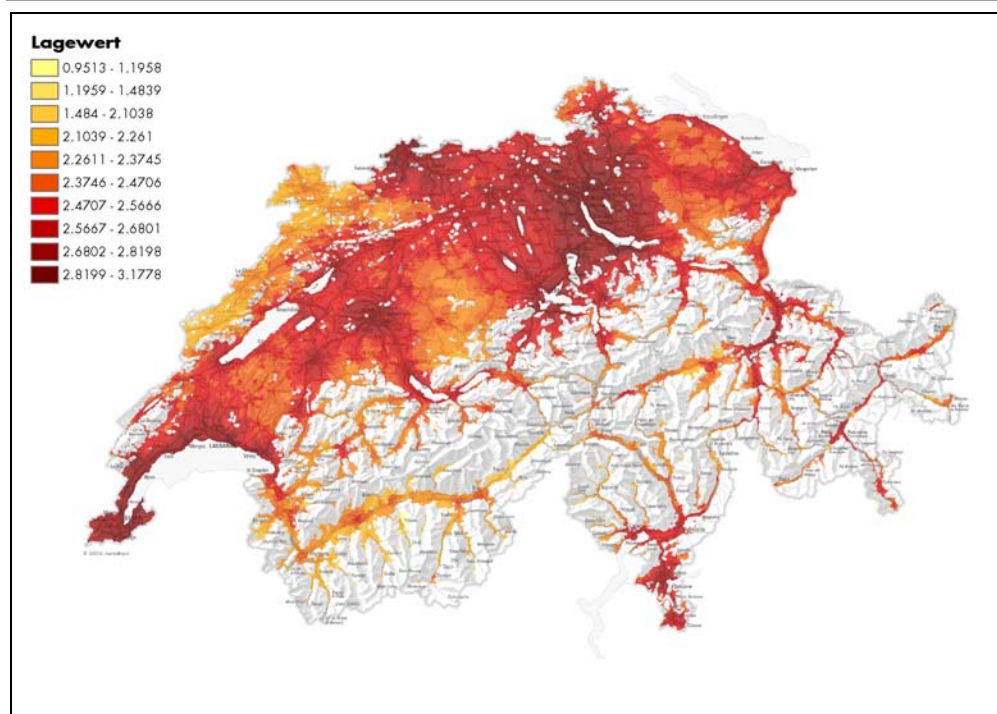
Die Lagewerte setzen sich aus Mikro- (z.B. Strassenlärm) und Makrolage (z.B. Steuerkraft) zusammen und werden nach derselben Methodik berechnet. Diese Lagewerte sind unabhängig von der spezifischen Wohnung und identisch für alle Wohnungen des gleichen Hektars. Die folgende Abbildung 13 zeigt die Lagewerte für alle bewohnbaren Hektare in der Schweiz.¹⁷ Je höher der Lagewert (dunkles Rot), desto bessere die Lage. Wie in der Abbildung er-

¹⁶ Eine Beispiel-Berechnung befindet sich im Dokument `manual_mpi_scoring.doc`, die separat geliefert wurde (siehe Anhang Tabelle 29).

¹⁷ Die Lagewerte werden separat als eine SAS Tabelle geliefert (siehe Anhang Tabelle 29).

sichtlich, dominiert die Makrolage die Mikrolage. Die Zentralität eines Wohnortes – gemessen anhand des Erreichbarkeitspotenzials von Arbeitsplätzen – hat einen entscheidenden Einfluss auf die Mietpreise (vgl. Abbildung 30). Jeder Wohnung kann durch ihre Adresse (Geo-Koordinaten) der entsprechende hektarspezifische Lagewert zugeordnet werden. Die qualitätsbereinigte Nettomiete ist die Summe von Struktur- und Lagewert.

Abbildung 13: Hektar – Lagewerte für die Schweiz



6.6 Fazit und Empfehlungen

Die gewählte, auch von Eurostat empfohlene Methode des "hedonic pricing" ist gut geeignet, um eine Qualitätsbereinigung der MPI-Berechnung durchzuführen. Die Umsetzung ist auf die Rahmenbedingungen des MPI abgestimmt und lässt sich in die quartalsweise Produktion integrieren. Die statistische Qualität des hedonischen Mietpreismodells kann als gut bezeichnet werden. Für die Modellanwendung auf die MPI-Daten wären vertiefte Analysen und grössere Fallzahlen in den einzelnen Zellen – besonders bei den 1- bzw. 6-Zimmerwohnungen jüngerer Baujahrs – notwendig. Die im Rahmen der MPI-Revision 2010 erfolgten zusätzlichen Massnahmen im Zusammenhang mit der Stichprobenaufstockung sind deshalb zu begrüssen.

Eine Modellkalibrierung bzw. Neuschätzung drängt sich kurzfristig nicht auf, zumal dafür aktuellere Mietstrukturdaten als die MSE 2003 notwendig wären. Alternativ ist eine Modellierung auf Basis gepoolter MPI-Daten zu überlegen. Die topografischen Lagefaktoren sind invariant über die Zeit. Aktualisierungen wären bei den Steuerdaten und den Erreichbarkeitspotenzialen (Betriebszählungsdaten) möglich, sofern die entsprechenden Daten vorliegen.

Literatur

- Andersson, H. e.a. (2009): *Property prices and exposure to multiple noise sources – hedonic regression with road and railway noise*; Environmental and Resource Economics; 45(1); 73-89.
- Banfi, S., Filippini, M. und Horehajova, A. (2008), *Valuation of Environmental Goods in Profit and Non-Profit Housing Sectors: Evidence from the Rental Market in the City of Zürich*, Swiss Journal of Economics and Statistics 144(4): 631-654.
- Baranzini, A. und Ramirez, J. V. (2005), *Paying for Quietness: The Impact of Noise on Geneva Rents*, Urban Studies 42(2):633-646.
- Baranzini, A., Schaerer, C. Ramirez, J. V. and Thalmann, P. (2006), *Feel it or Measure it: Perceived vs. Measured Noise in Hedonic models*, Geneva School of Business Administration, www.ssrn.com/abstract=937259.
- Baranzini, A. und Schaerer, C. (2007), *A Sight for Sore Eyes: Assessing the Value of View and Landscape Use on the Housing Market*, Geneva School of Business Administration, www.ssrn.com/abstract=981189.
- Baranzini, A., Schaerer, C. Ramirez, J. V. and Thalmann, P. (2008), *Do Foreigners Pay Higher Rents for the Same Quality of Housing in Geneva and Zurich?*, Swiss Journal of Economics and Statistics 144(4): 703-730.
- Bundesamt für Statistik [2009], *Landesindex der Konsumentenpreise : Revision 2010. Bericht zum Detailkonzept*, Version 4.0, Neuchâtel.
- Court, A.T. (1939): *Hedonic Price Indices – With Automotive Examples*, in The Dynamics of Automobile Demand; New York; General Motors Corporation; 99-117.
- Colwell, P.F. (1999): *Who was first – An examination of an early hedonic study*; Land Economics; 75(4); 620-626.
- Credit Suisse (2006), *Swiss Issue: Real Estate – Rents in Switzerland*.
- Djurdjevic, D., Eugster, Ch. And Haase, R. (2008), *Estimation of Hedonic Models Using a Multilevel Approach: An Application for the Swiss Rental Market*, Swiss Journal of Economics and Statistics 144(4): 679-702.
- De Haan, J. (2007), *Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches*, Statistics Netherlands, Division of Macro-Economic Statistics and Dissemination, September.
- Iten, R., Maibach, M. (1992), *Externe Kosten durch Verkehrslärm in Stadt und Agglomeration Zürich*, Swiss Journal of Economics and Statistics 128(1): 51-68.

- Greene, W. H. (2008), *Econometric Analysis*, 6th Edition, Pearson Prentice Hall.
- Haas, G.C. (1922): *Sales Prices as a Basis for Farm Land Appraisal*; Technical Bulletin 9; St. Paul; The University of Minnesota Agricultural Experiment Station; 57-77.
- Nelson, P. (2008): *Hedonic Property Value Studies of Transportation Noise – Aircraft and Road Traffic*; in Hedonic Methods in Housing Market Economics; Springer; 57-77.
- Rosen, S. (1974): *Hedonic Prices and Implicit Markets – Product Differentiation in Pure Competition*; Journal of Political Economy; 82; 34-55.
- Salvi, M. (2008), *Spatial Estimation of the Impact of Airport Noise on Residential Housing Prices*, Swiss Journal of Economics and Statistics 144(4): 577-606.
- Sirmans, S. G., Macpherson, D. A. und Zietz, E. N. (2005), The Composition of Hedonic Pricing Models, Journal of Real Estate Literature 13(1): 3-43.
- Soguel, N. (1994), *Evaluation monétaire des atteintes à l'environnement: une étude hédoniste et contingente sur l'impact des transports*, Neuchâtel, EDES.
- Sommer, H., Suter, S., Lenz, C. and Salvi, M. (2000), *Externe Lärmkosten des Verkehrs – Hedonic Pricing Analyse*, Bern, ECOPLAN.
- Statistik Stadt Zürich (2009), *Was Wohnqualität in Zürich kostet*, Juli.
- Thalmann, P. (1987), *Explication empirique des loyers lausannois*, Swiss Journal of Economics and Statistics (1): 47-70.
- Triplett, J. (2004), *Handbook on Hedonic indexes and Quality Adjustments in Prices Indexes*, OECD Science, Technology and Industry Working Papers.
- Waugh, F.V. (1928): *Quality Factors Influencing Vegetable Prices*; Journal of Farm Economics 10; 185-196.
- Yang, C. (2002): *Multiple Imputation for Missing Data: Concepts and New Development (Version 9.0)*, SAS Institute Inc., Rockville, MD
- Zürcher Kantonalbank (2004), *Preise, Mieten und Renditen – Der Immobilienmarkt transparent gemacht*.

7 Anhang

7.1 Verteilung der MSE-Beobachtungen pro Postleitzahlgebiet

Abbildung 14: Anzahl geokodierter Beobachtungen der MSE pro Postleitzahlgebiet

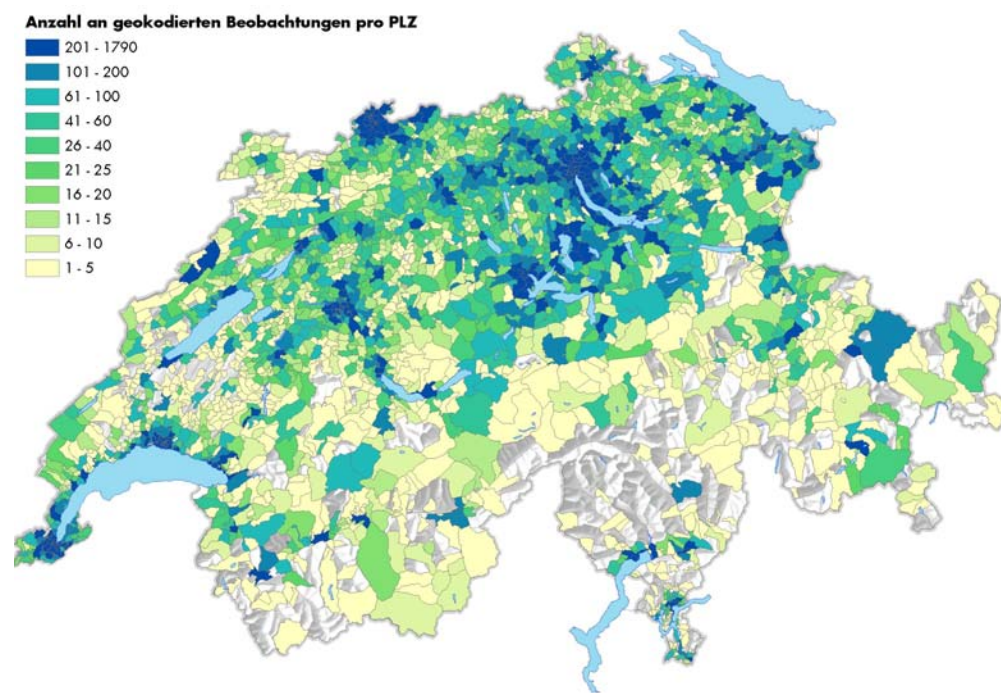
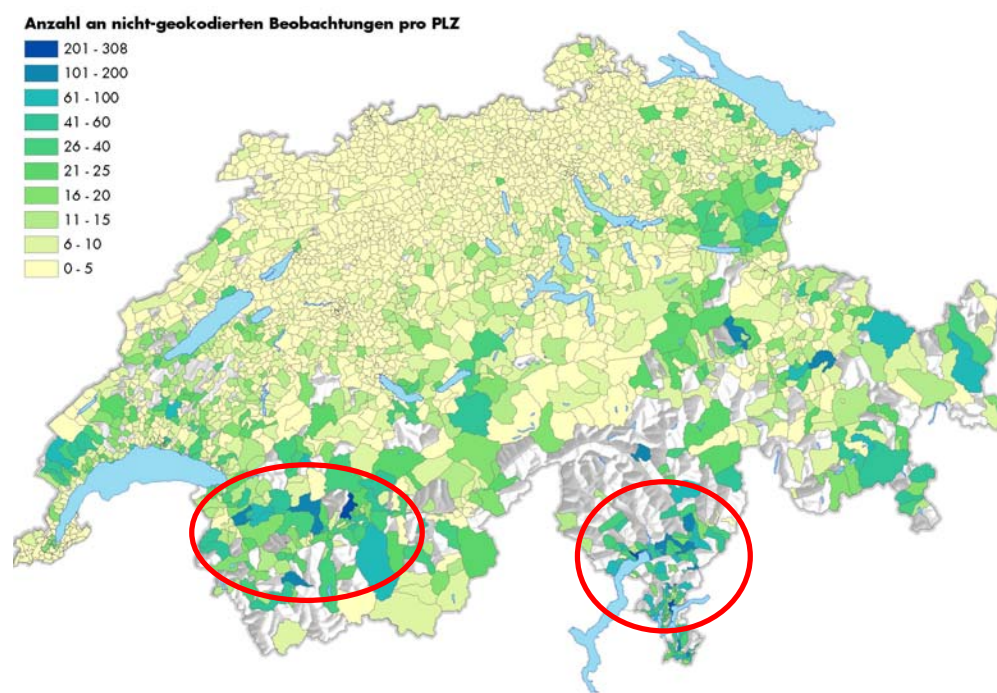
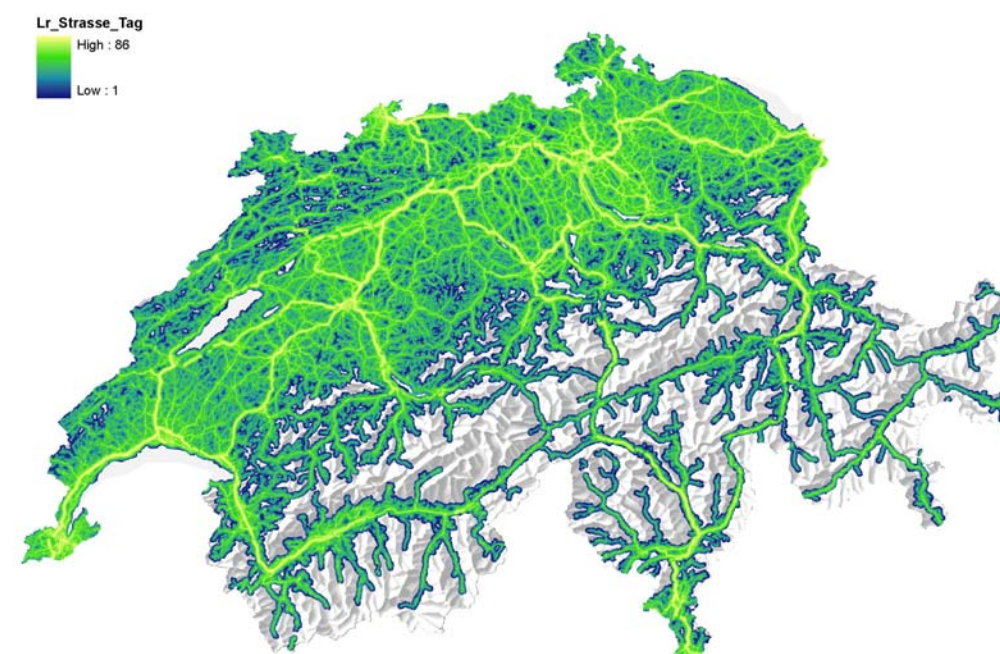


Abbildung 15: Anzahl nicht-geokodierter Beobachtungen der MSE pro Postleitzahlgebiet



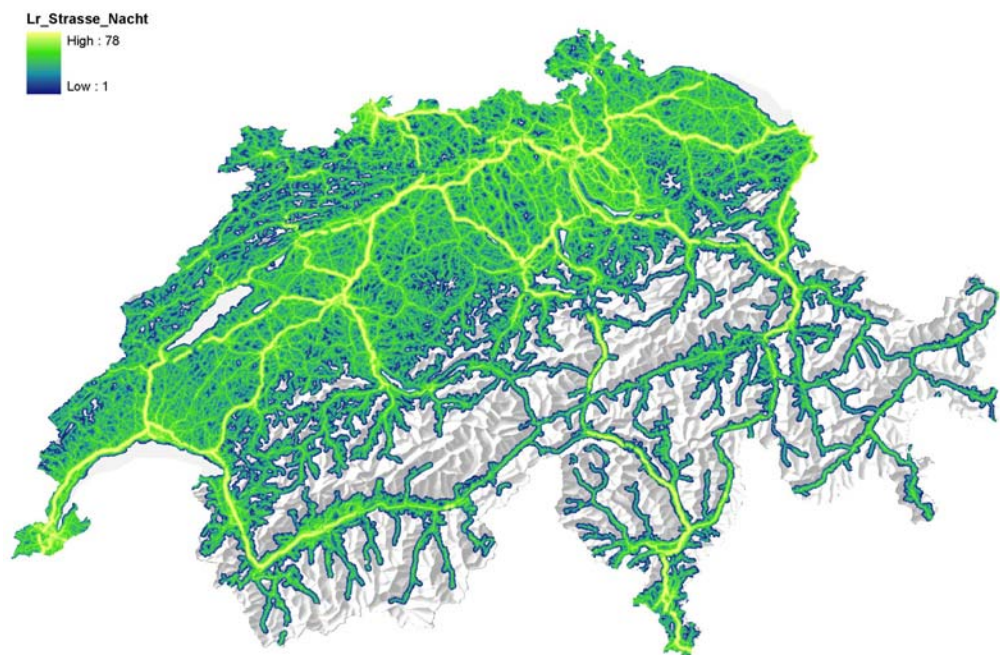
7.2 Geodaten

Abbildung 16: Strassenlärm Tag



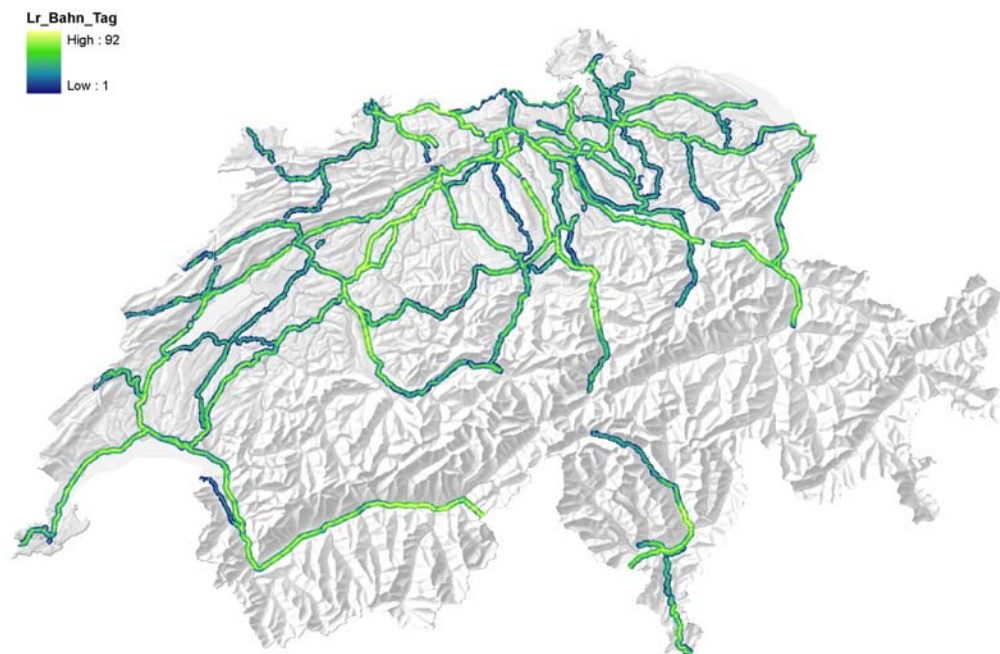
Datenquelle: BAFU

Abbildung 17: Strassenlärm Nacht



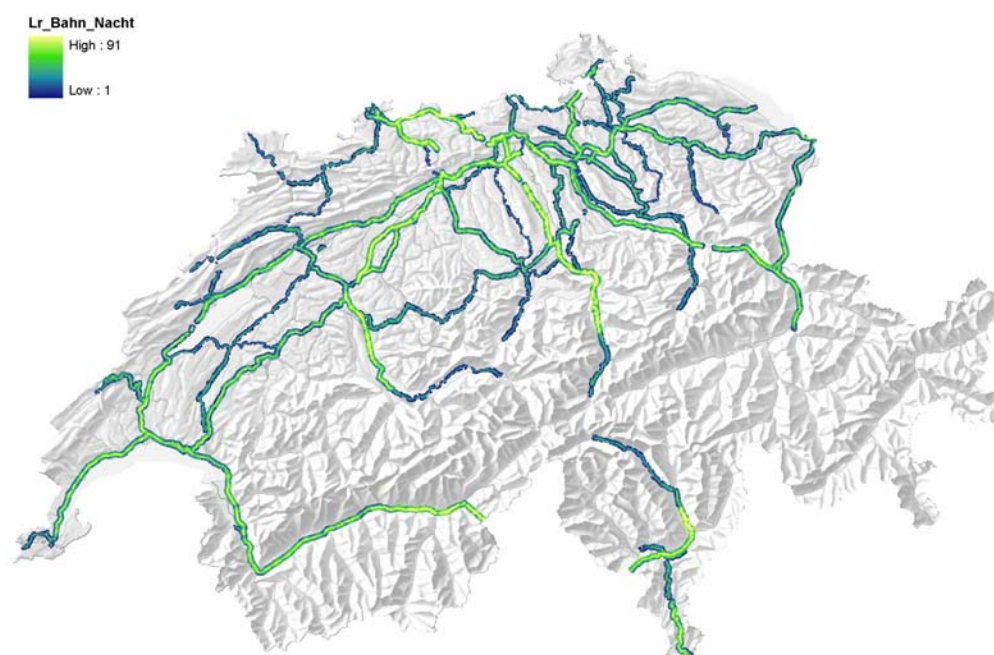
Datenquelle: BAFU

Abbildung 18: Bahnlärm Tag



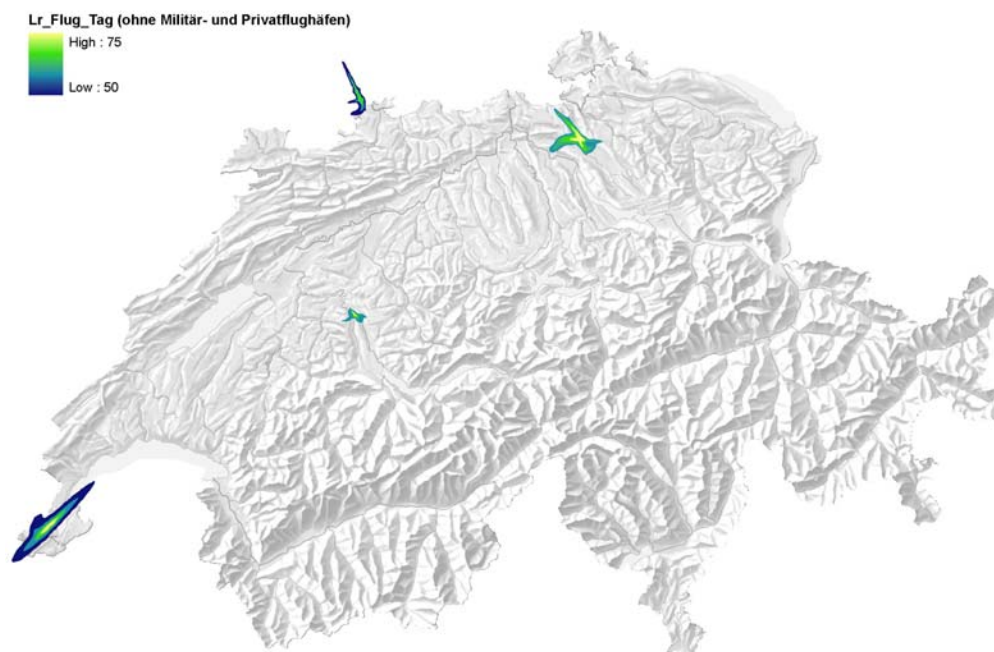
Datenquelle: BAFU

Abbildung 19: Bahnlärm Nacht



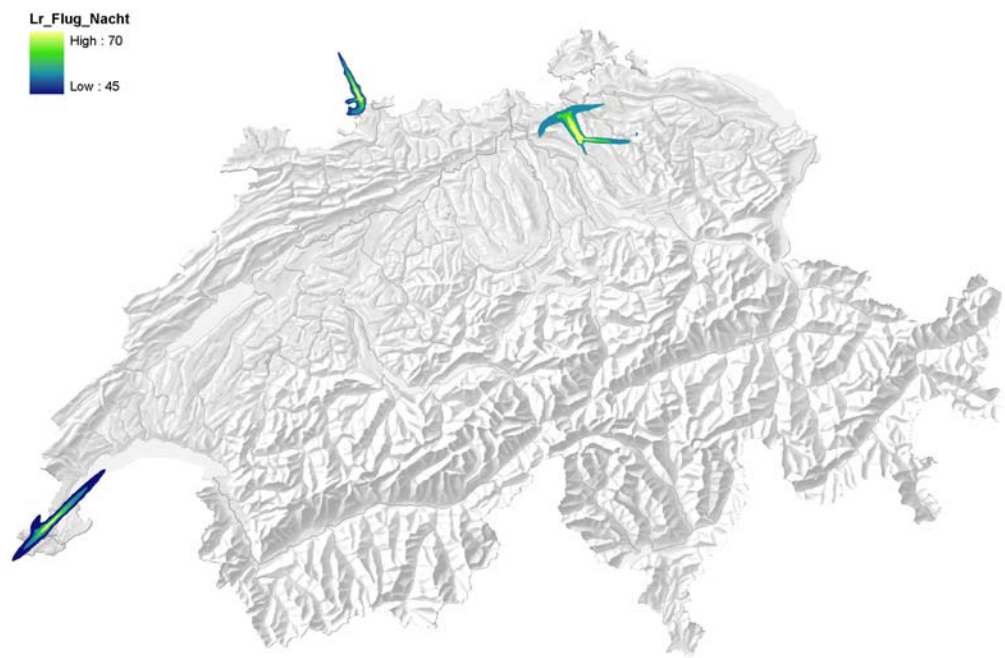
Datenquelle: BAFU

Abbildung 20: Fluglärm Tag (ohne Militär- und Privatflughäfen)



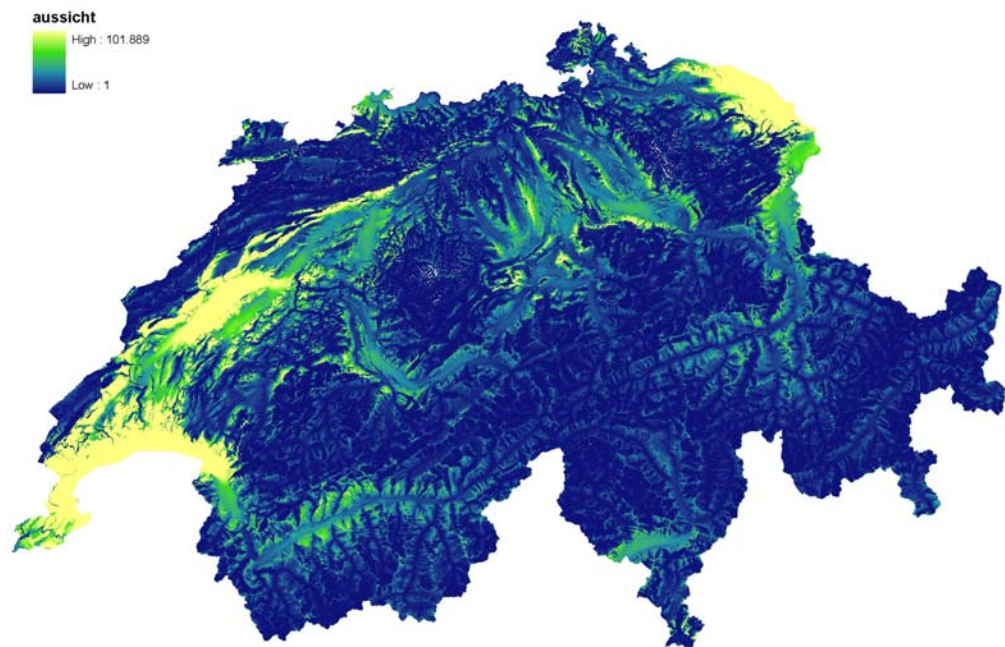
Datenquelle: BAFU

Abbildung 21: Fluglärm Nacht (alle Flughäfen)



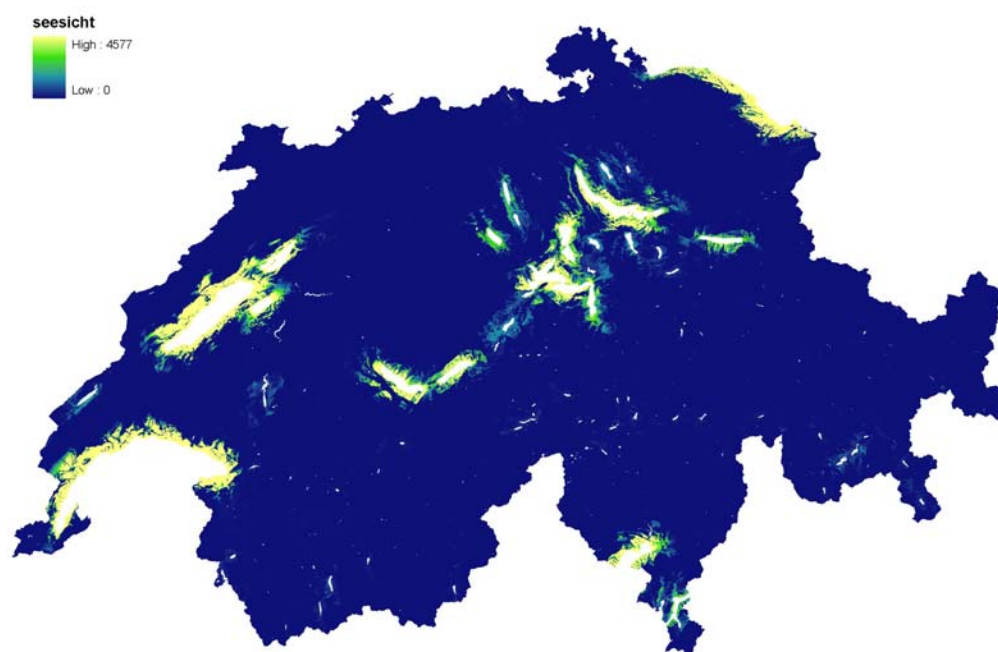
Datenquelle: BAFU

Abbildung 22: Aussicht



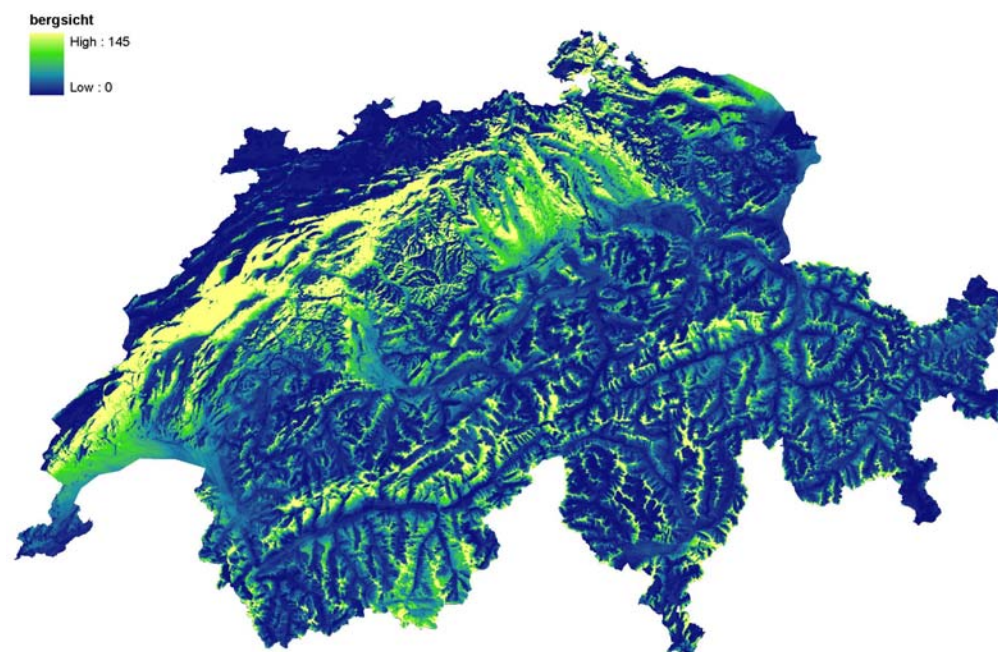
Datenquelle: EBP

Abbildung 23: Seesicht



Datenquelle: EBP

Abbildung 24: Bergsicht unter Einbeziehung von 217 Gipfeln



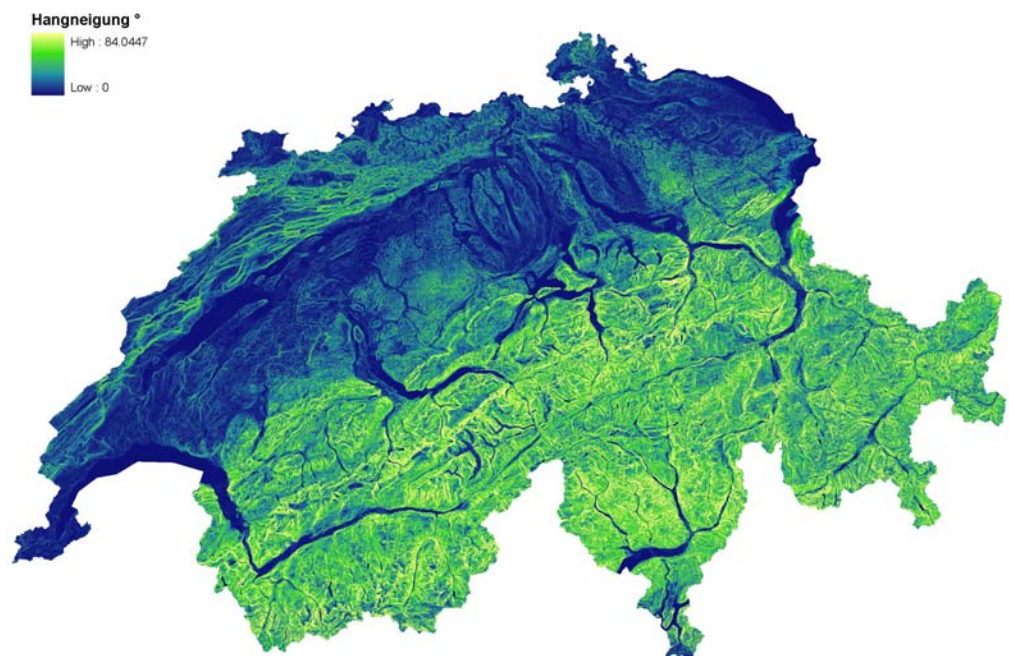
Datenquelle: EBP

Abbildung 25: Hangneigung in Prozent



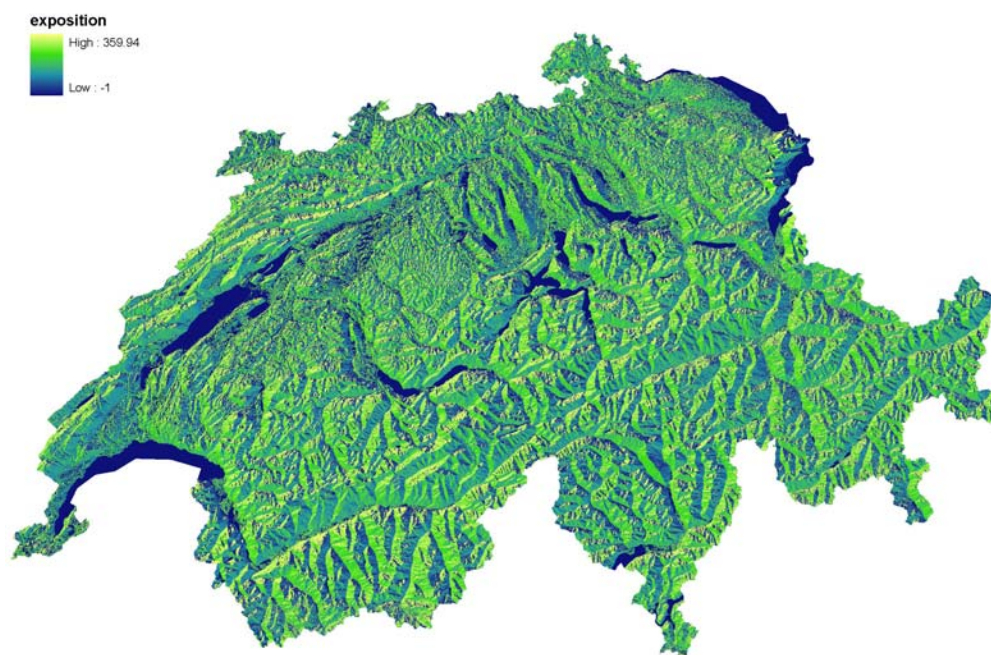
Datenquelle: EBP

Abbildung 26: Hangneigung in Grad



Datenquelle: EBP

Abbildung 27: Exposition



Datenquelle: EBP

Abbildung 28: Distanz zum See in km



Abbildung 29: Beispiel für die Gewichtung der Anzahl Arbeitsplätze

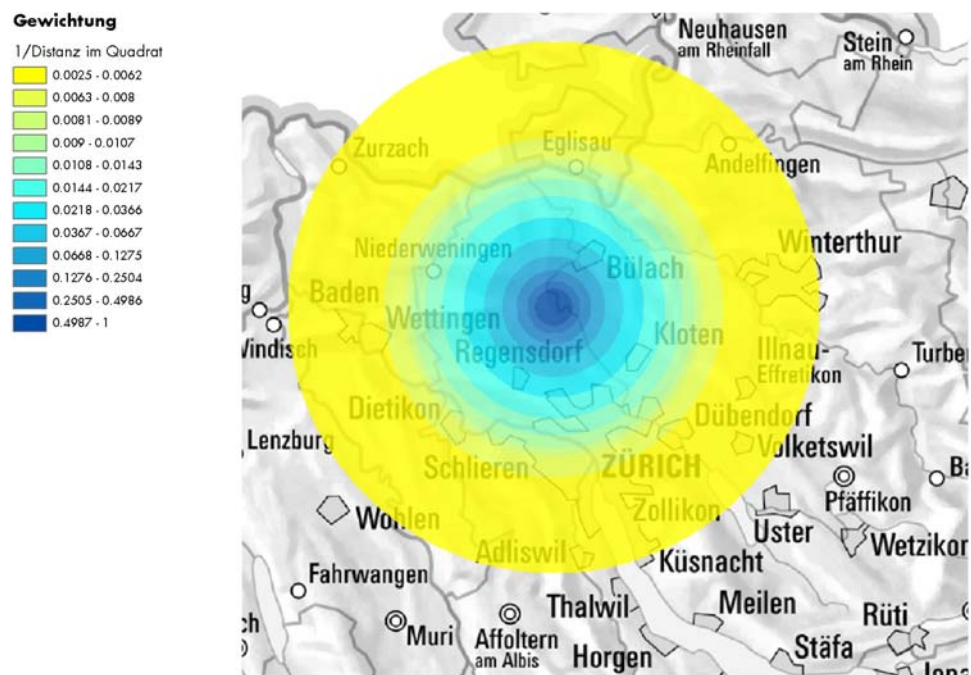
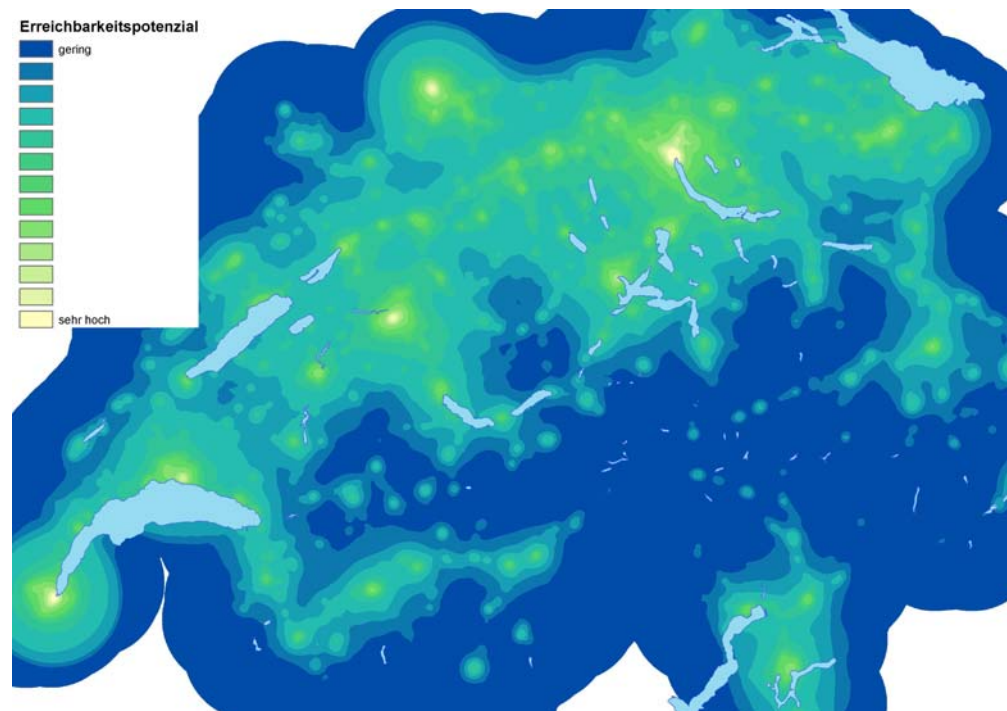


Abbildung 30: Erreichbarkeitspotenzial



The map illustrates the geographical spread of 'Zürcherdeutsch' in the Zurich region. The highest concentration of the dialect, represented by dark blue, is centered on the city of Zurich and its immediate surroundings, including areas like Regensdorf, Kloten, and the northern part of the region. The frequency decreases as one moves away from this central core, with yellow areas indicating lower frequency in the southern and western parts of the map. The map also shows various other towns and cities, such as Winterthur, Baden, and Wetztingen, providing a clear visual representation of the dialect's regional distribution.

Abbildung 32: Unterschiede in der Verteilung der Nettomieten für imputierte und nicht-imputierte Werte



7.4 Erweiterte Modellergebnisse

7.4.1 Schätzkoeffizienten

Tabelle 26: Modellergebnisse erweitert um Variance Inflation Factors (VIF) und 95%-Konfidenzintervalle

Objekteigenschaften	Schätzwert	Std. Abw.	VIF	95%L	95%U
Alter	-0.010160	0.000	17.5	-0.010400	-0.009930
Alter im Quadrat	0.000075	0.000	17.1	0.000073	0.000078
ln (Wohnfläche/Anzahl Zimmer)	0.3729	0.003	1.2	0.366	0.380
2-Zimmerwohnung	0.4105	0.005	4.8	0.402	0.419
3-Zimmerwohnung	0.6833	0.004	8.0	0.675	0.692
4-Zimmerwohnung	0.9002	0.005	7.9	0.891	0.909
5-Zimmerwohnung	1.0846	0.005	3.5	1.075	1.095
6-Zimmerwohnung	1.2122	0.007	1.6	1.198	1.226
Verweildauer (in Jahren)	-0.0077	0.000	1.1	-0.008	-0.008
Stockwerk	0.0045	0.000	1.2	0.004	0.005
Garage inbegriffen	0.0657	0.003	1.1	0.060	0.072
Attikawohnung	0.1659	0.005	1.0	0.157	0.175
Duplexwohnung	0.0670	0.003	1.1	0.061	0.073
Wohnbaugenossenschaft, Bund, Kanton, Gemeinde	-0.1667	0.002	1.1	-0.171	-0.162
Mikrovariablen					
Entfernung zu Seeufer < 200 m	0.0057	0.003	3.7	0.000	0.012
Distanz zum See (in km)	-0.0025	0.000	6.7	-0.003	-0.002
Seesicht:					
Kleine, 7 bis 100 ha (D)	0.0171	0.003	1.3	0.011	0.023
Mittlere, 101 bis 300 ha (D)	0.0474	0.004	1.5	0.041	0.054
Gute, > 300 ha (D)	0.0670	0.003	3.2	0.061	0.073
Bergsicht:					
Mittlere, 3 bis 12 Gipfel (D)	0.0043	0.002	1.9	0.000	0.008
Gute, >12 Gipfel (D)	0.0161	0.003	2.1	0.011	0.021
Neigung:					
4% - 9% (D)	0.0150	0.002	1.1	0.011	0.019
> 9% (D)	0.0341	0.003	1.1	0.029	0.039
Exposition:					
SuedWesten (D)	0.0090	0.002	1.2	0.005	0.013
NordWesten, nur Kanton GE (D)	0.0182	0.008	1.1	0.002	0.034
Lärm:					
Strassenlärm Nacht (über 45dB)	-0.0020	0.000	1.0	-0.003	-0.001
Bahnlärm Tag (über 50dB)	-0.0041	0.001	1.0	-0.006	-0.002
Beobachtungen	91'239				
Korrigiertes R ²	0.663				
Standardfehler der Schätzung (RMSE)	0.227				

Makrovariablen	Schätzwert	Std. Abw.	VIF	95%L	95%U
Erreichbarkeitspotenzial	0.0542	0.001	3.1	0.052	0.056
Steuerkraft pro 1 CHF	-0.2774	0.047	29.5	-0.369	-0.186
ln Steuersatz	0.5335	0.012	3.1	0.509	0.558
Bern	-0.0444	0.009	4.0	-0.061	-0.028
Lausanne	-0.1283	0.010	5.6	-0.147	-0.109
Genf	-0.0805	0.008	7.9	-0.096	-0.065
Kanton ZH ohne Stadt	-0.0480	0.004	3.1	-0.056	-0.040
Kanton BE ohne Stadt	-0.1131	0.008	7.5	-0.129	-0.097
Kanton VD ohne Stadt	-0.1222	0.008	8.1	-0.139	-0.106
Kanton GE ohne Stadt	-0.0986	0.009	5.7	-0.116	-0.082
Kanton LU	-0.1111	0.005	2.5	-0.121	-0.101
Kanton UR	-0.1888	0.017	1.1	-0.223	-0.155
Kanton SZ	-0.1553	0.011	2.8	-0.177	-0.134
Kanton OW	-0.1362	0.015	1.1	-0.166	-0.106
Kanton NW	-0.1009	0.012	1.2	-0.124	-0.077
Kanton GL	-0.1491	0.019	1.1	-0.185	-0.113
Kanton ZG	-0.1146	0.009	2.7	-0.132	-0.097
Kanton FR	-0.1374	0.009	2.3	-0.155	-0.119
Kanton SO	-0.1505	0.009	2.1	-0.168	-0.133
Kanton BS	-0.0277	0.008	6.6	-0.043	-0.013
Kanton BL	-0.0610	0.007	3.7	-0.075	-0.047
Kanton SH	-0.1567	0.009	1.3	-0.175	-0.138
Kanton AR	-0.1857	0.012	1.1	-0.209	-0.163
Kanton AI	-0.1429	0.038	1.0	-0.218	-0.068
Kanton SG	-0.1355	0.006	3.4	-0.147	-0.124
Kanton GR	-0.0635	0.008	1.4	-0.078	-0.049
Kanton AG	-0.1034	0.005	2.7	-0.114	-0.093
Kanton TG	-0.1874	0.007	1.8	-0.200	-0.174
Kanton TI	-0.1461	0.006	1.6	-0.158	-0.135
Kanton VS	-0.2723	0.010	1.7	-0.292	-0.252
Kanton NE	-0.3008	0.012	5.4	-0.324	-0.277
Kanton JU	-0.2388	0.018	1.3	-0.273	-0.204
Konstante	2.6260	0.080	0	2.468	2.784

D=Dummy Variable
ln = natürlicher Logarithmus

Abbildung 33: Histogramm der Residuen

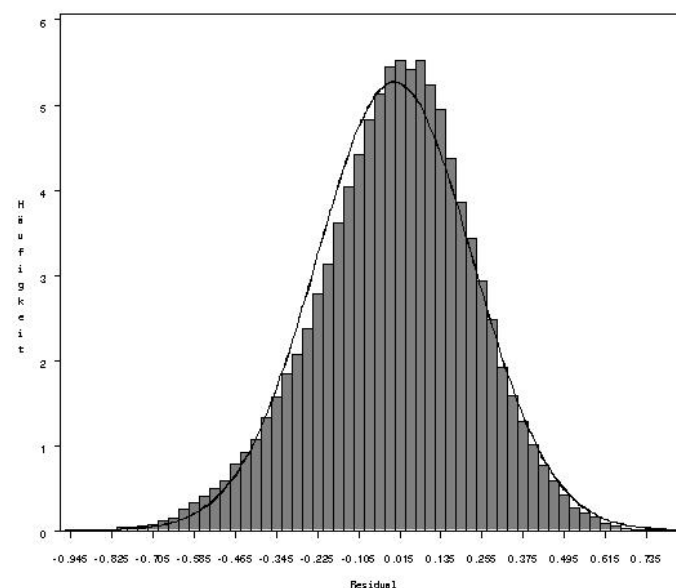


Abbildung 34: Residuen vs. Log(Nettomiete) Plot

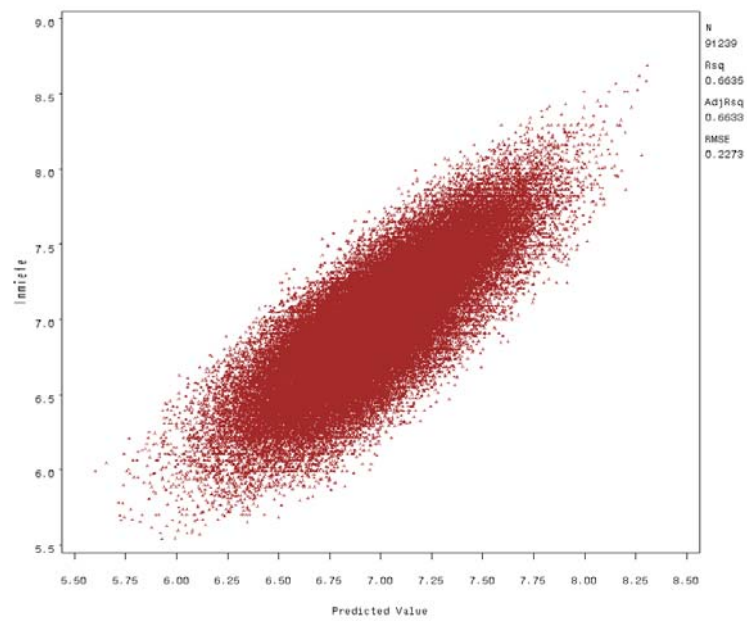


Tabelle 27: Vergleich der Koeffizienten der Mikro- und Makrovariablen aus verschiedenen Modellen

Mikrovariablen	Model AZ	Model BZ	Model CZ	Model FZ
Entfernung zu Seeufer < 200 m	0.0113	0.0171	0.0190	0.0061
Distanz zum See (in km)	-0.0030	-0.0034	-0.0037	-0.0026
Seesicht:				
Kleine, 7 bis 100 ha (D)	0.0188	0.0189	0.0176	0.0172
Mittlere, 101 bis 300 ha (D)	0.0509	0.0488	0.0503	0.0477
Gute, > 300 ha (D)	0.0712	0.0657	0.0659	0.0668
Bergsicht:				
Mittlere, 3 bis 12 Gipfel (D)	0.0057	0.0043*	0.0039*	0.0042
Gute, >12 Gipfel (D)	0.0151	0.0167	0.0175	0.0163
Neigung:				
4% - 9% (D)	0.0145	0.0148	0.0193	0.0150
> 9% (D)	0.0289	0.0307	0.0333	0.0343
Exposition:				
SuedWesten (D)	0.0104	0.0091	0.0092	0.0090
NordWesten, nur Kanton GE (D)	0.0201	0.0210	0.0151*	0.0187
Lärm:				
Strassenlärm Nacht (über 45dB)	-0.0019	-0.0023	-0.0021	-0.0020
Bahnlärm Tag (über 50dB)	-0.0045	-0.0037	-0.0035	-0.0040
Makrovariablen				
Erreichbarkeitspotenzial	0.060	0.052	0.050	0.055
Steuerkraft pro 1 CHF	0.519	0.504	0.496	0.526
In Steuersatz	-0.216	-0.383	-0.424	-0.287
Bern	-0.056	-0.032	-0.0186*	-0.043
Lausanne	-0.159	-0.124	-0.114	-0.127
Genf	-0.110	-0.069	-0.062	-0.079
Kanton ZH ohne Stadt	-0.046	-0.060	-0.072	-0.047
Kanton BE ohne Stadt	-0.125	-0.108	-0.113	-0.111
Kanton VD ohne Stadt	-0.139	-0.117	-0.123	-0.119
Kanton GE ohne Stadt	-0.125	-0.084	-0.075	-0.096
Kanton LU	-0.117	-0.110	-0.112	-0.110
Kanton UR	-0.189	-0.137	-0.137	-0.185
Kanton SZ	-0.154	-0.188	-0.204	-0.155
Kanton OW	-0.142	-0.143	-0.159	-0.135
Kanton NW	-0.122	-0.131	-0.142	-0.101
Kanton GL	-0.136	-0.162	-0.185	-0.146
Kanton ZG	-0.110	-0.130	-0.143	-0.114
Kanton FR	-0.150	-0.146	-0.153	-0.136
Kanton SO	-0.164	-0.149	-0.157	-0.148
Kanton BS	-0.041	-0.020	-0.01882	-0.028
Kanton BL	-0.068	-0.061	-0.067	-0.058
Kanton SH	-0.170	-0.168	-0.175	-0.155
Kanton AR	-0.192	-0.194	-0.203	-0.185
Kanton AI	-0.132	-0.0672*	-0.0927*	-0.142
Kanton SG	-0.143	-0.134	-0.141	-0.134
Kanton GR	-0.063	-0.076	-0.082	-0.061
Kanton AG	-0.112	-0.116	-0.127	-0.102
Kanton TG	-0.189	-0.202	-0.213	-0.186
Kanton TI	-0.181	-0.176	-0.193	-0.146
Kanton VS	-0.286	-0.267	-0.261	-0.270
Kanton NE	-0.318	-0.301	-0.304	-0.298
Kanton JU	-0.251	-0.229	-0.240	-0.236

*nicht signifikant, **Vorzeichen geändert

7.4.2 Ausreisser Analyse

Tabelle 28: Deskriptives der Ausreisserbeobachtungen im Vergleich zum Modellsample

Variable	Outliers, N=4'640				Sample, N=91'239			
	Mittel	Min	Max	Median	Mean	Min	Max	Median
Nettomiete	1'156	9	16'400	710	1'169	255	5'950	1'085
Wohnfläche	101	7	989	90	85	8	548	80
Anzahl Zimmer	4	1	6	4	3	1	6	3

7.5 Tabellenverzeichnis

Tabelle 1: Übersicht über die Stufen der Treffergenauigkeit des ZKB Algorithmus	13
Tabelle 2: Ergebnis des ZKB Algorithmus	13
Tabelle 3: Vergleich der Geokodierungen	14
Tabelle 4: Verteilung der Beobachtungen über die Entfernungsklassen (N=234'153)	14
Tabelle 5: Verteilung über die Entfernungsklassen	16
Tabelle 6: Übersicht über die verwendeten Geokoordinaten (Herkunft versus Differenzen)	17
Tabelle 7: Übersicht Steuerklassen.....	20
Tabelle 8: Mengengerüst zu den Selektionsschritten der MSE-Daten	21
Tabelle 9: Ausschlusskriterien und Mengengerüst (Zwischensample, N=106'691)	22
Tabelle 10: Stichprobenumfang	22
Tabelle 11: Analyse der Nettomiete (Häufigkeit)	23
Tabelle 12: Bauperiode, Abgleich von MSE2003 und GWR Daten.....	24
Tabelle 13: Durchschnittliche Nettomiete und Wohnfläche pro Zimmer	25
Tabelle 14: Missingsanalyse Eigentümertyp.....	30
Tabelle 15: Verteilung Insample Eigentümertyp.....	31
Tabelle 16: Verteilung Imputation Eigentümertyp gesamte Stichprobe	31
Tabelle 17: Datentransformationen, Strukturvariablen	32
Tabelle 18: Korrelationen zwischen Lärmvariablen und Nettomiete	33
Tabelle 19: Datentransformationen, Variablen der Makro- und Mikrolage.....	34
Tabelle 20: Definition der Variablen.....	36
Tabelle 21: Schätzergebnisse	37
Tabelle 22: Standardisierte Koeffizienten, sortiert nach relativer Stärke.....	39
Tabelle 23: Modellvergleiche: Erklärungsgehalt (Korrigiertes R^2) und Samplegrösse	41
Tabelle 24: Koeffizientenvergleich der Strukturvariablen aus verschiedenen Modellen	42
Tabelle 25: In-Sample und Out-of-Sample Auswertungen.....	43
Tabelle 26: Modellergebnisse erweitert um Variance Inflation Factors (VIF) und 95%- Konfidenzintervalle	58
Tabelle 27: Vergleich der Koeffizienten der Mikro- und Makrovariablen aus verschiedenen Modellen	60
Tabelle 28: Deskriptives der Ausreisserbeobachtungen im Vergleich zum Modellsample.....	60
Tabelle 29: Beiliegende SAS-Daten und Codes.....	62

7.6 Abbildungsverzeichnis

Abbildung 1: Unterteilung einer MPI-Stichprobenzelle für die Qualitätsbereinigung	9
Abbildung 2: Verteilung der Beobachtungen über die Entfernungsklassen (N=165'380)	15
Abbildung 3: Verteilung der Nettomieten (Histogramm)	24
Abbildung 4: Durchschnittliche Nettomiete pro Zimmer, Vergleich für Kanton Zürich, Zug und Tessin	25
Abbildung 5: Eigentübertypverteilung in der Stichprobe	26
Abbildung 6: Verteilung Eigentübertyp pro Kanton	26
Abbildung 7: Verweildauer in der Stichprobe	27
Abbildung 8: Weitere Wohnungsmerkmale	28
Abbildung 9: Häufigkeiten und durchschnittliche Nettomieten verschiedener Wohnungstypen	28
Abbildung 10: Durchschnittliche Nettomieten und Häufigkeiten der 3-Zimmerwohnungen nach Kanton	29
Abbildung 11: Residuenplot – Residuen und prognostizierte Werte	40
Abbildung 12: Schematische Darstellung der Berechnung	44
Abbildung 13: Hektar – Lagewerte für die Schweiz	45
Abbildung 14: Anzahl geokodierter Beobachtungen der MSE pro Postleitzahlgebiet	48
Abbildung 15: Anzahl nicht-geokodierter Beobachtungen der MSE pro Postleitzahlgebiet	49
Abbildung 16: Strassenlärm Tag	49
Abbildung 17: Strassenlärm Nacht	50
Abbildung 18: Bahnlärm Tag	50
Abbildung 19: Bahnlärm Nacht	51
Abbildung 20: Fluglärm Tag (ohne Militär- und Privatflughäfen)	51
Abbildung 21: Fluglärm Nacht (alle Flughäfen)	52
Abbildung 22: Aussicht	52
Abbildung 23: Seesicht	53
Abbildung 24: Bergsicht unter Einbeziehung von 217 Gipfeln	53
Abbildung 25: Hangneigung in Prozent	54
Abbildung 26: Hangneigung in Grad	54
Abbildung 27: Exposition	55
Abbildung 28: Distanz zum See in km	55
Abbildung 29: Beispiel für die Gewichtung der Anzahl Arbeitsplätze	56
Abbildung 30: Erreichbarkeitspotenzial	56
Abbildung 31: Detailkarte Erreichbarkeitspotenzial	57
Abbildung 32: Unterschiede in der Verteilung der Nettomieten für imputierte und nicht-imputierte Werte	57
Abbildung 33: Histogramm der Residuen	58
Abbildung 34: Residuen vs. Log(Nettomiete) Plot	59

7.7 SAS-Daten und Programme

Tabelle 29: Beiliegende SAS-Daten und Codes

SAS-Daten	Beschreibung
koef_fuer_scoring_red	Die Strukturkoeffizienten zur Berechnung qualitätsbereinigter Mieten für die MPI-Daten
mikromakrolage_schweiz	Lagewerte Schweiz (für jeden Hektar)
SAS-Programmcode	Beschreibung
mpiscoring_final.sas	Programm zur Berechnung qualitätsbereinigter Mieten für die MPI-Daten