

Dokumentation für den *Mietpreisindex*

Francisco Amaral, Philipp Jungwirth,
Jonas Zdrzalek and Steffen Zetzmann *

12.06.2025

*Amaral: Universität Zürich, Kiel Institut für Weltwirtschaft, francisco.amaral@df.uzh.ch; Jungwirth: Kiel Institut für Weltwirtschaft, philipp.jungwirth@ifw-kiel.de; Zetzmann: University of Mannheim, Kiel Institut für Weltwirtschaft, steffen.zetzmann@ifw-kiel.de; Zdrzalek: Kiel Institut für Weltwirtschaft, jonas.zdrzalek@ifw-kiel.de.

Inhaltsverzeichnis

1	Einleitung	3
2	VALUE Marktdaten	4
3	Index-Methodik	5
3.1	Variablen	7
3.2	Hedonische Preisindizes - Stadt	8
3.3	Deutschlandweiter Gesamtindex - Gewichtung	8
3.4	Inflationsbereinigung	9
3.5	Rolling-Windows	10
4	Summary Statistics	12

1 Einleitung

Die Erstellung eines neuen Mietpreisindex erfolgte durch eine Analyse einer Vielzahl wertvoller Informationen über Mietpreisangebote in 19 deutschen Städten und einem Landkreis. Im Interesse der Transparenz und zur Einhaltung strenger wissenschaftlicher Praktiken haben wir die Einzelheiten der Daten auf Mikroebene und die Indexmethodik für diese Indizes in diesem Dokument genauestens beschrieben.

Die herangezogenen Daten stammen von VALUE Marktdaten¹, welche Mietinserate von diversen Portalen sammeln und aufbereiten. Der Datensatz umfasst mehrere Millionen Mietinserate und reicht bis ins Jahr 2012 zurück. Wir berechnen mit diesen Daten nicht nur Indizes auf Städteebene, sondern auch einen gewichteten deutschlandweiten Index. Die Berechnung erfolgt auf monatlicher, quartalsweiser sowie jährlicher Frequenz. Mehr dazu finden Sie in der nachstehenden Ausführung der Methodik.

Ein Überblick über die Städte, für welche uns Daten zur Verfügung stehen finden Sie hier angeführt:

- | | | |
|--------------|---------------------|--------------------|
| • Berlin | • Erfurt | • München |
| • Bonn | • Frankfurt am Main | • Potsdam |
| • Chemnitz | • Hamburg | • Rhein-Erft-Kreis |
| • Dortmund | • Karlsruhe | • Stuttgart |
| • Dresden | • Köln | • Wiesbaden |
| • Duisburg | • Leipzig | |
| • Düsseldorf | • Lübeck | |

¹<https://www.value-marktdaten.de>; Persönlichen Dank an Sebastian Hein für die Kooperation.

2 VALUE Marktdaten

Wir bekommen Daten zu Wohnungsanzeigen von VALUE Marktdaten. Das Unternehmen sammelt und verarbeitet Anzeigen von Online-Plattformen und Immobilienanbietern. Dafür haben sie einen Algorithmus entwickelt, der zufällig Anzeigen von den Plattformen auswählt, um eine möglichst repräsentative Auswahl zu erstellen. So wird sichergestellt, dass Inserate erfasst werden, die nur eine kurze Zeit, aber auch Inserate die eine längere Zeit online sind. Dadurch wird sichergestellt, dass keine Verzerrung entsteht, die durch eine bestimmte chronologische Reihenfolge der Anzeigen entstehen könnte. In den Daten sehen wir, wann die Anzeigen veröffentlicht und entfernt wurden, Adressen (falls verfügbar), grobe Infos zur Lage wie Postleitzahl oder Stadtteil, die geforderten Mieten und verschiedene Merkmale des Objekts. Der Datensatz startet mit Januar 2012 wird laufend aktualisiert.

3 Index-Methodik

Der ideale Mietpreisindex würde den Mietpreisanstieg einer repräsentativen, unveränderten Immobilie erfassen, die in jedem Zeitraum vermietet wird. Leider sind Wohnungen oft sehr heterogen und einige werden über einen langen Zeitraum nicht neu vermietet. Dies bedeutet, dass sich die Stichprobe der vermieteten Immobilien von Periode zu Periode erheblich verändert. Um sich dem oben beschriebenen idealen Preisindex anzunähern, müssen wir daher die Heterogenität berücksichtigen. In der aktuellen Literatur werden zwei Hauptmethoden zur Erreichung dieses Ziels identifiziert (Balk et al., 2014). Die erste ist die so genannte 'repeat-sales'-Methode. Der Grundgedanke dieser Methode besteht darin, die Preisentwicklung ein und derselben Immobilie im Laufe der Zeit zu nutzen, um einen Preisindex zu erstellen. Diese Methode ist natürlich auf einen sehr liquiden Immobilienmarkt beschränkt. Werden jedoch Wohnungen nicht regelmäßig neu vermietet, ist diese Methode nicht durchführbar. Die zweite Methode verwendet sogenannte hedonische Regressionen, um einen Preisindex zu erstellen. Der Kerngedanke besteht darin, den Mietpreis für eine Reihe von Immobilienmerkmalen zu kontrollieren und einen qualitätsbereinigten Preisindex im Zeitverlauf zu schätzen.

Rolling-Window-Zeit-Dummy-Hedonische Regression. Aufgrund des detailreichen Datensatzes, haben wir uns für die Verwendung hedonischer Regressionsmethoden entschieden, um die Indizes auf Stadtebene zu konstruieren. Insbesondere haben wir hedonische Regressionen mit rollierenden Zeitfenstern auf der Grundlage einer log-linearen Spezifikation verwendet. Eine solche log-lineare Spezifikation erlaubt es die Koeffizienten als prozentuale Änderung interpretieren. In unserer Basisspezifikation regressieren wir die logarithmierte Kaltmiete der Immobilie i auf eine Reihe von Immobilienmerkmalen (x_i) und einen Zeitdummy (D_τ) unter Verwendung einer logarithmierten Spezifikation:

$$\ln(P_t^i) = \beta^0 + \sum_{\tau=0}^T \gamma_\tau D_\tau^i + \sum_{k=1}^K (\beta^k * x^{k,i}) + \epsilon_t, \quad (1)$$

wobei ϵ_t der Fehlerterm ist.

Der Dummy-Parameter „Zeit“ ermöglicht es uns, den Einfluss der „Zeit“ auf den Preis zu bewerten, während wir für die Eigenschaften der Immobilie kontrollieren.

Unsere Kontrollvariablen bezeichnen wir mit x_i . Diese Variablen sind für alle Städte identisch und beinhalten Attribute wie die Wohnfläche, das Baujahr und die Anzahl der Zimmer.

Der Zeit-Dummy-Rolling-Window-Ansatz hat den entscheidenden Vorteil, dass er eine kontinuierliche Aktualisierung der Koeffizienten (β^k) im Zeitablauf ermöglicht. Dies bedeutet, dass die Methode Veränderungen in den Auswirkungen von Immobilienmerkmalen auf den Preis berücksichtigen kann, z. B. ob sich ein zusätzlicher Quadratmeter Wohnfläche im Jahr 2015 anders auf den Mietpreis auswirkte als heute. Ein wichtiger Aspekt dieser Methode ist die Größe dieses sogenannten „Fensters“. Wir haben uns dabei für einen Zeitraum von zwei Jahren und somit für eine hohe Flexibilität entschieden.

Die Beschreibung aller Variablen, die in dieser Dokumentation genannt werden, befindet sich in Tabelle 1. Dort wird zusätzlich zu einer Definition auch erläutert, um welche Art von Variable es sich handelt.²

Datenbereinigung. VALUE Marktdaten bereitet die Daten schon bereinigt vor und sorgt mit einem Algorithmus dafür, dass unter anderem keine Inserate mehrfach vorkommen. Dies wird in unserer Auswertung jedoch erneut überprüft und Datenpunkte die wiederholt auftreten entfernt. Wir verwerfen Daten, die über dem 99. Perzentil oder unter dem 1. Perzentil für Kaltmiete und Wohnfläche fallen. Zudem schließen wir Inserate die länger als 180 Tage aktiv waren in der Indexberechnung aus. Die Zeitspanne wie lange ein Inserat aktiv ist, ergibt sich aus dem erstmaligen auftreten und dem Zeitpunkt an dem ein Inserat nicht mehr veröffentlicht ist. Der Algorithmus der VALUE Marktdaten erkennt Angebote als Duplikate, welche mit Unterbrechung erneut inseriert werden. Die Dauer des Inserats wird somit automatisch angepasst. Da es sich hierbei nicht um Vertragsdaten sondern lediglich um Angebotsdaten handelt, können wir nicht sichergehen, dass der inserierte Preis auch der letztendliche Mietpreis ist. Um jedoch möglichst nahe an diesem tatsächlich vereinbarten Mietpreis zu sein ziehen wir den zuletzt inserierten Preis heran.³

²Hauptsächlich wird zwischen kontinuierlichen und kategorischen Variablen unterschieden.

³Historisch hat sich jedoch gezeigt, dass solche Änderungen sehr selten vorkommen und meist marginal sind.

3.1 Variablen

Tabelle 1: Beschreibung der für die hedonischen Regressionen verwendeten Variablen (1/2)

	Variable Name	Variable Definition	Details
1	AusstattungsKlassifikation	Ausstattungsklassifikation (5 Klassen)	Kategorische Variable
2	Badezimmer	Ausstattung des Badezimmers	Kategorische Variable
3	BaujahrKategorie	Kategorie des Baujahrs	Kategorische Variable
4	Balkon	Existiert ein Balkon und/oder eine Loggia	Kategorische Variable
5	Dachgeschoss	Handelt es sich um eine Immobilie im Dachgeschoss	Dummy-Variable
6	Etage	Etage der Einheit im Gebäude	Kategorische Variable
7	Garten	Garten, Gartenanteil oder Garten- nutzung ist gegeben	Kategorische Variable
8	GästeWC	Existiert ein Gäste-WC	Dummy-Variable
9	HeizungArt	Heizungsart nach Energieträger	Kategorische Variable
10	HeizungSystem	Kategorie des Heizungssystems	Kategorische Variable
11	Keller	Gebäude ist unterkellert oder zur Einheit gehört ein Keller	Kategorische Variable
12	Küche	Welche Art von Küche ist vorhanden	Kategorische Variable
13	Lift	Ein Personenaufzug ist vorhanden	Dummy-Variable
14	Objektyp	differenzierter Objektyp	Kategorische Variable
15	PLZ	Postleitzahl	Kategorische Variable
16	Zustand	Bau- und Nutzungszustand	Kategorische Variable
17	Zustandsklassifikation	In welchem Zustand befindet sich das Objekt	Kategorische Variable
18	ZimmerAnzahl	Anzahl der im Objekt vorhandenen Zimmer	Kategorische Variable
19	WNF	Wohnfläche des Objekts in m^2	Kontinuierliche Variable

3.2 Hedonische Preisindizes - Stadt

$$\begin{aligned}\ln(P_{i,t}) = & \beta^0 + \sum_{\tau=0}^5 \gamma^\tau D_i^\tau \\ & + \beta^1 \cdot WNF_i + \beta^2 \cdot WNF_i^2 \\ & + \beta^3 \cdot PLZ_i + \beta^4 \cdot HeizungArt_i \\ & + \beta^5 \cdot HeizungSystem_i + \beta^6 \cdot BaujahrKategorie_i \\ & + \beta^7 \cdot Etage_i + \beta^8 \cdot ZimmerAnzahl_i \\ & + \beta^9 \cdot Zustandsklassifikation_i + \beta^{10} \cdot Zustand_i \\ & + \beta^{11} \cdot Ausstattungsklassifikation_i + \beta^{12} \cdot Objekttyp_i \\ & + \beta^{13} \cdot Gästetoilette_i + \beta^{14} \cdot Balkon_i \\ & + \beta^{15} \cdot Lift_i + \beta^{16} \cdot Küche_i \\ & + \beta^{17} \cdot Dachboden_i + \beta^{18} \cdot Keller_i \\ & + \beta^{19} \cdot Dachgeschoss_i + \beta^{20} \cdot Garten_i \\ & + \beta^{21} \cdot Bad_i + \epsilon_{i,t}\end{aligned}\tag{2}$$

3.3 Deutschlandweiter Gesamtindex - Gewichtung

Neben den städtischen Indizes berechnen wir unter anderem einen bundesweiten Mietpreisindex. Durch die Gewichtung der einzelnen Städte errechnen wir einen repräsentativen Index für die gesamte Bundesrepublik Deutschland. Für diese Berechnung verwenden wir den Fisher-Preisindex (I_{FI}). Dieser wird durch das geometrische Mittel des Laspeyres-Index (I_{LA}) und des Paasche-Index (I_{PA}) berechnet.

Fischer-Index

$$I_{FI} = (I_{LA} + I_{PA})^{1/2}$$

Laspeyres-Index

$$I_{LA} = \frac{\sum_{i=1}^N p_{it} \cdot w_{i0}}{\sum_{i=1}^N p_{i0} \cdot w_{i0}}$$

Paasche-Index

$$I_{PA} = \frac{\sum_{i=1}^N p_{it} \cdot w_{it}}{\sum_{i=1}^N p_{i0} \cdot w_{it}}$$

Indexgewichte Die Indexpreise aus den jeweiligen Städten werden für die Berechnung von ILA , sowie IPA mit dem Gewicht w versehen. Diese Gewichtung wird aus dem Volumen (v_i) der Stadt in der entsprechenden Zeitperiode t durch das Gesamtvolumen in t berechnet. Volumen wird von uns als Anzahl der Inserate in t multipliziert mit der durchschnittlichen Kaltmiete in t .

$$w_{it} = \frac{v_{it}}{\sum_{i=1}^N v_{it}}$$

$$v_{it} = \sum_{i=1}^N \text{Inserate}_{it} \cdot \overline{\text{Kaltmiete}_{it}}$$

3.4 Inflationsbereinigung

Um auch die allgemeine Inflation zu berücksichtigen berechnen wir neben den nominalen Indizes auch reale Werte. Dafür verwenden wir den Verbraucherpreisindex für Deutschland. Unsere Datenquelle dafür sind die Datensätze des Internationalen Währungsfonds (IWF)⁴.

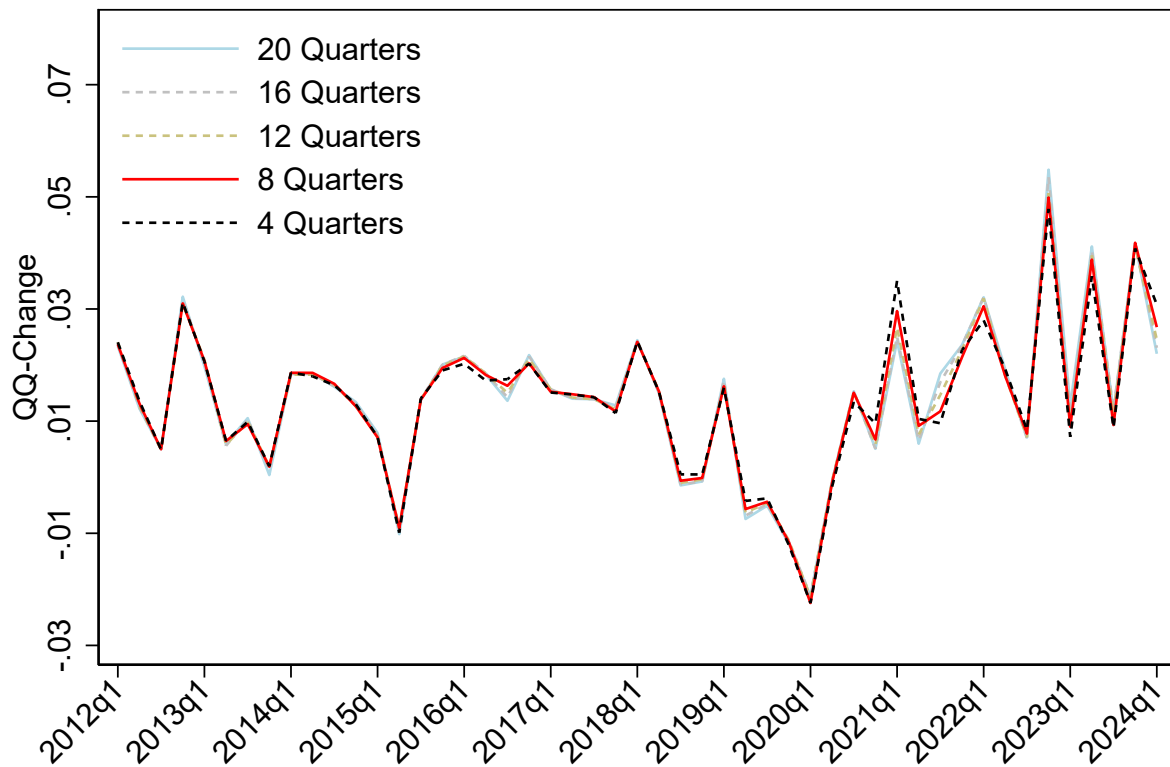
⁴<https://betadata.imf.org/>

3.5 Rolling-Windows

Wie bereits in der Einleitung erwähnt, verwenden wir eine rollierende Fenster (rolling-windows) Methode für die hedonische Regression. Der Vorteil einer solchen Methode ist es, dass ein sich verschiebendes Zeitfenster in der Lage ist strukturelle Datenänderung zu erkennen. Somit fallen ältere Beobachtungen aus der Regression und beeinflussen die Koeffizienten nicht mehr. Dennoch ist es wichtig den Zeitraum des rollierenden Fensters nicht zu klein oder zu groß zu wählen. Eine zu kurze Periode könnte zu einem sehr volatilen Ergebnis führen, in welchen kurzfristige Ausreißer die Resultate stark beeinflussen. Eine zu lange Periode hingegen kann dazu führen, dass die Indizes eine sehr schlechte Reaktionsfähigkeit aufweisen und dadurch aktuelle Trends nicht abbilden. Die nachstehenden Abbildungen zeigen die Mietpreisindexentwicklung sowie die prozentuale Entwicklung für Berlin. Es handelt sich dabei um Indizes, welche Quartalsdaten abbilden. In beiden Abbildungen werden Indizes abgebildet, die jeweils mit rollierenden Fenstern von 4, 8, 12, 16 und 20 Quartale berechnet wurden.



Berlin Percentage Change - Rolling Windows



In den fünf dargestellten Indizes lässt sich erkennen, dass über einen längeren Zeitraum hinweg, unabhängig von der Anzahl der rollierenden Fenster, ein sehr ähnlicher Trend abgebildet wird. Betrachtet man jedoch die prozentuale Veränderung zwischen den Quartalen, zeigt sich ein differenzierteres Bild. Insbesondere der Index aus vier Quartalen (schwarz gestrichelt) weist häufig eine deutlich höhere Volatilität auf als die anderen Indizes. Aus diesem Grund haben wir uns entschieden, einen Index mit rollierenden Fenstern zu konstruieren, der ein zweijähriges Zeitfenster umfasst. Diese Periodenlänge hat den Vorteil, dass sie aktuelle Veränderungen widerspiegelt, gleichzeitig jedoch weniger anfällig für Ausreißer ist.

4 Summary Statistics

Tabelle 2: Statistischer Überblick über die Kaltmiete pro Quadratmeter per Stadt (2012q1-2024q1).

	N	Mittelwert	SD	Min	Median	Max
Berlin	530817	10.26	4.39	2.36	9.11	48.48
Bonn	73831	10.01	2.54	2.47	9.69	35.63
Chemnitz	127153	5.35	0.99	2.35	5.21	16.10
Dortmund	136906	7.33	2.01	2.35	7.01	27.25
Dresden	234177	7.69	1.89	2.36	7.35	40.00
Duisburg	109612	6.26	1.46	2.36	6.00	25.88
Düsseldorf	169104	10.62	2.78	2.40	10.05	42.19
Erfurt	62253	7.48	1.66	2.38	7.25	22.78
FFM	165029	13.58	3.90	2.36	13.04	49.58
Hamburg	338978	11.81	3.58	2.38	11.29	49.09
Karlsruhe	41344	10.34	2.59	2.38	10.00	30.00
Köln	180903	11.12	3.19	2.38	10.68	39.75
Leipzig	315662	6.73	1.98	2.35	6.39	31.00
Lübeck	46853	8.48	2.40	2.50	8.08	26.19
München	201905	17.58	5.06	2.35	16.94	49.74
Münster	61397	10.16	2.76	2.38	9.86	47.50
Potsdam	31783	9.69	2.81	2.39	9.30	35.00
REK	60928	8.64	2.34	2.50	8.33	41.11
Stuttgart	97685	13.03	3.63	2.35	12.67	45.00
Wiesbaden	68539	10.50	2.42	2.38	10.00	25.25
Total N = 3054859						

Literatur

Balk, B., Diewert, W. E., Fenwick, D., Prud'homme, M., & de Haan, J. (2014). Handbook on residential property prices indices (rppis). Eurostat: Methodologies & Working Papers.