Eine räumliche Alternative der Modellierung und Darstellung von Immobilienpreisen für Immobilienportale

Lasse SCHEELE und Harald SCHERNTHANNER
Universität Potsdam, Potsdam · harald.schernthanner@uni-potsdam.de

Zusammenfassung

Viele Immobilienportale nutzen seit Jahren die hedonische Regression zur Preismodellierung und verwenden deren nicht räumlichen Ergebnisse als Grundlage für eine suboptimale räumliche Darstellung von Immobilienpreisen. Die von Portalbetreibern seit mehr als einer Dekade gesammelten Immobilienpreisdaten werden für diverse Analyse- und Darstellungszwecke genutzt. Die zur Kartenerstellung und -darstellung verwendeten Methodiken geben jedoch nur sehr unvollständig Auskunft über die tatsächlichen Immobilienpreise. Ansätze der räumlichen Analyse werden von Portalbetreibern kaum genutzt. Der Beitrag gibt eine Übersicht über den Status quo der raumanalytischen Grundlage und der kartographischen Darstellung von Immobilienpreisen auf Immobilienportalen und untersucht die Anwendung von räumlich-statistischen Methoden zur Modellierung von Immobilienpreisen als Alternative zu den traditionell verwendeten nicht-räumlichen Verfahren. Es erfolgt im Beitrag eingangs eine Übersicht über den räumlich analytischen und kartographischen State of the Art großer Immobilienportale. Hierbei wird auf Mängel der verwendeten Analyse- und Darstellungsformen eingegangen. Im zweiten Teil des Beitrages werden geeignete Alternativen der Modellierung und Darstellung von Immobilienpreisen identifiziert. Hierzu zählen das Ordinary Kriging, das Kriging with External Drift und das Ordinary Co-Kriging. Diese drei Methoden werden exemplarisch mit einem Fokus auf die geeignetste Parameterwahl mit Mietpreisdaten des Portalbetreibers Immobilienscout24[1] anhand eines Untersuchungsgebiets in Berlin-Kreuzberg auf ihre Performanz hin untersucht und verglichen. Die Umsetzung erfolgt im Statistikprogramm R. Abschließend erfolgt eine Diskussion und eine Einschätzung der Potenziale der operationellen Nutzung der vorgeschlagenen Alternative.

1 Status quo der (räumlichen-) Analyse und Kartendarstellung von Immobilienpreisen

1.1 Übersicht über Methoden der Immobilienpreisanalyse und -darstellung von Immobilienportalen

In diesem Beitragsteil wird eine tabellarische Übersicht (siehe Tabelle 1) über den Status quo (räumlich-) statistischer Immobilienpreisanalyse und der kartenbasierten Darstellung

[1] Die Daten wurden zu Forschungszwecken innerhalb der Transparenzoffensive der Immobilienwirtschaft von Immobilienscout24 zur Verfügung gestellt. Immobilienscout24 hat mit ~ 1,5 Million Datenpunkten die Deutschlandweit größte Immobiliendatenbank für Wohnimmobilien.

Strobl, J., Blaschke, T., Griesebner, G. & Zagel, B. (Hrsg.) (2014): Angewandte Geoinformatik 2014.
© Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-543-0.
Dieser Beitrag ist ein Open-Access-Beitrag, der unter den Bedingungen und unter den Auflagen der Creative Commons Attribution Lizenz verteilt wird (http://creativecommons.org/licenses/by/3.0/).

von Immobilienpreisen auf den jeweils drei nach Webzugriffen (nicht nach Angeboten) größten Immobilienportalen in den USA, in Deutschland und dem nach Angeboten größten Schweizer Portal gegeben. Laut dem E-Business Guide *eBizMba* (EBIZMBA 2014), der den Traffic, der von *Alexa.com* statistisch erfassten Webseitenaufrufe, auswertet und kategorisiert, sind in folgender Reihenfolge: *Zillow.com, Trulia.com, homes.yahoo.com* die meistaufgerufenen Immobilienportale in den USA. In Deutschland sind laut *Alexa.com: Immobilienscout 24, Immowelt und Immonet* die am häufigst aufgerufenen Immobilienportale. Komplementär, aufgrund der detaillierten Beschreibung der verwendeten hedonischen Regression, wurde das größte Schweizer Portal *Homegate* in die Darstellung mitaufgenommen. Als Standardmethode zur Preisanalyse verwenden die meisten Portalbetreiber die hedonische Regression. Informationen über die von Portalbetreibern zur Preismodellierung verwendeten Analysemethoden zu bekommen, erweist sich als schwierig. Beispielhaft für andere Immobilienportale ist das Schweizer Portal *Homegate* zu nennen, das sehr ausführlich über die verwendete Methode der statistischen Preismodellierung informiert. *Homegate* misst in Zusammenarbeit mit der Zürcher Kantonalbank monatlich die Änderungen der Schweizer Immobilienpreise und bildet einen sogenannten Angebotsmietindex über die Methode der hedonische Regression (HOMEGATE 2013).

Tabelle 1: Übersicht über Methoden der Räumlichen Statisik und Kartendarstellung

Immobilienportal	(Räumliche) Statistik	Kartendarstellung von Immobilienpreisen
Zillow.com	Hedonisches Regressionsmodell: Zillow Price Estimate	Pin Map der Angebotspreise. Listendarstellung geschätzter Preise
Trulia.com	Hedonisches Regressionsmodell: Monitoring von Miet- und Kaufpreisen	Pin map von Miet- und Kaufpreisen
Homes.yahoo.com	Keine Angaben	Pin map von Miet- und Kaufpreisen
Immobilienscout24.de	Hedonisches Regressionsmodell: IMX	Mapping geschätzter Miet- und Kaupreise gemappt auf willkürliche geographischen Bezugseinheit
Immowelt.de	Keine Angaben	Pin Map
Immonet.de	Keine Angaben	Keine Kartenansicht
Homegate.ch	Hedonisches Regressionsmodell: Mietindex	Pin Map

1.2 Hedonische Regression als nicht räumliche Standardmethode der Immobilienpreismodellierung

Die hedonische Regression zur Modellierung von Immobilienpreisen, ist bei Portalbetreibern die am weitesten verbreitete Methode. Es erfolgt eine Preisschätzung von Produkten mit laufender Veränderung des Preis-Leistungsverhältnis, bzw. einer Preisberechnung über Qualitätsanpassung (GORDON 1990). Umgemünzt auf den Immobiliensektor erfolgt die Preisermittlung einerseits über die Verfügbarkeit von Immobilien, andererseits über die Qualität der Immobilie. Der Ansatz der hedonischen Regression vernachlässigt räumliche Autokorrelation, die direkte Beeinflussung von Objekten durch benachbarte Objekte wird in der hedonischen Regression nicht berücksichtigt.

1.3 Status quo der Kartendarstellung von Immobilienangeboten und Immobilienpreisanalysen

Die Darstellung von Immobilienangeboten erfolgt in den meisten Portalen über eine Verortung als Positionssignatur bzw. als „Pin map“. Hierbei wird häufig das Symbol eines Hauses gewählt. Signaturen werden meist nicht maßstabsabhängig skaliert, was zu einer Überdeckung bzw. Überlappung von Signaturen führt. Teils werden beim Wechsel von größeren in kleinere Maßstabsebenen die Anzahl der aggregierten Angebote an einer Position in eine geometrische Signatur geschrieben. Eine weitere oft gewählte Form der Darstellung erfolgt über die Angabe des Preises als absoluter Zahlenwert (Zahlenwertsignatur) innerhalb geometrischer Formen, wie z. B. innerhalb eines Rechtecks. Auch hierbei treten häufig Probleme mit überlappenden Signaturen auf.

Ergebnisse von Immobilienpreismodellierungen werden in seltenen Fällen in Kartenform dargestellt. Oft wird ein Durchschnittspreis als Choroplethenkarte auf eine Bezugsgeometrie, wie z. B. eine administrative Einheit „gemappt“. Die tatsächlichen räumlichen Verhältnisse von Immobilienpreisen werden nicht dargestellt.

2 Räumlich-statistische Alternativen

2.1 Methodenauswahl und -beschreibung

Im zweiten Teil des Beitrages werden geeignete Alternativen der Modellierung und Darstellung von Immobilienpreisen identifiziert. Es erfolgte eine Literaturrecherche in wissenschaftlichen Untersuchungen zur Modellierung von Immobilienpreisen unter Berücksichtigung räumlicher Autokorrelation (vgl. CHICA-OLMO 2007, BOURASSA et al. 2010, TSUTSUMI et al. 2011 und WONG et al. 2013). Zu den von den Autoren als am geeignetsten identifizierten Methoden der Preisermittlung gehören das Ordinary Kriging (OK), das Kriging with External Drift (KED) und das Ordinary Co-Kriging (OCK), welche bei räumlich korrelierten Daten eine koordinatengenaue Wertmodellierung erlauben. In dieser Arbeit werden diese Verfahren als Alternative zur hedonischen Regression vorgestellt und evaluiert.

Hierfür wurden 637 Anzeigendaten für Mietwohnungen des Immobilienportals ImmobilienScout24 für ein 4,5 km² großes Gebiet in dem Berliner Ortsteil Kreuzberg aus dem dritten Quartal 2007 ausgewählt. Für diese Anzeigen und die zugehörige Kaltmiete pro Quadratmeter konnten zum Signifikanzwert $p > 0{,}05$ signifikante Moran’s *I*- und Geary’s *c*-Werte von 0,67 und 0,35 ermittelt werden, was auf eine positive räumliche Autokorrelation schließen lässt. Die genannten Kriging-Varianten können also zur Modellierung des Anzeigenpreises genutzt werden.

Beim Kriging wird der Wert an einem Punkt modelliert, indem die räumliche Konstellation der umgebenden Beobachtungspunkte und die zugehörigen Werte mittels eines theoretischen Variogramms in Bezug gebracht werden (eine genaue Beschreibung der Verfahren findet sich z. B. bei LI & HEAP 2008). Das Variogramm modelliert die Semivarianz in Abhängigkeit der Entfernung und wird durch die Wahl einer Modellfunktion und den zugehörigen Modellparametern Nugget, Partial Sill und Range definiert. Es gilt also, die räumliche

Struktur der Datengrundlage mittels einer geeigneten Parameterwahl abzubilden. In dieser Arbeit wird daher die Modellqualität der betrachteten Methoden für verschiedene Parameterkonstellationen untersucht.

Beim OK setzt sich der zu modellierende Wert aus einer Trend- und einer Residuenkomponente zusammen und wird lediglich mittels Informationen zu der Primärvariable berechnet. Als Trend wird hier ein lokaler Mittelwert verwendet, während sich der Residuenanteil aus den räumlich-gewichteten Abweichungen der benachbarten Beobachtungspunkte von diesem lokalen Mittelwert ergibt. Die Modellierung mittels KED verläuft ähnlich, wobei die Trendkomponente mittels einer linearen Regression aus am Modellierungspunkt zur Verfügung stehenden Sekundärvariablen auf der Grundlage der Sekundärinformationen an den Beobachtungspunkten berechnet wird. Hierfür wurden entsprechend vergleichbarer Studien Informationen zur Wohnfläche in Quadratmeter, zum Alter der Immobilie, zur Zimmeranzahl, zur Etage und zur Art der Heizung verwendet (vgl. CHICA-OLMO 2007, Bourassa et al. 2010 und WONG et al. 2013). Mit dem OCK werden ebenfalls die genannten Sekundärvariablen verwendet, wobei hierbei auch deren räumliche Autokorrelation und die räumliche Korrelation zwischen den Primär- und Sekundärvariablen genutzt wird. Es werden also Variogramme und Kreuz-Variogramme für alle vorliegenden Variablen erstellt. Beim OCK müssen entsprechend, anders als beim KED, die Sekundärinformationen nicht an den Modellierungspunkten vorliegen.

2.2 Modellevaluierung und -vergleich

Um die Modellqualität in Abhängigkeit der Modellparameter zu evaluieren und die verschiedenen Verfahren miteinander vergleichen zu können, wurden Wertebereiche für die genannten Modellparameter ausgewählt und die Modellqualität für alle Parameterkombinationen mittels einer 5-Fold-Cross Validation erfasst. Als Modellfunktion wurden die Sphärische, Gaussche und Exponentielle genutzt. Zur Evaluierung des OCKs wurde lediglich die Modellfunktion und der Range-Wert vorgegeben. Die Werte der übrigen Parameter Nugget und Partial Sill wurden iterativ mittels einer gewichteten Kleinste-Quadrate-Schätzung berechnet. Dadurch ergeben sich für OK und KED jeweils 5130 verschiedene Modellläufe und für OCK lediglich 45. Durch die Umsetzung einer Cross Validation konnten beobachtete und modellierte Anzeigenpreise verglichen werden, wobei für die Bewertung der verschiedenen Verfahren für der Mean Error (ME), der Root Mean Square Error (RMSE), die mittlere relative absolute Abweichung und der Anteil mit einer Abweichung von weniger als 10 % analysiert wurden. Zudem wurde die Rechendauer der Methoden erfasst und verglichen. Die Auswertung hat gezeigt, dass in der Regel für jede Modellfunktion und Datendichte ein methodenunabhängiger optimaler Range-Wert existiert, bei dem die besten Ergebnisse erzielt werden. Dieser ist bei sphärischen Modellen am größten und nimmt bei einem Übergang zu exponentiellen und gausschen Modellen jeweils ab. Ein weiterer wichtiger Aspekt ist die gegenseitige Beeinflussung der übrigen Modellparameter. So sollte eine Erhöhung des Nuggets ein Anheben des Partial Sills und eine Verkürzung der Range bedingen.

Der Vergleich der räumlich statistischen Verfahren untereinander mit einer hedonischen Regression hat zunächst gezeigt, dass alle Kriging-Varianten der nicht-räumlichen Regression bezüglich der Modellierungsgenauigkeit überlegen sind, auch wenn die Regression das schnellste Verfahren ist (vgl. Abb. 1). Darüber hinaus besitzt das KED die größte Performanz, mit ähnlichen guten Ergebnissen beim OK. Das OCK fällt besonders durch die langen Laufzeiten negativ auf.

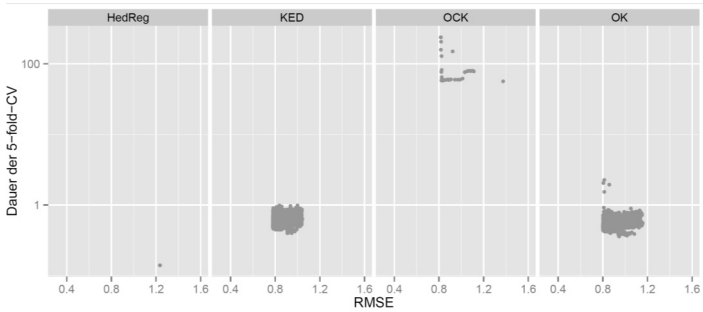

Abb. 1: Modellvergleich bezüglich der Rechendauer und des RMSE

2.3 Räumlich statistische Immobilienpreiskarte

Die räumlich statistischen Modellierungsverfahren lassen sich für eine hausgenaue Schätzung des Angebotspreises von Immobilien nutzen. Hierfür zeigt Abbildung 2 eine Beispielkarte, die mittels der über OpenStreetMap verfügbaren Gebäudegeometrien erzeugt wurde. Dabei wurde der Anzeigenpreis für jeden Gebäudemittelpunkt modelliert in das jeweilige Polygon geschrieben. Die benötigten Sekundärinformationen wurden aus den vorliegenden Daten ermittelt. Die resultierende Standardmietwohnung hat eine Wohnfläche von 77,1 m², ein Alter von 46,8 Jahren, besitzt 2,6 Zimmer, liegt in der Etage 2,5 und besitzt eine Zentralheizung.

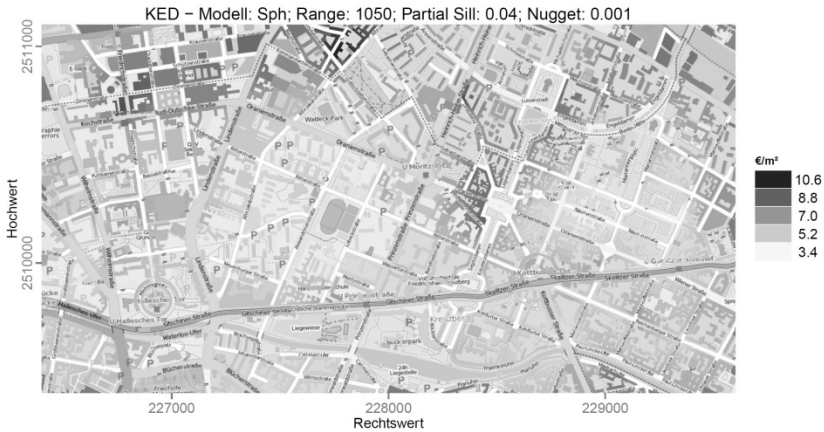

Abb. 2: Immobilienpreiskarte auf Grundlage des nach RMSE besten KED-Modells

3 Ergebnisse

Im operationellen Umfeld werden räumlich-statistische Analysemethoden von Portalbetreibern nicht genutzt, die kartographische Darstellung von Immobilienangeboten ist als suboptimal einzustufen. Keines der im Rahmen des Beitrags analysierten Immobilienportale nutzt

operationell Methoden der räumlichen Statistik, eine Kartendarstellung einer räumlich-statistischen Modellierung existiert bisher noch nicht. Die im Beitrag vorgeschlagene räumlich-statistische Alternative versteht sich als Startpunkt räumlich-statistische und kartographische Methoden aus der Wissenschaftsdomäne in das operationelle Umfeld von Portalbetreibern zu überführen. Diesbezüglich sind von den Autoren weitere Aktivitäten geplant.

Literatur

ALEXA (2014), Top Sites in Germany. http://www.alexa.com/topsites/countries/DE (17.04.2014).

BOURASSA, S. C., CANTONI, E. & HOESLI, M. (2010), Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods. Journal of Real Estate Research, 32 (2), 139-160.

CHICA-OLMO, J. (2007), Prediction of housing location price by a multivariate spatial method: cokriging. Journal of Real Estate Research, 29 (1), 95-114.

COURT, A. T. (1939), 'Hedonic price indexes with automobile examples'. The dynamics of the automobile demand, 99-117.

EBIZMBA, THE EBUSINESSGUIDE (2014), The 15 Most Popular Real Estate Websites. http://www.ebizmba.com/articles/real-estate-websites (17.04.2014).

GORDON, R. J. (1990), 'The Measurement of Durable Goods Prices', National Bureau of Economic Research Monograph, 1. Aufl., University of Chicago Press, Chicago.

GRILICHES, Z. (1961), 'Hedonic Price Indexes for Automobiles: An Econometric of QualityChange', in 'The Price Statistics of the Federal Goverment', National Bureau of Economic Research, General Series, 73, 173-196.

HOMEGATE (2014), Homegate Angebotsmietindex. bit.ly/1hPh2fd (17.04.2014).

IMMOBILIENSCOUT24 (2014), Immobilienindex IMX. http://www.immobilienscout24.de/immobilienbewertung/immobilienindex.html (17.4.2014)

IMMONET (2014), http://www.immonet.de/ (17.04.2014).

IMMOWELT (2014), http://www.immowelt.de/ (17.04.2014).

LI, J. & HEAP, A. D. (2008), A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, Record 2008/23.

TRULIA (2014), Trulia price and rent monitor. http://info.trulia.com/trulia-price-and-rent-monitor (17.04.2014).

TSUTSUMI, M., SHIMADA, A. & MURAKAMI, D. (2011), Land price maps of Tokyo Metropolitan Area. Procedia – Social and Behavioral Sciences, 21, 193-202.

WONG, S. K., YIU, C. Y. & CHAU, K. W. (2013), Trading volume-induced spatial autocorrelation in real estate prices. The Journal of Real Estate Finance and Economics, 46 (4), 596-608.

YAHOO HOMES (2014), https://homes.yahoo.com/ (17.04.2014).

ZILLOW (2014), What´s the Zillow Home Value Index? http://www.zillow.com/WIKIPAGES/What's-the-Zillow-Home-Value-Index/ (17.04.2014).