



Prof. Dr. Ralf Guckel

studierte Mathematik an der Universität Leipzig. Im Anschluss an seine Promotion im Jahr 2001 arbeitete er zunächst als Dozent und Projektentwickler für Bildungs- und Beratungsunternehmen. Seit dem Jahr 2011 ist er als Dozent für Mathematik und Statistik an der Berufsakademie Sachsen, Staatliche Studienakademie Leipzig, tätig, seit 2017 als Professor. Interessengebiete seiner wissenschaftlichen Tätigkeit sind freie Randwertprobleme für partielle Differentialgleichungen, die inferenzstatistische Interpretation von Konzentrationsmaßen und die statistische Analyse von Mietpreisdaten.

KONTAKT: Berufsakademie Sachsen / Staatliche Studienakademie Leipzig
Schönauer Straße 113a / 04207 Leipzig / ralf.guckel@ba-leipzig.de

Methodik der preisstatistischen Auswertung einer Angebotsdatenbank für Mietwohnungen

Ralf Guckel

Seit dem Jahr 2016 erscheint jährlich der „PISA-Marktbericht“, der den Zustand des Leipziger Wohnungsmarktes dokumentiert. Es handelt sich um ein Gemeinschaftsprojekt der PISA Immobilienmanagement GmbH & Co. KG und der Staatlichen Studienakademie Leipzig. Die Analyse des Mietwohnungsmarktes beruht auf einer statistischen Auswertung der Angebotsdatenbank von Immobilienscout24. Der vorliegende Text beschreibt die hierfür verwendete Methodik. Die Daten kommen durch eine Sekundärerhebung zustande und weisen zahlreiche Qualitätsmängel auf. Dies erfordert verschiedene Vorverarbeitungsschritte sowie Kompromisse bei der Durchführung der Preisvergleiche.

Since 2016, the „PISA-Marktbericht“, which documents the state of the Leipzig housing market, is published annually. It is a joint project of the PISA Immobilienmanagement GmbH & Co. KG and the Staatliche Studienakademie Leipzig. The analysis of the rental housing market is based on a statistical evaluation of the offer database of Immobilienscout24. The present text describes the methodology used for this purpose. The data come from a secondary survey and have many quality defects. This requires various preprocessing steps as well as compromises in making the price comparisons.

Problem- und Zielstellung

Die Erhebung, Analyse und Interpretation von Immobilienmarktdaten ist für verschiedene Interessengruppen von erheblicher Bedeutung. Im Mittelpunkt der Aufmerksamkeit stehen oftmals Preisvergleiche. Ein wesentlicher Aspekt ist hierbei die Heterogenität von Immobilienbeständen in Bezug auf nutzwertbestimmende Merkmale wie z. B. Größe, Lage, Ausstattung, Zustand usw., die teilweise nur schwer operationalisierbar sind. Dies führt bei der Durchführung von Preisvergleichen zu einer Reihe ernsthafter methodischer Schwierigkeiten. Die meisten dieser Probleme treten bei der Analyse von Mietpreisen in besonderem Maße in Erscheinung, und zwar aufgrund der im Vergleich zu Kaufpreisen schwerer kontrollierbaren Datenqualität. Insbesondere die Herausforderungen regionaler und örtlicher Mietpreisstatistiken unter Einschluss kleinräumiger Vergleiche stel-

len daher einen aktuellen Forschungsgegenstand der Wirtschafts- und Sozialstatistik dar.

Aus methodischer Sicht genießt der im Rahmen der Verbraucherpreismessung ermittelte Index der Nettokaltmiete die größte Autorität unter den Mietpreisstatistiken, vgl. [2]. Die örtlichen Mietspiegel gemäß § 558 BGB – vgl. z. B. [11], [12] – besitzen dagegen aufgrund fehlender Standards bereits eine deutlich geringere Autorität. Generell ist der Nutzen beider Statistiken für Markteinschätzungen begrenzt, weil die erfassten Mietpreise gerade nicht den Marktpreisen entsprechen und auch kein direkter Bezug zu Angebot oder Nachfrage besteht. Aus diesem Grund sind insbesondere in den letzten Jahren zahlreiche weitere Statistiken entstanden, begünstigt auch durch die zunehmende Digitalisierung und Vernetzung der Immobilienbranche. Bislang hat sich jedoch hinsichtlich der Methodik der Marktbeobachtung keine einheitliche, allgemein anerkannte Heran-

gehensweise durchgesetzt. Eine gute Übersicht über die verschiedenen methodischen Ansätze und Kritikpunkte vermitteln z. B. das Lehrbuchkapitel [7, 816 ff.] sowie das u. a. von Peter von der Lippe verfasste Gutachten [14].

Im Jahr 2016 stellte die PISA Immobilienmanagement GmbH & Co. KG (kurz: PISA) erstmals ihren Bericht zum Leipziger Wohnungsmarkt vor, vgl. [4]. Selbst gestelltes Ziel war u. a. die Sichtbarmachung des örtlichen Mietpreisniveaus für verschiedene Klassen von Wohnimmobilien einschließlich kleinräumiger Disparitäten. Für die Konzeption und Erstellung des Folgeberichtes wurde die Staatliche Studienakademie Leipzig um Unterstützung gebeten. Zukünftig ist eine periodisch wiederkehrende gemeinsame Veröffentlichung geplant. Die seitens PISA in eine Kooperation gesetzten Erwartungen lassen sich wie folgt zusammenfassen:

- I: wissenschaftliche Autorisierung der Methodik,
- II: Verringerung des für die statistische Auswertung erforderlichen Aufwands,
- III: Gewährleistung größtmöglicher Kontinuität zum Vorjahresbericht.

Ein Vergleich des PISA-Marktberichtes 2016 mit den eingangs skizzierten methodischen Problemen macht deutlich, dass ein Zielkonflikt vor allem zwischen den Erwartungen I und II sowie zwischen I und III besteht. Vor diesem Hintergrund muss der PISA-Marktbericht 2017 – vgl. [5] – zwangsläufig als Kompromisslösung verstanden werden.

Zielstellung der vorliegenden Arbeit ist die Beschreibung der für den PISA-Marktbericht 2017 verwendeten Methodik. Abschließend erfolgt eine kritische Zusammenfassung des Vorgehens. Die vorgestellten Auswertungen (vgl. Abbildungen 1, 2 und 3) tragen exemplarischen Charakter. Auf eine Interpretation wird daher bewusst verzichtet.

Konzeptionelle Vorüberlegungen

Im PISA-Marktbericht 2016 wurde die Angebotsdatenbank der Online-Plattform Immobilienscout24 als Datenbasis verwendet. Die Frage nach einer Alternative stellte sich zumindest kurzfristig nicht, so dass auch für den Bericht des Jahres 2017 auf diese Quelle zurückgegriffen wurde. Es handelt sich also grundsätzlich um die Analyse der über einen Angebotskanal bzw. ein Marktfenster veröffentlichten Mietangebote. Unklar bleibt dabei, welcher Anteil der angebotenen Wohnungen tatsächlich zu den veröffentlichten Konditionen vermietet werden konnte, und wann dies ggf. geschah. Im Unterschied zum Vorjahresbericht erfolgte nunmehr eine Auswertung aller Angebote der Jahre 2014, 2015 und 2016, um wenigstens kurzfristige Veränderungen abbilden zu können.

Von den 79 von Immobilienscout24 zur Verfügung gestellten Merkmalen kommen grundsätzlich nur diejenigen für eine Auswertung in Frage, die eine möglichst eindeutige Identifizierung und Interpretation

der Mietwohnungsangebote gestatten. Für die Nichtberücksichtigung eines Merkmals kommen folgende Gründe in Betracht:

- Das Merkmal ist redundant, weil sich die Merkmalswerte eindeutig aus den Werten anderer Merkmale rekonstruieren lassen. Beispiel: Das Merkmal `sc_id` ist zusammengesetzt aus `skid_id` (Stadt/Kreis) und `bgid` (Ortsteil).
- Das Merkmal ist sachlich irrelevant. Beispiel: Das Merkmal `edatum` enthält den Zeitpunkt der Erstellung des betreffenden Datensatzes in der Datenbank.
- Die Merkmalswerte sind Freitexte, für die der Auswertungsaufwand mittels semantischer Normalisierung in keinem sinnvollen Verhältnis zum Nutzen steht. Beispiel: Im Merkmal `mietekauti` werden Höhe und ggf. Zahlungsmodalitäten der Mietkaution erfasst.
- Die Merkmalswerte sind nicht eindeutig definiert und somit subjektiv interpretierbar. Beispiel: Das Merkmal `aqid` beschreibt die Qualität der Ausstattung durch die nicht näher spezifizierten, ordinalen Ausprägungen 1, ..., 5 (und NULL). Die Zuordnung der Werte durch den Anbieter ist willkürlich, zumal es sich hierbei um die aggregierte Beurteilung einer im Grunde sehr komplexen Eigenschaft handelt.
- Das Merkmal ist in der Datenbasis indifferent oder nahezu indifferent verteilt. Beispiel: Das dichotome Merkmal `foerderung` beschreibt, ob es sich bei dem betreffenden Objekt um geförderten Wohnraum handelt.
- Das Merkmal besitzt sehr viele fehlende Werte, die nicht sinnvoll mittels Imputation ergänzt werden können. Beispiel: Das Merkmal `schlafzimmer` enthält die Anzahl der Schlafzimmer mit einer Fehlquote von 64%.

Aus den genannten Gründen wurden vor Beginn der eigentlichen Auswertung insgesamt 58 der ursprünglich 79 Merkmale für eine spätere Nichtberücksichtigung gekennzeichnet. Nach einigen naheliegenden Umbenennungen verblieben die folgenden 21 auswertbaren Merkmale:

```

Primärschlüssel:      obid
Mietpreis [€/m²]:    nettokaltmiete
Zeitliche Einordnung: zeitraum laufzeittage
Lage:                 ortsteil postleitzahl strasse
                       hausnummer etage
Ausstattung, Qualität: wohnflaeche zimmeranzahl
                       wohnungstyp objektzustand
                       aufzug einbaukueche
                       gaestewc balkon garten
                       parkplatz parkplatzpreis
Nachfrage, Interesse: liste_match
  
```

Die inhaltliche Bedeutung der meisten dieser Merkmale ist offensichtlich. Zwar gehen die wenigsten dieser Merkmale explizit in die

statistische Endauswertung des Marktberichtes ein. Sie erweisen sich jedoch zum Teil bei der Konsolidierung der Datenbasis und bei der Bereinigung um Ausreißer, Inkonsistenzen und Redundanzen als nützlich. Sie stehen darüber hinaus für zusätzliche statistische Auswertungen des Kooperationspartners zur Verfügung.

Das Merkmal `objektzustand` ist – ähnlich wie das Merkmal `aqid` – durch subjektiv interpretierbare, diesmal jedoch nur nominal skalierte Merkmalswerte gekennzeichnet. Für die Auswertung wurde daher eine Dichotomisierung mit den Ausprägungen `erstbezug` (nach Neubau oder Sanierung) und `bestand` vorgenommen.

Die folgenden vom Datenanbieter zur Verfügung gestellten Merkmale sind grundsätzlich für eine Analyse der Nachfragedisparität geeignet:

Merkmal	Bedeutung
<code>hits</code>	Anzahl der Exposéansichten
<code>click_schnellkontakte</code>	Anzahl der Schnellkontakte
<code>click_customer</code>	Anzahl der E-Mails an den Anbieter
<code>click_weitersagen</code>	Anzahl der E-Mails an Dritte
<code>click_url</code>	Anzahl der Klicks zur Homepage des Anbieters
<code>liste_show</code>	Anzahl, wie oft das Objekt auf einer aufgerufenen Seite der Ergebnisliste enthalten war
<code>liste_match</code>	Anzahl, wie oft das Objekt auf eine Suchanfrage gepasst hat

Die Merkmale `click_*` erweisen sich jedoch als nahezu indifferent, die damit verbundenen Aktionen werden von den Datenbankanwendern offenbar kaum genutzt. Die Werte der Merkmale `hits` und `liste_show` hängen davon ab, welche Umfänge und welche Sortierreihenfolgen die Ergebnislisten besitzen, in denen das betreffende Angebot enthalten ist. Eine objektive Messung der Nachfrage ist damit nicht möglich. Das Merkmal `liste_match` beschreibt in gewissem Sinn die Anzahl der (passiven) Nachfragen nach dem betreffenden Objekt. Diese Interpretation erlaubt es, Verteilungen dieser (passiven) Nachfrage zu ermitteln und mit den Verteilungen des Angebots zu vergleichen. Es zeigt sich, dass sich die auf diesem Weg ermittelten Nachfrage-Verteilungen hoch signifikant von den Verteilungen des Angebots unterscheiden (χ^2 -Homogenitätstests zum Signifikanzniveau $\alpha=0,001$). Trotzdem ist die Aussagekraft der auf diese Weise generierten Nachfrage-Verteilungen und Nachfrage-Parameter begrenzt, denn sie bilden letztlich nur ein recht unspezifisches Interesse der Datenbanknutzer ab. Gleichwohl wurde das Merkmal `liste_match` in die Liste der auszuwertenden Merkmale aufgenommen.

Konsolidierung und Vorverarbeitung

Die zur Analyse vorgesehenen Rohdaten wurden von PISA in Form von zwei Excel-Dateien mit folgenden Parametern bereitgestellt:

Datei	Zeitraum	Datensätze	Merkmale
I	01/2014 - 06/2015	88.859	79
II	07/2015 - 12/2016	74.640	79

Es handelt sich hierbei um Sekundärerhebungen, deren Repräsentativität für die Grundgesamtheit aller Mietwohnungsangebote intransparent bleibt. Ob bei der Erstellung der Datenbankabbilder Transformations- oder Aggregationsprozesse durchlaufen wurden, bleibt ebenfalls offen.

Die Konsolidierung der Mietangebote wurde durch die ungewöhnliche Aufteilung der Rohdaten auf die Dateien I und II erschwert. Das Merkmal `zeitraum` (mit Textwerten im Format JJJJMM bis JJJJMM) beschreibt den Zeitraum, innerhalb dessen das betreffende Objekt als Angebot auf dem Portal sichtbar war. Das Merkmal `laufzeittage` (mit ganzzahligen Werten) beschreibt die Anzahl der Tage, an denen das betreffende Angebot (innerhalb des zuvor genannten Zeitraums) sichtbar war. Aus dem Merkmal `zeitraum` lässt sich eine obere und eine untere Schranke für die maximale Anzahl der Laufzeittage berechnen. Es zeigt sich, dass die Werte des Merkmals `laufzeittage` in 98,35% der Fälle in dem berechneten Korridor liegen. In den verbleibenden 1,65% der Fälle ist die Anzahl der Laufzeittage kleiner, was sich als untere Schranke für den Anteil der nicht zusammenhängenden Angebotszeiträume interpretieren lässt. Eine genauere zeitliche Lokalisierung der Angebote ist nicht möglich. Vor diesem Hintergrund wurde eine ursprünglich ins Auge gefasste monatliche Auswertung wieder verworfen.

Die eigentliche Aufteilung der Mietangebote auf die Jahre 2014, 2015 und 2016 erfolgte durch Zusammenführen der Dateien I und II und Auswertung des Merkmals `zeitraum` mit Hilfe der Excel-Funktionen `RECHTS`, `LINKS` und `WERT`. Nach Entfernung der im vorangegangenen Abschnitt als nicht relevant qualifizierten Merkmale sowie der nunmehr nicht mehr benötigten Merkmale `zeitraum` und `laufzeittage` transformierte sich der Datenbestand schließlich auf die folgenden drei Dateien:

Datei	Zeitraum	Datensätze	Merkmale
III ₁	01/2014 - 12/2014	58.316	19
III ₂	01/2015 - 12/2015	66.212	19
III ₃	01/2016 - 12/2016	48.772	19

Innerhalb der Dateien I und II sind die einzelnen Datensätze durch den mutmaßlichen Primärschlüssel `obid` eindeutig bestimmt. Innerhalb der Dateien III_{1,2,3} ist dies jedoch nicht immer der Fall: Es treten 587 `obid`-Werte in Datei III₁ und 6170 Werte in Datei III₂ (sowie 0 Werte in Datei III₃) doppelt auf. Die dadurch repräsentierten Objekte sind in beiden Dateien I und II enthalten. In der Mehrzahl der Fälle (>6362) sind die Angebotszeiträume dabei nicht zusammenhängend. In sehr wenigen Fällen ist darüber hinaus zu erkennen, dass identische `obid`-Werte nicht notwendigerweise identische Lage-Merkmale der betreffenden Wohnungen implizieren. Es musste also prinzipiell davon ausgegangen werden, dass durch die Dynamik der Datenbanknutzung bereits vor der Erhebung in geringem Umfang Informationen verloren gegangen sind.

Weil für einen späteren Zeitpunkt ohnehin eine Bereinigung um redundante Mietangebote vorgesehen war, wurde in diesem Stadium auf eine Bereinigung verzichtet und lediglich das Merkmal `obid` aus der Datenbasis entfernt. Bei den verbleibenden 18 Merkmalen erfolgte zusätzlich eine Vereinheitlichung hinsichtlich der Wertekodierung und -Formatierung:

Lfd. Nr.	Merkmal	Format	Fehlquote	Imputation
1	<code>nettokaltemiete</code>	Zahl (0,00)	0,00%	
2	<code>ortsteil</code>	Text (Liste)	0,00%	
3	<code>postleitzahl</code>	00000	0,00%	
4	<code>strasse</code>	Text	0,03%	N
5	<code>hausnummer</code>	Text	2,03%	N
6	<code>etage</code>	Zahl (0)	16,58%	N
7	<code>wohnflaeche</code>	Zahl (0,00)	0,00%	
8	<code>zimmeranzahl</code>	Zahl (0,0)	0,00%	
9	<code>wohnungstyp</code>	Text (Liste)	0,00%	
10	<code>objektzustand</code>	dichotom (J/N)	0,00%	
11	<code>aufzug</code>	dichotom (J/N)	17,51%	J
12	<code>einbaukueche</code>	dichotom (J/N)	23,49%	J
13	<code>gaestewc</code>	dichotom (J/N)	12,09%	J
14	<code>balkon</code>	dichotom (J/N)	13,59%	J
15	<code>garten</code>	dichotom (J/N)	23,43%	J
16	<code>parkplatz</code>	dichotom (J/N)	95,51%	J
17	<code>parkplatzpreis</code>	Zahl (0,00)	86,10%	J
18	<code>liste_show</code>	Zahl (0)	0,00%	

Ein Blick auf die konsolidierten Daten offenbarte eine Vielzahl von Datenqualitätsmängeln. Insbesondere fehlende, falsche, kryptische und widersprüchliche Werte (vgl. [3, 41]) machen eine Vorverarbeitung zwingend erforderlich. Als besonders gravierend erweist sich das folgende Problem: Erfahrungsgemäß werden Mietwohnungen nicht selten von mehreren Anbietern am Markt platziert. Diese Mehrfachangebote sind zu bereinigen, um ein realistisches Abbild zu gewährleisten. Diese Bereinigung wird jedoch durch sehr viele falsche

Adressangaben erschwert. Selbst nach naheliegenden Vereinfachungen (Entfernung von „White Spaces“, Normierung von Umlauten etc.) stehen bspw. den offiziellen 3.032 Straßennamen der Stadt Leipzig immerhin noch 5.348 verschiedene Ausprägungen des Merkmals `strasse` gegenüber. Die Korrektur erfolgte durch eine Kombination verschiedener manueller sowie automatischer Normalisierungsschritte:

- 351 Datensätze enthalten eine nicht eindeutig der Stadt Leipzig zuzuordnende `postleitzahl`. Eine Sichtprobe machte deutlich, dass es sich hierbei in der überwiegenden Mehrzahl der Fälle um einfache Zahlendreher oder fälschlich als zu Leipzig gehörende Umlandgemeinden handelt. Aufgrund der insgesamt geringen Fallzahlen wurden die notwendigen Korrekturen manuell durchgeführt.
- Für die Korrektur der Straßennamen wurde ein auf der Jaccard-Metrik beruhendes Verfahren zum unscharfen Vergleich zweier Zeichenketten verwendet, vgl. z. B. [8, 15]. Das Excel Add-In „Fuzzy Lookup“ stellt eine Implementierung dieses Verfahrens zur Verfügung. Damit konnten die Ausprägungen der Datenbasis bis auf sehr wenige Ausnahmen automatisch durch die zugehörigen korrekten Einträge aus dem Straßenverzeichnis der Stadt Leipzig ersetzt werden. Die Qualitätssicherung bestand in einer manuellen Sichtung anhand einer Zufallsstichprobe.
- In einem letzten Schritt wurden die Datenbankeinträge anhand des Leipziger Straßenabschnittsverzeichnisses auf Inkonsistenzen hinsichtlich des Zusammenhangs der Merkmale `ortsteil`, `postleitzahl`, `strasse` und `hausnummer` geprüft.
- Die Entfernung der Mehrfachangebote erfolgte im Anschluss an diese Normalisierung durch eine Filterung über die Merkmale 1-8. Die Werte des Merkmals `liste_show` wurden dabei über alle Mehrfachangebote gemittelt. Es zeigte sich, dass dieser Bereinigungsprozess sehr stabil gegenüber einer Erweiterung der Filterung auf die dichotomen Qualitätsmerkmale (mit oder ohne Imputation) ist.

Nach Abschluss der Konsolidierung und Vorverarbeitung standen die folgenden Daten für die eigentliche Auswertung zur Verfügung:

Datei	Zeitraum	Datensätze	Merkmale
IV1	01/2014 - 12/2014	31.325	18
IV2	01/2015 - 12/2015	29.527	18
IV3	01/2016 - 12/2016	28.761	18

Der Datenbestand wurde durch die Bereinigungsprozesse also nahezu halbiert.

Auswertung

Hinter einer Preisstatistik steht der Wunsch, jeder geographischen Region und jedem Zeitraum ein Preisniveau zuzuordnen, um interlokale und intertemporale Preisvergleiche zu ermöglichen. Wünschenswert ist ein absolutes Preisniveau in Gestalt einer Maßzahl, deren Interpretation über einen Sachbezug erfolgt. Genau dies erweist sich jedoch im Fall heterogener Güter als sehr komplexes Problem. In der statistischen Praxis wird der Begriff Preisniveau daher in der Regel als relative Größe aufgefasst, indem

- entweder auf eine lediglich relative Interpretation der gewählten Maßzahl geachtet wird oder
- Preisindizes berechnet werden, vgl. z. B. [13].

Als primäres Preismerkmal für Mietwohnungen wird normalerweise die auf 1 m² Wohnfläche und einen Monat bezogene Nettokaltmiete verwendet. Im PISA-Marktbericht 2016 erfolgte eine Berechnung ungewichteter arithmetischer Mittelwerte dieses Merkmals. Das derart festgelegte Preisniveau besitzt jedoch keinerlei Sachbezug, da es sich bei Stückpreisen um Verhältniszahlen handelt. Im Fall von beobachteten Mietpreisen $p^{(1)}, \dots, p^{(n)}$ wäre allenfalls eine mit den zugehörigen Wohnflächen $F^{(1)}, \dots, F^{(n)}$ gewichtete Mittelwertbildung sachlich interpretierbar: Die Verhältniszahl

$$\bar{p} = \frac{\sum_i F^{(i)} p^{(i)}}{\sum_i F^{(i)}}$$

beschreibt den auf 1 m² Wohnfläche entfallenden Mietpreis für den gesamten Wohnungsbestand. Allerdings ist diese Mittelung sehr anfällig gegenüber „Ausreißern“ bei den beobachteten Preisen und Flächen, vor allem bei kleinen Datenbeständen. Außerdem stellt sich

dabei die Frage, weshalb ausgerechnet das wohnwertbestimmende Merkmal „Wohnfläche“ als Vergleichsmerkmal herangezogen wird. Verwendet man anstelle der Wohnfläche andere wohnwertbestimmende Merkmale, so wäre auch die Verwendung des Preismerkmals $p = P/F$ grundsätzlich in Frage zu stellen. Dies führt zum „stochastischen“ Zugang zu Preisvergleichen (vgl. [10]), bei dem zunächst das primäre Preismerkmal P als Regressionsfunktion der wohnwertbestimmenden Sekundärmerkmale S_1, \dots, S_m und einem Residuum R dargestellt wird: $P = F(S_1, \dots, S_m; R)$. Der eigentliche Preisvergleich besteht dann im Vergleich von Regressionsmodellen. Ist bspw. $P = p_1 S_1 + \dots + p_m S_m + R$, so lassen sich die Regressionsparameter p_1, \dots, p_m als die im jeweiligen Orts-, Zeit- und Sachbezug vorliegenden Preis-Bewertungen (im Sinne von Stückpreisen) der Sekundärmerkmale auffassen. Dies erfordert jedoch ein Verständnis des Zusammenhangs zwischen Mietpreis und wohnwertbestimmenden Merkmalen. Ein solcher Zugang kam jedoch wegen der fehlenden operationalisierten Lagemerkmale im vorliegenden Fall von vornherein nicht in Betracht.

Für den PISA-Marktbericht 2017 wurde stattdessen als Kompromisslösung der Median des Merkmals `nettokaltmiete` als relativ zu interpretierendes Preisniveau verwendet. Dieser Lageparameter ist stabil gegenüber Extremwerten. Es besteht nicht die Gefahr der unsachgemäßen Fehlinterpretation als „Mittelwert“. Da das Merkmal `nettokaltmiete` außerdem eine linkssteile Verteilung aufweist, scheint der Median auch aus statistischer Sicht besser zur Beschreibung eines „typischen“ Wertes der Verteilung geeignet zu sein. Als ergänzende Streuungsparameter wurden zudem die 5%-, 25%-, 75%- und 95%-Quantile berechnet und in Form einfacher Box-Plots visualisiert, vgl. Abbildung 1. (Um die Aussagekraft zu erhöhen, wurden in die Abbildungen 1-3 die Ergebnisse des Jahres 2017 mit aufgenommen.)

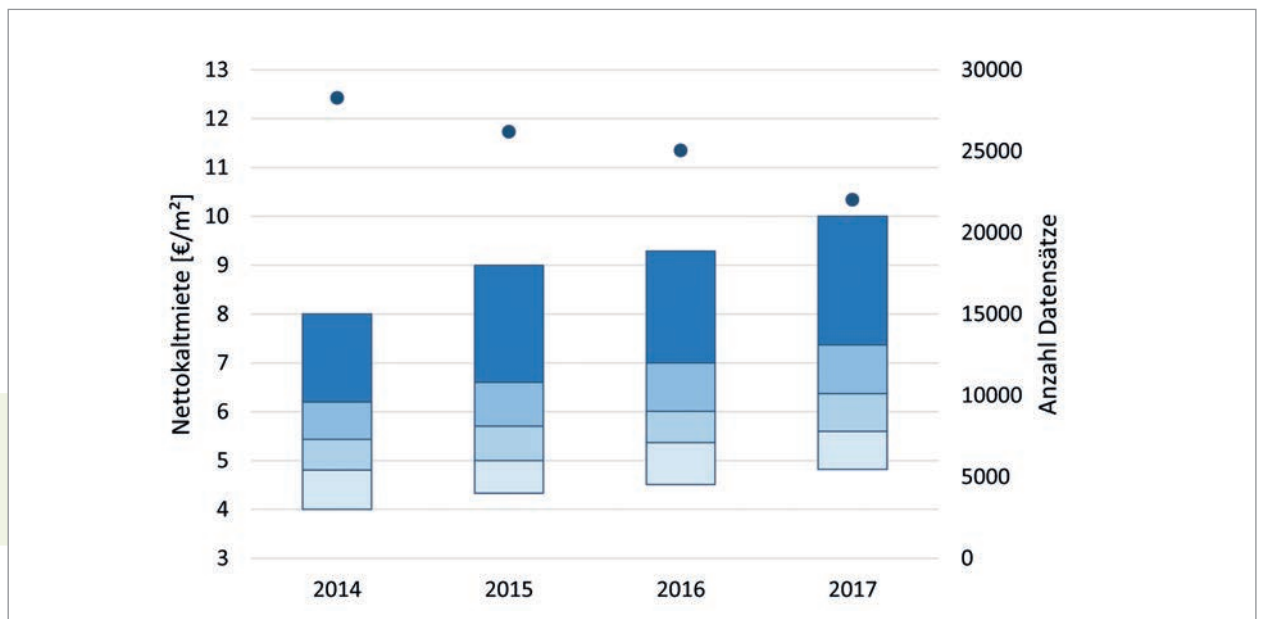


Abbildung 1: Intertemporaler Preisvergleich (Stadt Leipzig, Bestand)



Es wurde weder eine intertemporäre noch eine interlokale Qualitätsbereinigung durchgeführt. Zwar existiert eine Reihe (vorwiegend dichotomer bzw. diskreter) Ausstattungs- und Zustandsmerkmale. Operationalisierte Lagemerkmale waren jedoch kurzfristig nicht verfügbar. Die bei der Berechnung des Leipziger Mietspiegels (siehe [12]) verwendete Zurückführung der Lageeinschätzung auf die Bodenrichtwerte ist aus methodischer Sicht zu kritisieren: Die teilweise Erklärung des Mietpreises durch Lagemerkmale, die aus Kaufpreisen gewonnen werden, stellt letztlich einen methodischen Zirkelschluss dar. Die alternativ in Erwägung gezogene Lagecharakterisierung durch den Abstand zum Zentrum ist zwar grundsätzlich besser geeignet, sie wird aber der Komplexität des Merkmals „Lage“ nicht gerecht.

Neben den für den PISA-Marktbericht 2017 durchgeführten Preisvergleichen erfolgte darüber hinaus eine Reihe weiterer statistischer Auswertungen. Zwei ausgewählte Beispiele sollen im Folgenden kurz beschrieben werden.

Der nichtlineare statistische Zusammenhang zwischen den Merkmalen wohnfläche und nettokaltmiete ist bekannt, vgl. z. B. [1, 25ff.]. Die zeitliche Dynamik dieses Zusammenhangs wurde anhand der vorliegenden Daten mit Hilfe segmentierter Regressionsmodelle untersucht. Die Nettokaltmiete p (gemessen in €/m²) wird dabei als stetige, stückweise lineare Regressionsfunktion der Wohnfläche F aufgefasst: $p = a_i + b_i \cdot F + R$, $i = 1, 2$. Die Schätzung der Modellparameter a_i , b_i wurde unter Verwendung der Kleinste-Quadrate-Methode mit dem Excel-Add-In „Solver“ realisiert. Das Ergebnis ist Abbildung 2 zu entnehmen. Für jedes Jahr sind neben den Regressionsmodellen und Strukturbrüchen auch die Mediane der Wohnfläche und der Nettokaltmiete dargestellt. Auffällig ist die unterschiedliche Dynamik der Preisabhängigkeit bei den „kleinen“ bzw. „großen“ Wohnungen.

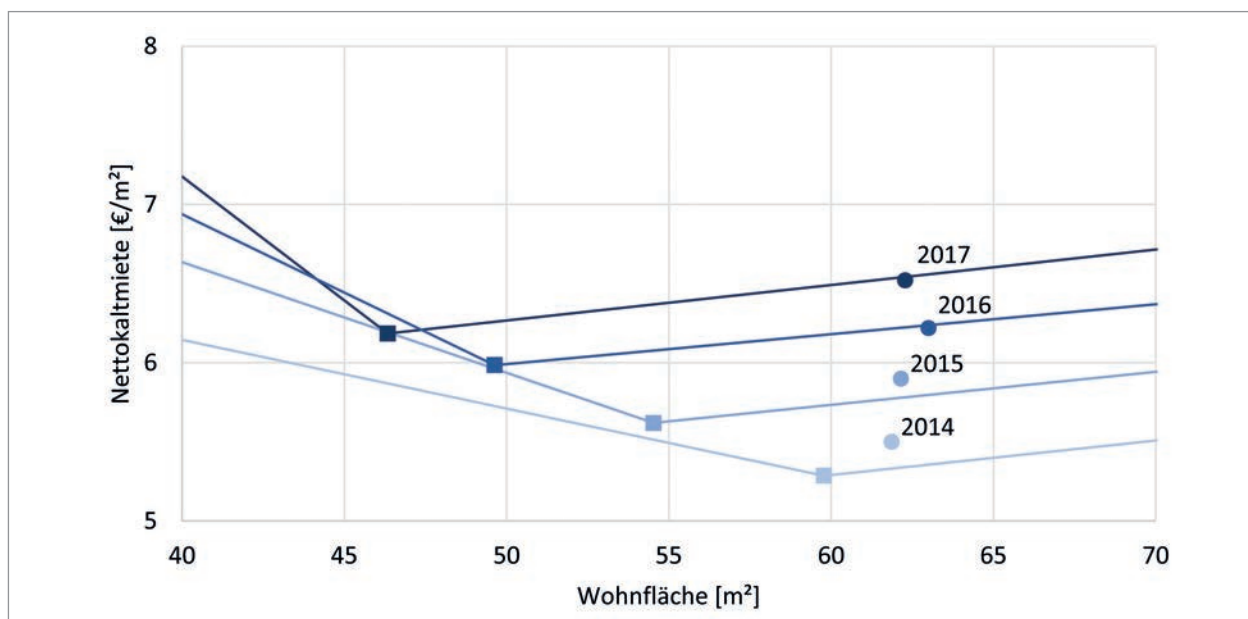


Abbildung 2: Zusammenhang zwischen Wohnfläche [m²] und Nettokaltmiete [€/m²]

Die eingangs erwähnte Möglichkeit, durch Auswertung des Merkmals `liste_match` das Interesse an den angebotenen Mietwohnungen abzubilden, wurde ebenfalls weiterverfolgt. Dazu wurden die mit den Werten dieses Merkmals gewichteten Mediane berechnet und den entsprechenden Angebotsparametern gegenübergestellt. Abbil-

dung 3 zeigt das Ergebnis für die Stadt Leipzig insgesamt.

Um die Frage beantworten zu können, ob diese Entwicklung, insbesondere ob das scheinbar konvergierende Trendlinienpaar, einen Aspekt des zugrundeliegenden Mietwohnungsmarktes abbildet, werden längerfristige Beobachtungen erforderlich sein.

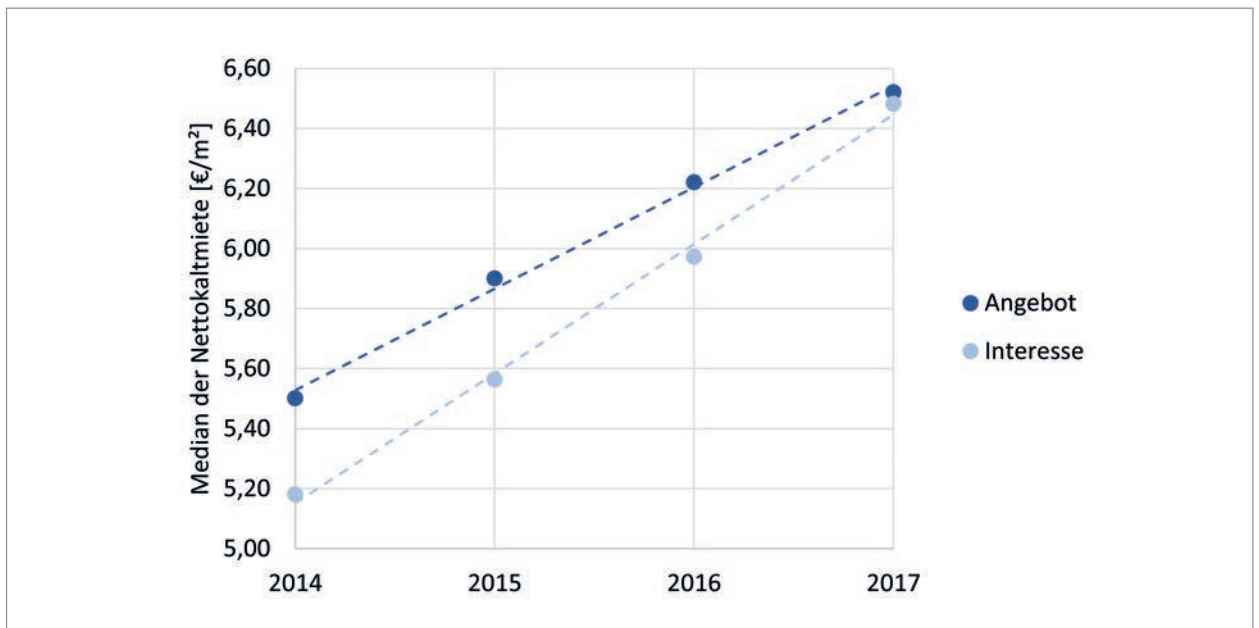


Abbildung 3: Vergleich der Preisentwicklung von Angebot und Interesse

Zusammenfassung

Die Ergebnisse des PISA-Marktberichtes 2017 erscheinen zwar plausibel und werden von den lokalen Marktakteuren akzeptiert. Allerdings sollten die (aus der Beschaffenheit der vorliegenden Daten sowie den Nebenbedingungen II und III resultierenden) methodischen Schwächen nicht außer Acht gelassen werden:

- Es erfolgt ausschließlich eine Analyse der Angebotspreise. Für eine verlässliche Markteinschätzung wäre darüber hinaus ein möglichst direkter Vergleich mit den Nachfrage- und Neuvertragspreisen erforderlich.
- Die Repräsentativität der ausgewerteten Stichprobe für die Grundgesamtheit aller Angebote bleibt intransparent.
- Als Vergleichswert wird der Median der auf 1 m² Wohnfläche bezogenen Nettokaltmiete gewählt. Er ist unter den gegebenen Umständen als sinnvollste Alternative unter den klassischen Lageparametern anzusehen. Diese Wahl ist jedoch rein statistisch, nicht jedoch sachlich motiviert. Außerdem findet keine hedonische Preisbereinigung statt.

Die beiden ersten Kritikpunkte lassen sich innerhalb des bestehenden konzeptionellen Rahmens nicht auflösen. Für die Umsetzung beider Aspekte des dritten Kritikpunktes – sachliche Begründung des Preisniveaus sowie hedonische Preisbereinigung – ist ein tieferes Verständnis des Zusammenhangs zwischen Mietpreis und wohnwertbestimmenden Merkmalen erforderlich. Insbesondere wären hierfür operationalisierte Lagemerkmale der betreffenden Mietwohnungen zu erfassen. Hierfür scheint der in [6] vorgeschlagene Ansatz vielversprechend zu sein. Die Grundidee besteht darin, aus den Adressdaten einer Wohnung objektive Lagemerkmale mit Hilfe frei zugänglicher GIS-Systeme (bspw. OpenStreetMap) zu generieren. Neben Umweltparametern (wie z. B. Luftverschmutzung, Lärmbelastung, Bebauungsdichte, Durchgrünungsdichte) kommen auch weitere Merkmale, wie z. B. der Abstand von oder die Wegzeit zu bestimmten Ankerpunkten (Zentrum, Erholungsgebiete, ÖPNV-Knotenpunkte) in Frage. Das Verfahren wurde bereits erfolgreich auf mehrere westdeutsche Großstädte angewendet und könnte nun für die Stadt Leipzig angepasst werden.

Literaturverzeichnis

- [1] Fahrmeir, L.; Kneib, T.; Lang, S. (2009): Regression. Springer-Verlag, Berlin, Heidelberg, 2. Aufl..
- [2] Goldhammer, B. (2016): Die neue Mietenstichprobe in der Verbraucherpreisstatistik. *Wirtschaft und Statistik* 5, 86-101.
- [3] Müller, R. M.; Lenz, H.-J. (2013): Business Intelligence. Springer-Verlag, Berlin, Heidelberg.
- [4] PISA Immobilienmanagement GmbH & Co. KG (Hrsg.): Marktbericht 2016.
- [5] PISA Immobilienmanagement GmbH & Co. KG (Hrsg.): Marktbericht 2017.
- [6] Promann, J. (2012): Die Berücksichtigung des Wohnwertmerkmals Lage in den Mietspiegeln der deutschen Großstädte. Josef Eul Verlag, Lohmar. (Diss., Universität Stuttgart, 2012)
- [7] Rottke, N. B.; Voigtländer, M. (Hrsg.) (2017): Immobilienwirtschaftslehre – Ökonomie. Springer Gabler, Wiesbaden.
- [8] Runkler, T. A. (2010): Data Mining. Springer Vieweg, Wiesbaden, 2. Aufl..
- [9] Schernthanner, H. (2017): Räumliche Analyse und Visualisierung von Mietpreisdaten. Springer Spektrum, Wiesbaden. (Diss., Universität Potsdam, 2016)
- [10] Selvanathan, E. A.; Prasada Rao, D. S. (1994): Index Numbers - A Stochastic Approach. Macmillan Press, Basingstoke, London.
- [11] Stadt Leipzig, Sozialamt (Hrsg.): Leipziger Mietspiegel 2016.
- [12] Stadt Leipzig, Sozialamt (Hrsg.): Leipziger Mietspiegel 2016 – Dokumentation.
- [13] Von der Lippe, P. (2007): Index Theory and Price Statistics. Peter Lang Internationaler Verlag der Wissenschaften, Frankfurt a. M..
- [14] Von der Lippe, P.; Breuer, C. C. (2010): Konzept für ein wirtschaftliches System periodischer regionaler Preisindizes – Möglichkeiten zur Gewinnung regionaler Daten über Mieten und Immobilienpreise. In: Möller, J.; Hohmann, E.; Huschka, D. (Hrsg.): Der weiße Fleck – zur Konzeption und Machbarkeit regionaler Preisindizes. IAB-Bibliothek 324. Institut für Arbeitsmarkt- und Berufsforschung. Nürnberg, 167-261.