

## Definition von homogenen hedonischen Preisregionen für Österreich mittels SKATER-Algorithmus

Marco Helbich

GIScience, Geographisches Institut, Universität Heidelberg

Berliner Straße 48, 69120 Heidelberg, Deutschland

e-mail: marco.helbich@geog.uni-heidelberg.de

Wolfgang A. Brunauer

IRG Immobilien Rating GmbH

Taborstr. 1-3, 1020 Wien, Österreich

e-mail: wolfgang.brunauer@irg.at

### Zusammenfassung:

Dieser Artikel präsentiert einen datengetriebenen Ansatz zur Ableitung von räumlich homogenen Regimen auf Basis eines lokalen hedonischen Immobilienpreismodells für Einfamilienhäuser in Österreich. Die mit Geographically Weighted Regression geschätzten räumlich variierenden Koeffizienten werden mithilfe von Hauptkomponentenanalyse und (räumlichen) Clusteralgorithmen in homogene, kohärente Regionen zusammengefasst. Die so generierten Regionen bzw. Teilmärkte erlauben die Modellierung räumlicher Heterogenität bei gleichzeitiger Vermeidung von großer Volatilität in den Parameterschätzungen wie in rein lokalen Modellen. Die Ergebnisse finden Anwendung in der automatisierten Bewertung von Kleinimmobilien in Österreich.

## 1 Einleitung

Die Anwendung von automatisierten Bewertungssystemen bei Hypothekendarlehen hat in den letzten Jahren zunehmende Akzeptanz erfahren. Ein wesentlicher Grund dafür liegt in den Eigenkapitalvorschriften des Basler Ausschusses für Bankenaufsicht („Basel II“), die seit dem Jahr 2007 in den EU-Mitgliedsstaaten für alle Kredit- und Finanzdienstleistungsinstitute bindend sind. Konkret fordert die in Österreich gültige Solvabilitätsverordnung, dass der Wert einer Wohnimmobilie mindestens alle drei Jahre überprüft werden muss (in volatilen Märkten in kürzeren Zyklen). Immobilienpreisindizes stellen in diesem Zusammenhang ein adäquates Mittel dar, um Sicherheitenportfolien zu bewerten. Aus diesem Grund finden automatisierte Immobilienbewertungsmodelle im Bankenbereich zunehmend Verbreitung.

Eine besondere Rolle kommt dabei sogenannten „hedonischen Immobilienpreismodellen“ zu (basierend auf den Konzepten von Lancaster 1966 und Rosen 1974). In der angewandten Ökonomie werden Immobilien als aus Einzeleigenschaften zusammengesetzte Güter angesehen, wobei jede dieser Eigenschaften ihren spezifischen Preis besitzt (McDonald 1997). Hedonische Immobilienpreismodelle erklären daher Preise durch die Lage- und Ausstattungsmerkmale der jeweiligen Immobilien, üblicherweise im Rahmen multipler Regressionsmodelle. Dabei ist zwischen „Struktureigenschaften“ der Immobilie selbst sowie deren Lageeigenschaften bzw. räumlichen Externalitäten zu unterscheiden.

Da es sich bei der Immobilie also um ein „Bündel“ von Eigenschaften handelt, die sich nicht getrennt handeln lassen und die zudem räumlich gebunden sind, ergeben sich zwei wesentliche Problemstellungen bei hedonischen Immobilienpreismodellen: Erstens ist zu erwarten, dass die marginalen Preiseffekte der Charakteristika nicht linear sind. Eine übliche Modellspezifikation um Heteroskedastizität in den Residuen und Nichtlinearität in den Effekten zu reduzieren ist die (semi-)log-Form (Malpezzi 2002), die zudem dem multiplikativen Charakter der Immobilieneigenschaften entspricht. Die zweite Herausforderung ist die sich durch räumliche Gebundenheiten der Immobilie ergebende räumliche Heterogenität (d.h. Variabilität des Effektes wertbestimmender Eigenschaften), die bei Nichtberücksichtigung zu einer Verletzung von Modellannahmen und lokal zu erhebliche verzerrten Parameterschätzungen führen kann. Aus diesem Grund müssen räumliche Effekte explizit in der Modellstruktur berücksichtigt werden. Gegenwärtig finden beispielsweise räumliche bzw. raum-zeitliche Modelle von Anselin (1988), Brunsdon et al. (1996), Wood (2006), Diggle und Ribeiro (2007) oder LeSage und Pace (2009) ihre Anwendung. Obwohl räumliche Heterogenität, bei der es sich im Fall von Immobilienmärkten um „Teilmärkte“ handelt, eine bedeutende Rolle bei der Modellierung von Hauspreisen spielt (Bourassa et al. 2007), werden räumliche Cluster meist exogen (z.B. auf Basis von Verwaltungsgrenzen) definiert.

Ziel dieses Beitrages ist daher, eine datengetriebene/endogene Definition von räumlichen Teilmärkten für den österreichischen Einfamilienhausmarkt zu finden. Diese homogenen Modellregionen bieten damit den Ansatzpunkt für effizientere Modelle mit räumlichen Interaktionsparametern ohne den Nachteil großer Volatilität in den Parameterschätzungen.

Dieser Artikel ist wie folgt strukturiert: In Abschnitt 2 werden das Untersuchungsgebiet und der zugrundeliegende Datensatz vorgestellt. Abschnitt 3 skizziert die angewandte Methodik und diskutiert die resultierenden Modellergebnisse, und in Abschnitt 4 schließlich wird eine Einordnung der Ergebnisse in den theoretischen Kontext vorgenommen und der praktische Nutzen einer solchen Regionalisierung erörtert.

## 2 Untersuchungsgebiet und Datengrundlage

Der Datensatz umfasst die Kaufpreise und das Kaufdatum von 3892 geokodierten österreichischen Einfamilienhäusern mit ihren Eigenschaften (Fläche, Qualität, Alter etc.) für den Zeitraum von 1998 bis 2009. Die Daten wurden im Zuge von Kreditvergaben der Bank Austria UniCredit Group erhoben und dienen zur Bewertung von Immobiliensicherheiten. Abbildung 1 visualisiert die räumliche Verteilung der Beobachtungen einerseits anhand koordinativ verorteter Standorte und andererseits auf Basis von Dichtewerten einer Kerndichteschätzung. Interpretativ repräsentieren Zellen mit heller Kolorierung eine hohe Standortdichte und dunkel eingefärbte Zellen eine geringe. Allgemein ist ein, der Topographie folgendes, Ost-West-Gefälle und eine Konzentration der Objekte auf urbane Gebiete, wie zum Beispiel Wien, Linz und Graz, erkennbar.

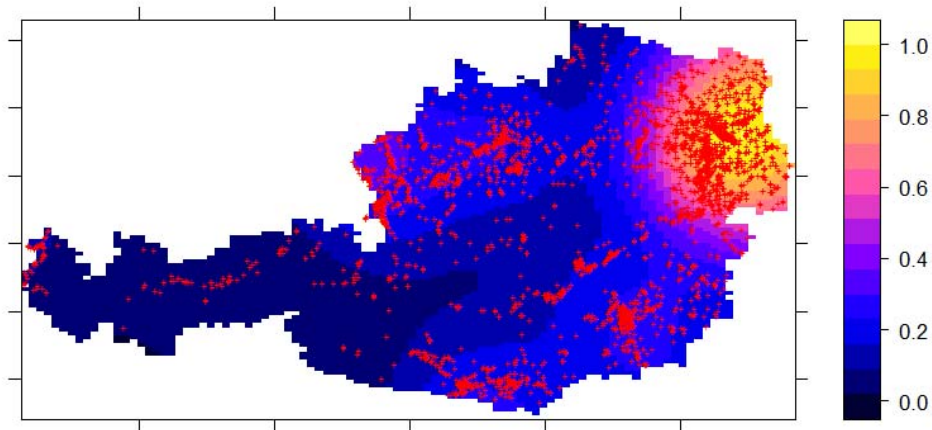


Abbildung 1: Räumliche Verteilung der Einfamilienhaus-Standorte ( $n=3892$ ) als Punktsignatur und Dichtewerte

Zusätzlich zu den Immobilieneigenschaften umfasst der Datensatz Lage- bzw. Nachbarschaftseigenschaften der amtlichen Statistik (Statistik Austria), die auf demographischen und sozio-ökonomischen Eigenschaften (wie etwa Akademikerquote oder Arbeitslosenrate) beruhen und ergänzend auf drei hierarchischen administrativen Einheiten (Zählsprenkel-, Gemeinde- und Bezirksebene) für das Jahr 2001 bzw. 2009 zur Verfügung stehen. Sogenannte Compositional Effects (Goldstein 2003) erlauben es, räumliche Heterogenität anhand von hierarchischen Strukturen zu berücksichtigen. Dabei werden jeweils die Abweichungen einer administrativen Hierarchiestufe (z.B. Gemeinde) vom Mittel der Nächsthöheren (z.B. Bezirk) gebildet. Der Effekt der resultierenden Variable kann als Hierarchieebenen-spezifischer Differenzeffekt vom Effekt der jeweils übergeordneten Ebene interpretiert werden.

### 3 Methodik

Wie in Abschnitt 1 beschrieben, ist das Ziel dieser Untersuchung, homogene Preisregionen für die hedonische Modellbildung abzuleiten. Um dieses Vorgehen zu motivieren, zeigt Abschnitt 3.1.1 die aus einer globalen Modellierung – ohne Berücksichtigung lokaler Parametervariation – resultierenden Probleme (insb. Heteroskedastizität und Autokorrelation) sowie Lösungsansätze auf. Zur Erfassung der räumlichen Heterogenität bietet sich daher ein lokales Modell an, das jedoch zu stark volatilen Parameterschätzungen führt und daher für den praktischen Einsatz von Preisprognosen nur bedingt geeignet ist (Abschnitt 3.1.2). Aufbauend auf den Ergebnissen eines adaptierten lokalen Modells werden in Abschnitt 3.2 Regionen ableitbar, innerhalb derer die Parametervariation gering ist. Das Ergebnis ist eine datengetriebene Teilmarktdefinition für den österreichischen Einfamilienhausmarkt.

#### 3.1 Hedonische Preismodellierung

##### 3.1.1 Globale Modelle<sup>1</sup>

Wie in Abschnitt 1 beschrieben werden hedonische Preismodelle üblicherweise als (semi-)log Modelle spezifiziert. Dabei wird stets die abhängige Variable (der Kaufpreis) logarithmiert, die erklärenden Variablen gehen linear (kategoriale Variablen wie die Qualität des Gebäudes oder solche, die nicht strikt positiv sind, wie etwa das Alter) oder logarithmiert (kontinuierliche Variablen wie die Wohnfläche) in das Modell ein. In solch einem Modell kann der Koeffizient der logarithmierten erklärenden Variable als Elastizität interpretiert werden (d.h. die prozentuelle Änderung der abhängigen Variable bei einer Änderung der erklärenden Variable um 100%) bzw. bei einer nichtlogarithmierten erklärenden Variable approximativ als semi-Elastizität (d.h. die prozentuelle Änderung der abhängigen Variable bei einer Änderung der erklärenden Variable um eine Einheit).

Das Ausgangsmodell bildete also ein log-lineares Ordinary Least Square (OLS) Modell mit 24 vorselektierten Variablen, die wiederum über Minimierung des Akaike Information Criteria (AIC) schrittweise auf 19 Variablen reduziert wurden. Die Ergebnisse zeigen, dass ein schlechter Zustand des Hauses den Preis um 6%, die schlechte Ausstattung des Bades und der Heizung den Preis um 7% bzw. 10% reduziert. Im Gegensatz erhöht das Vorhandensein eines Kellers den Preis um 12%. Hinsichtlich der Nachbarschaftseigenschaften zeigte sich, dass eine um 1 Prozentpunkt höhere Arbeitslosenquote zu einer Preisreduktion von rund 2% führt, aber urbane Gebiete einen positiven Effekt (Elastizität von +7%) auf den Preis aus-

---

<sup>1</sup> Da im Kern der Betrachtungen die Ableitung regionaler Cluster steht, werden bei sämtlichen Modellen die Schätzergebnisse nicht im Detail diskutiert.

üben. Eine Überprüfung der zugrundeliegenden Modellannahmen verwies jedoch einerseits auf signifikant autokorrelierte Residuen (Moran's  $I = 0,167$ ;  $p < 0,001$ ) – dies bewirkt zwar keine Verzerrung der geschätzten Koeffizienten, jedoch werden die Standardfehler unter- und die  $t$ -Werte überschätzt – andererseits auf Heteroskedastizität (Breusch-Pagan Teststatistik (BP) = 161,523;  $p < 0,001$ ).

Im Falle von Heteroskedastizität besteht die Möglichkeit, diese implizit in die Modellstruktur über einen Feasible Generalized Least Square Ansatz (FGLS, Fahrmeir et al. 2009) zu integrieren, was abermals zu heteroskedastischen (BP: 163,417;  $p < 0,001$ ) und autokorrelierten (Moran's  $I: 0,181$ ;  $p < 0,001$ ) Residuen führte.

Da diese Modellverletzungen ebenso ein Resultat autokorrelierter Eingangsdaten sein können, wurden in einem nächsten Schritt räumliche Effekte explizit in das Modell integriert. In der Ökonometrie werden hierzu oftmals Simultaneous Autoregressive Modelle (Anselin 1988) verwendet. Diese Modellklasse beinhaltet sowohl das Spatial Lag, als auch das Spatial Error Model. Anhand der Lagrange Multiplier Statistik fiel die Wahl auf Zweiteres.

Zusammenfassend liefern das FGLS und das Spatial Error Model ähnliche Ergebnisse wie die OLS-Schätzung. Bei einem Vergleich der Modellgüte wies FGLS, gemessen am AIC, die bessere Performanz auf. In Summe zeigten aber sämtliche globale Modelle Schwachstellen hinsichtlich der Modellannahmen.

### 3.1.2 Lokale Modelle

Globale Modelle basieren auf der Annahme, dass die Beziehungen zwischen abhängiger und den unabhängigen Variablen im Untersuchungsgebiet räumlich konstant sind. Ist dies nicht der Fall, kann dies zu Heteroskedastizität und Autokorrelation führen. Ein Ansatz um Heterogenität in den Parameterschätzungen zu modellieren, ist die Geographically Weighted Regression (GWR, Brunsdon et al. 1996). Im vorliegenden Fall wurden – dem Prinzip der Parsimonie folgend – nur jene Variablen für die Schätzung des lokalen Modells herangezogen, die im globalen Modell höchste Signifikanz aufwiesen. Da 7 der 16 Variablen (Tab. 1) auf Basis der Teststatistik nach Leung et al. (2000) keine räumliche signifikante Variation in den Parameterschätzungen aufwiesen, konnten diese als global konstant angenommen werden. Zu diesem Zweck mussten die globalen und lokalen Effekte in einem mehrstufigen Verfahren getrennt geschätzt werden. In der Literatur ist dieses Vorgehen als Mixed GWR (MGWR, Fotheringham et al. 2002) angeführt. Die Prüfung der Residuen ergab, dass ein Großteil der Autokorrelation der Residuen ( $I = 0,019$ ;  $p = 0,013$ ) mittels einer lokalen Modellierung absorbiert werden konnte und ließ den Schluss zu, dass Heteroskedastizität in den globalen Modellen auf eine tatsächlich vorhandene Variation in den Parametern rückführbar ist. Tabelle 1 listet die resultie-

renden Parameterschätzungen<sup>2</sup> auf, wobei im oberen Teil die Effekte der Variablen mit variierenden Koeffizienten aufgeführt sind, im unteren Teil jene der räumlich konstanten Variableneffekte, außerdem vergleicht sie die globalen OLS-Schätzungen (linke Hälfte) mit den lokalen Schätzungen aus der MGWR (rechte Hälfte). Dabei zeigt sich bei den meisten signifikant variierenden Effekten, dass die ersten und dritten Quantile der lokalen Schätzungen außerhalb der Konfidenzintervalle der globalen Schätzungen liegen, was erneut die Wichtigkeit der Modellierung räumlicher Heterogenität betont.

Tabelle 1: Parameterschätzungen: fixe und variierende Koeffizienten\*

		global			lokal		
		Schätz.	Std.fehl.	t-Stat.	1. Quantil	Median	3. Quantil
variabel	gem_kaufind_09***	0,005	0,001	4,497	0,000	0,003	0,006
	gem_abi***	0,012	0,003	3,897	0,010	0,016	0,020
	gem_alter_index***	-0,045	0,005	-8,635	-0,040	-0,030	-0,025
	ln_gem_dichte***	0,066	0,009	7,459	0,038	0,075	0,083
	cond_house_3*	-0,044	0,019	-2,236	-0,098	-0,085	-0,004
	attic_dum**	-0,061	0,019	-3,137	-0,063	-0,039	-0,032
	lnarea_plot***	0,092	0,020	4,527	0,040	0,112	0,151
	lnarea_total***	0,466	0,029	15,906	0,375	0,423	0,518
	age_num***	-0,006	0,001	-11,205	-0,006	-0,005	-0,005
fix	intercept	-0.007	0.009	-0.828			
	gem_alq_09	-1.232	1.735	0.478			
	zsp_abi_dgem***	0.008	0,002	3.939			
	heat_3***	-0.132	0.038	-3.491			
	bath_3	-0.017	0.035	-0.484			
	cellar_dum***	0.117	0.022	5.368			
	garage_3***	-0,094	0,020	-4.765			
	terr_dum*	0.044	0.019	2.257			

\* gem\_kaufind\_09 = Kaufkraftindex je Gemeinde 2009, gem\_abi = Akademikeranteil je Gemeinde 2009, gem\_alter\_index = durchschnittliches Alter der Einwohner pro Gemeinde, ln\_gem\_dichte = Log der Anzahl der Personen pro ha pro Gemeinde, cond\_house\_3 = Schlechter Zustand des Hauses, attic\_dum = Dachgeschoss (Dummy), lnarea\_plot = Grundstücksfläche, lnarea\_total = Wohnfläche, age\_num = Gebäudealter, gem\_alq\_09 = Arbeitslosenquote je Gemeinde 2009, zsp\_abi\_dgem = Abweichung Akademikeranteil je Zählsprengel vom Gemeindemittel 2001, heat\_3 = Schlechte Qualität

<sup>2</sup> Um die Rechenzeit zu Reduzieren wurde ein Sample von 35% (1393 Objekte) herangezogen.

des Heizungssystems (Dummy), bath\_3 = Niedere Qualität des Bades (Dummy), cellar\_dum = Keller (Dummy), garage\_3 = niedere Qualität oder keine Garage (Dummy), terr\_dum = Terrasse (Dummy) / Signif.: 0 \*\*\*\* 0.001 \*\*\* 0.01 \*\* 0.05

Die räumlich signifikant variierenden Parameterschätzungen für jede Immobilie wurden in weitere Folge mittels Ordinary Kriging (Isaaks und Srivastava 1989) interpoliert und bilden nun den Ausgangsdatensatz zur Ableitung der räumlichen Taxonomie für den Einfamilienhausmarkt.

## 3.2 Ableitung der räumlichen Taxonomie

### 3.2.1 Hauptkomponentenanalyse

Das zugrundeliegende MGWR Modell ergab 9 räumlich variierende Variableneffekte (Tab. 1). Als Vorprozessierung zur Clusteranalyse wurde die Variation dieser Effekte auf Hauptkomponenten reduziert und nicht relevante Information eliminiert, was eine präzisere Diskriminierung der Cluster ermöglicht. Die resultierenden Hauptkomponenten repräsentieren eine Linearkombination der Eingangsdaten und besitzen die Eigenschaft, dass diese orthogonal aufeinander stehen und somit unkorreliert sind (Reimann et al. 2008). Da meist bereits wenige Hauptkomponenten einen Großteil der in den Daten vorhandenen Varianz erklären, findet gleichfalls eine Dimensionsreduktion der Eingangsdaten statt. Im vorliegenden Fall wurden die 9 Eingangsvariablen auf vier Hauptkomponenten, die kumulativ rund 97% der Gesamtvarianz erklären, reduziert. Abbildung 2 stellt dabei Loadings und Scores der ersten beiden Hauptkomponenten (erklären ca. 84% der Varianz) in Beziehung.



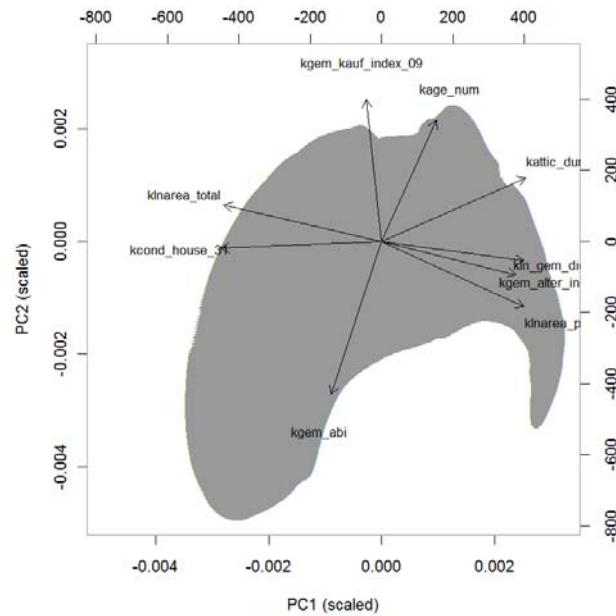


Abbildung 2: Visualisierung der ersten beiden Hauptkomponenten als Biplot

Es zeigt sich einerseits eine gewisse Gruppierung der Variablen und andererseits, dass etwa die Variablen *Inarea\_total*, *cond\_house\_3*, *In\_gem\_dichte*, *gem\_alter\_index* und *Inarea\_plot* ähnliche Ladungen aufweisen und folglich ähnliche Informationen repräsentieren. Da die Taxonomie auf Basis der administrativen Einheiten der Gemeinde vollzogen werden soll (um eine nachträgliche Verknüpfung mit weiteren Daten zu ermöglichen), wurden die Rasterzellenwerte der vier Hauptkomponenten mittels Mittelwertbildung aggregiert. Die Dimension der resultierenden und zur Clusterung herangezogenen Matrix beträgt schlussendlich 2379 Zeilen (Gemeinden) mal 4 Spalten (Hauptkomponenten).

### 3.2.2 Clusteranalyse

Ziel der Clusteranalyse ist es, die einzelnen Entitäten in Gruppen zu diskriminieren (Everitt 1993). Dabei soll die Ähnlichkeit der Objekte innerhalb eines Clusters maximiert und jene zwischen den Clustern minimiert werden. Die Literatur bietet hierzu zahlreiche Ansätze wie beispielsweise hierarchische, probabilistische, graphenbasierte, dichte-basierte oder modellbasierte Algorithmen an (Han und Kamber 2006). Schwachpunkte dieser Algorithmen sind, dass sie a) keine zusammenhängenden Regionen ableiten und b) die resultierenden Regionen scharfe Clustergrenzen aufweisen. Zweiteres wurde in der geographischen Literatur bereits diskutiert (Leung

1987) und ist durch die Anwendung von fuzzy Algorithmen (Bezdek 1981; Kaufman und Rousseeuw 1990) handhabbar. Um zusammenhängende Regionen zu erzielen, muss ein Algorithmus Nachbarschaftsbeziehungen zwischen den Objekten bei der Clusterung mitberücksichtigen (Openshaw und Rao 1995). Hierzu bieten sich etwa der SKATER-Algorithmus (Assuncao et al. 2006) an. Das Akronym SKATER steht dabei für "Spatial 'K'luster Analysis by Tree Edge Removal". Die Nachbarschaftsbeziehung zwischen Objekten wird dabei über einen Graphen implementiert, der die einzelnen Objekte (deren Zentroide) mittels Kanten und Knoten verbindet. Dabei sind die einzelnen Kanten mit Kosten versehen, die die Unähnlichkeit zwischen den Objekten ausdrückt und sich aus deren Attributwerten errechnet. Über das Zerschneiden und Stutzen des Graphen erhält man schließlich räumlich zusammenhängende homogene Cluster. Weitere Details zum Algorithmus sind Assuncao et al. (2006) zu entnehmen. Abbildung 3 visualisiert die resultierende Partitionierung für eine unterschiedliche Anzahl an Clustern.

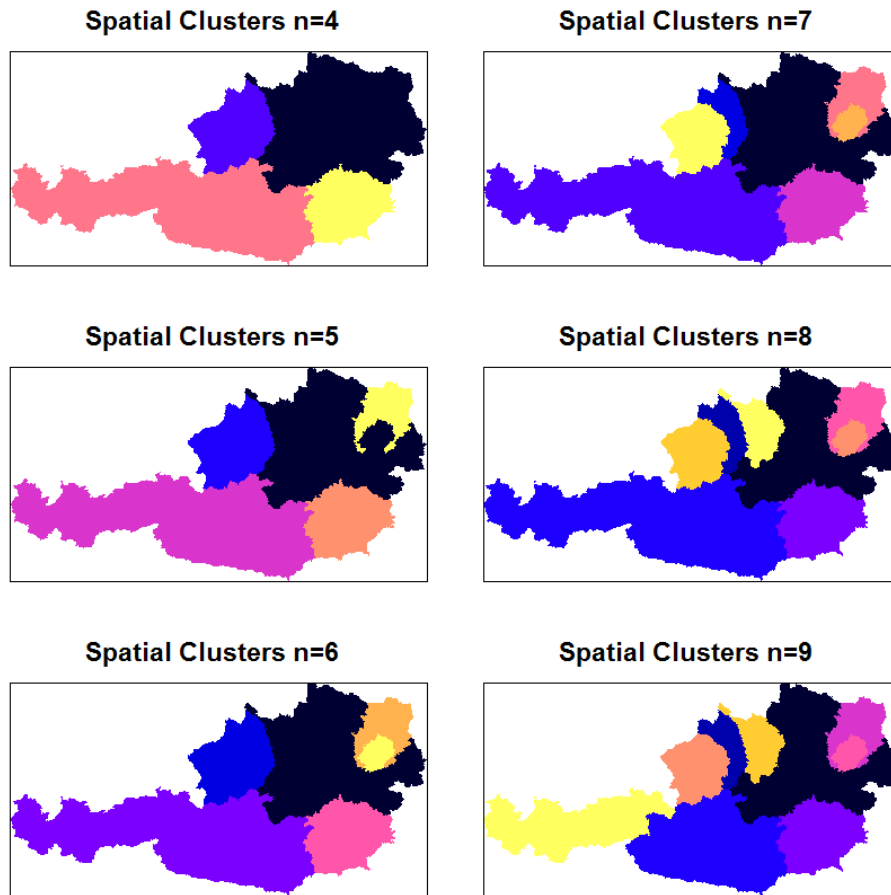


Abbildung 3: Räumliche Regime der Parameterschätzungen

Interpretativ erscheint das Clusterergebnis mit sechs Regionen (Abb. 3, links unten) am adäquatesten und wird nun kurz diskutiert. Bei dieser Partitionierung werden erstmals die Bundeshauptstadt Wien und Teile des Stadt-Umlandes als separater Cluster ausgewiesen, was der Sonderstellung Wiens am Immobilienmarkt entspricht. Umgeben wird dieser Cluster einerseits von einer Region die in etwa das Weinviertel repräsentiert und andererseits einem Cluster der grob das restliche Bundesland Niederösterreich und Teile des Bundeslands Oberösterreich umfasst. Einen vierten Cluster formen der Oberösterreichische sowie Salzburger Zentralraum, ein hinsichtlich des Immobilienmarktes höchst dynamischer Bereich. Gleiches ist für die südwestlich gelegene Region um Graz zu sagen. Den flächenmäßig umfangreichsten Cluster bilden die östlichen und südlichen Bundesländer Vorarlberg, Tirol, Kärnten und Teilen der westlichen Steiermark.

## 4 Resümee

Primäres Ziel dieser Untersuchung war die Ableitung räumlicher Teilmarktdefinitionen auf Basis eines lokalen hedonischen Regressionsmodells. Hintergrund dafür ist die Unzulänglichkeit globaler Modelle, räumliche Parametervariationen adäquat zu modellieren. Es wurde eine zweistufige Vorgehensweise vorgeschlagen, in der zunächst ein (mixed) GWR-Modell geschätzt wird und daraus die variierenden Parameter abgeleitet werden. Da die Ergebnisse der GWR, zum Beispiel im Rahmen von automatisierten Bewertungsmodellen, schlecht operationalisierbar sind, wird nach einer Möglichkeit gesucht, einen „Bias-Varianz-Trade off“ zwischen Variabilität in den Parametern und Robustheit in der Parameterschätzung zu schaffen. Zu diesem Zweck werden die räumlich variierenden Parameterschätzungen aus der GWR (nach Orthogonalisierung über Hauptkomponentenanalyse) als Inputvariablen für eine räumliche Clusteranalyse mittels SKATER-Algorithmus herangezogen, die kohärente räumliche Teilmarktdefinitionen ergeben. Auf diese Weise erhält man eine methodengestützte und datengetriebene Spatial Regime Definition, die im Anschluss für hedonische hierarchische Regressionsmodelle mit räumlich variierenden Parametern verwendet werden kann.

## 5 Literatur

- Anselin, L (1988): *Spatial Econometrics. Methods and Models*. Kluwer, Dordrecht
- Assuncao R, Neves M, Camara G, Freitas C (2006): Efficient Regionalization Techniques for Socio-Economic Geographical Units Using Minimum Spanning Trees. In: *International Journal of Geographical Information Science*, 20, S. 797-811
- Bezdek J (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York
- Brunsdon C, Fotheringham S, Charlton M (1996): Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. In: *Geographical Analysis*, 28, S. 281-298
- Bourassa S, Cantoni E, Hoesli M (2007): Spatial Dependence, Housing Submarkets, and House Price Prediction. In: *The Journal of Real Estate Finance and Economics* 35, S. 143-160
- Diggle P, Ribeiro P (2007): *Model-Based Geostatistics*. New York: Springer
- Everitt B (1993): *Cluster Analysis*. Arnold, London
- Fahrmeir L, Kneib T, Lang S (2009): *Regression: Modelle, Methoden und Anwendungen*. Springer, Berlin
- Fotheringham S, Brundson C, Charlton M (2002): *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, Wiley
- Goldstein, H (2003): *Multilevel Statistical Models*. Arnold: London
- Han J, Kamber M (2006): *Data Mining. Concepts and Techniques*. Morgan Kaufmann, San Francisco
- Isaaks E, Srivastava M (1989): *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford
- Kaufman L, Rousseeuw P (1990): *Finding Groups in Data. An Introduction to Cluster Analysis*. Wiley, New York
- Lancaster K (1996): A New Approach to Consumer Theory. In: *Journal of Political Economy*, 74, S. 132-157
- LeSage J, Pace K (2009): *Introduction to Spatial Econometrics*. CRC Press: Boca Raton
- Leung Y (1987): On the Imprecision of Boundaries. In: *Geographical Analysis*, 19, S. 125-151

Leung Y, Mei C-L, Zhang W-X (2000): Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. In: Environment and Planning A, 32, S. 9-32.

Malpezzi S (2002): Hedonic Pricing Models: A Selective and Applied Review. URL <http://econpapers.repec.org/RePEc:wop:wisule:02-05>

McDonald J (1997): Fundamentals of Urban Economics. Prentice Hall: Upper Saddle River

Openshaw S, Rao L (1995): Algorithms for Reengineering 1991 Census Geography. In: Environment and Planning A, 27, S. 425-446

Reimann C, Filzmoser P, Garrett R, Dutter R (2008): Statistical Data Analysis Explained. Applied Environmental Statistics with R. Wiley, Chichester

Rosen S (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. In: Journal of Political Economy, 82, S. 34-55

Wood S (2006): Low-Rank Scale-invariant Tensor Product Smooths for Generalized Additive Mixed Models. In: Biometrics, 62, S. 1025-1036