

# Der neue Regionale Preisindex (RPI) der amtlichen Statistik

Ludwig von Auer (Universität Trier)  
Sebastian Weinand (Deutsche Bundesbank)<sup>1</sup>

Düsseldorf, Juni 2018

---

<sup>1</sup>Diese Präsentation gibt die persönliche Meinung des Autors wieder und stellt nicht notwendigerweise die Sichtweise der Deutschen Bundesbank dar.

# 1 Vorbemerkungen

- Meine Vortragstitel der letzten drei Jahre:
  - 2015 (Schwerin) Regionaler Preisvergleich in Deutschland:  
Blaupause Bayern
  - 2016 (Eltville) Interregionaler Preisvergleich in Deutschland:  
Der aktuelle Stand
  - 2017 (Berlin) Interregionaler Preisvergleich in Deutschland:  
Erste Ergebnisse
- Nach zweijähriger Antragsphase hatten alle Statistischen Landesämter einem neuartigen Anonymisierungsprozess der Mikrodaten der Verbraucherpreisstatistik (Mai 2016) zugestimmt.
- Die Arbeit mit den Daten begann im September 2016 am Forschungsdatenzentrum (FDZ) Standort Hessen.

## 2. Datenaufbereitung

- Deutschland wird in 402 Regionen (295 Landkreise, 107 kreisfreie Städte) zerlegt.
- Insgesamt liegen uns 382.272 Preisbeobachtungen für 651 Güterkategorien vor.
- Für jede Preisbeobachtung muss überprüft werden, ob exakt das gleiche Produkt (inklusive Geschäftstyp) auch in einer anderen Region beobachtet wurde.
- Nur dann liegt eine für den regionalen Vergleich verwertbare Information vor (puristischer Messansatz).
- Für jede Güterkategorie wird eine eigene Preismatrix ermittelt.
- Sie besitzt immer 402 Spalten (Regionen), aber die Zeilenanzahl variiert.

	GT	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	3										0.89									
2	4					0.49	0.49		0.49		0.49			0.49	0.49	0.49	0.49			
3	2										0.89			0.89						
4	5										1.69									
5	4	0.98	0.98	0.98	0.98					0.98	0.98	0.98						0.98	0.98	0.98
6	5															0.49				
7	2										0.99									
8	2										1.69									
9	3										1.69									
10	3							2.59												
11	2													1.69			1.69			
12	4			1.79		1.99	1.75				1.75			1.79						
13	3														1.79					
14	5																			3.19
15	5			1.49		1.89														1.89
16	3													2.09					1.89	
17	3										1.99									
18	2																		1.99	
19	3															1.89				
20	3											1.99								

Tabelle 1: Preistableau für die Güterkategorie Reis.

	GT	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
3	2										0.89			0.89						
7	2										0.99									
8	2										1.69									
11	2													1.69			1.69			
18	2																1.99			
1	3										0.89									
9	3										1.69									
10	3							2.59												
13	3														1.79					
16	3													2.09					1.89	
17	3										1.99									
19	3															1.89				
20	3											1.99								
2	4					0.49	0.49		0.49		0.49			0.49	0.49	0.49	0.49			
5	4	0.98	0.98	0.98	0.98					0.98	0.98	0.98	0.98					0.98	0.98	0.98
12	4			1.79		1.99	1.75				1.75			1.79						
4	5										1.69									
6	5														0.49					
14	5																			3.19
15	5			1.49		1.89														1.89

Tabelle 2: Preistableau für Reis; nach Geschäftstypen sortiert.

## 3 Allgemeines Vorgehen

- Für Produkte (z.B. Uncle Ben's, Beutelreis 500g, Langkorn) liegen keine Ausgabengewichte vor.
- Für jede Güterkategorie liegen Ausgabengewichte vor (z.B. Reis 0,039%).
- Für zwei Drittel der Güterkategorien liegen sogar Ausgabengewichte der Geschäftstypen vor.
- Ausgabengewichte sind regional einheitlich (z.B. kein höheres „Biergewicht“ in München).
- Die Ausgabengewichte dürfen nicht am FDZ Standort Hessen verwendet werden („externe Daten“).
- Dies schränkt die methodischen Optionen ein.
- Die Mietdaten (5 der 651 Güterkategorien) werden gesondert ausgewertet.

- Berechnungsverfahren im Überblick

**Schritt 1:** Berechnung regionaler Mietniveaus; im FDZ.

**Schritt 2:** Berechnung regionaler Preisniveaus für jede Güterkategorie-Geschäftstyp-Kombination (z.B. Reis im SB Warenhaus, Reis im Supermarkt, ...); im FDZ.

**Schritt 3:** Berechnung regionaler Preisniveaus für jede Güterkategorie (z.B. Reis); Ausnutzung der Ausgabengewichte der Geschäftstypen; außerhalb FDZ.

**Schritt 4:** Berechnung regionaler Preisniveaus; Ausnutzung der Ausgabengewichte der Güterkategorien; außerhalb FDZ.

- Schritt 1 wird mit dem *hedonischen Regressionsansatz* durchgeführt.
- Für die Schritte 2, 3 und 4 wird der Regressionsansatz der *Country Product Dummy Methode* (CPD-Methode) eingesetzt.



## 4 Regionale Mietniveaus (Schritt 1)

- Es liegen 15.582 Mietbeobachtungen vor (davon 315 für Einfamilienhäuser; sie wurden ausgeschlossen).
- Die Mietbeobachtungen decken 381 der 402 Regionen ab.
- Manche Mietbeobachtungen sind unvollständig.
- Deshalb wird für 15 Regionen lediglich die durchschnittliche Quadratmetermiete berechnet.
- Die Mietbeobachtungen der anderen 366 Regionen gehen in die hedonische Regression ein.

- Sie erklärt die Miete einer Wohnung  $i$  aus Region  $r$  durch die folgenden Wohnungseigenschaften:
  - Region (365 Dummy-Variablen  $region_i^r$ )
  - Wohnungsgröße ( $flaeche_i$ )
  - Dauer des Mietverhältnisses ( $dau_i$ )
  - Ausstattungsqualität (zwei Dummy-Variablen  $ausst_{ai}$ )
  - Lage (drei Dummy-Variablen  $lage_{li}$ )
  - öffentlicher Vermieter (Dummy-Variable  $soz_i$ )
  - Küche (Dummy-Variable  $kueche_i$ )

- Die Regressionsgleichung lautet:

$$\begin{aligned} \ln rent_i &= \alpha + \sum_{r=1}^{366} \beta_{0r} region_i^r + \beta_1 \ln (flaeche_i) \\ &\quad + \beta_2 soz_i + \beta_3 \ln (dau_i) + \beta_4 \ln (dau_i) soz_i \\ &\quad + \sum_{a=1}^2 \beta_{5a} ausst_{ai} + \sum_{l=1}^3 \beta_{6l} lage_{li} + \beta_7 kueche_i + u_i \end{aligned}$$

mit  $\sum_{r=1}^{366} \hat{\beta}_{0r} = 0$  und

$$region_i^r = \begin{cases} 1 & \text{wenn Wohnung } i \text{ in Region } r \\ 0 & \text{wenn Wohnung } i \text{ nicht in Region } r \end{cases}$$

- Es wird eine Kleinst-Quadrat-Schätzung durchgeführt.
- Auf Basis der Schätzergebnisse  $(\hat{\alpha}, \hat{\beta}_{01}, \dots, \hat{\beta}_7)$  wird für jede der 366 Regionen die (logarithmierte) Miete berechnet, welche für eine Wohnung mit folgenden Eigenschaften zu erwarten ist:
  - 65 m<sup>2</sup>
  - privat vermietet
  - 7-jähriges Mietverhältnis
  - mittlere Ausstattung
  - mittlere Lage
  - Einbauküche

- Für die 15 Regionen mit unvollständigen Daten wird der berechnete Quadratmeterpreis mit 65 multipliziert und logarithmiert.
- Die Ergebnisse werden mit den Ergebnissen aus der hedonischen Regression verknüpft.
- Es ergibt sich ein Vektor der regionalen Mietniveaus

$$\widehat{\ln P}_{\text{miete}} = \left( \widehat{\ln P^1}_{\text{miete}} \dots \widehat{\ln P^{402}}_{\text{miete}} \right)$$

- Der Vektor hat Lücken, denn für 21 Regionen lagen keine Mietpreisbeobachtungen vor.

## 5 Regionale Preisniveaus für Güterkategorie-Geschäftstyp-Kombinationen (Schritt 2)

- Es wird eine bestimmte Güterkategorie-Geschäftstyp-Kombination betrachtet (z.B. Reis verkauft im Supermarkt)
- $P^r$  ist das Preisniveau der Region  $r$ ,  
 $\pi_i$  ist der Wert des Produktes  $i$  (der betrachteten Güterkategorie-Geschäftstyp-Kombination)
- In der CPD-Methode wird unterstellt, dass

$$preis_i^r = P^r \pi_i \varepsilon_i^r, \quad (1)$$

wobei  $\ln \varepsilon_i^r \sim N(0, \sigma^2)$ .

- (1) kann auch in folgender Form geschrieben werden:

$$\ln price_i^r = \sum_{s=1}^R \ln P^s \text{ region}^s + \sum_{j=1}^N (\ln \pi_j) \text{ produkt}_j + u_i^r, \quad (2)$$

mit  $\sum_{s=1}^R \widehat{\ln P^s} = 0$  und

$$\text{region}^s = \begin{cases} 1 & \text{wenn } r = s \\ 0 & \text{wenn } r \neq s \end{cases}$$

$$\text{produkt}_j = \begin{cases} 1 & \text{wenn } i = j \\ 0 & \text{wenn } i \neq j \end{cases}$$

- Kleinst-Quadrat-Schätzung der Gleichung (2) liefert die geschätzten logarithmierten regionalen Preisniveaus  $\widehat{\ln P^r}$  ( $r = 1, \dots, 402$ ).
- Sie werden im Vektor

$$\widehat{\ln P} = \left( \widehat{\ln P^1} \ \widehat{\ln P^2} \ \dots \ \widehat{\ln P^{402}} \right)$$

zusammengefasst und sind so normiert, dass das durchschnittliche Preisniveau gleich 1 beträgt.

- Für jede Güterkategorie-Geschäftstyp-Kombination ergibt sich ein eigener Vektor (insgesamt 1648 Vektoren).
- Fast alle Vektoren sind lückenhaft, viele sogar leer.
- Für Reis existieren vier Geschäftstypen. Es liegen also vier Vektoren vor.



## 5. Regionale Preisniveaus für Güterkategorie-Geschäftstyp-Kombinationen (Schritt 2)

	GT	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
3	2										0.89			0.89						
7	2										0.99									
8	2										1.69									
11	2													1.69			1.69			
18	2																1.99			
1	3										0.89									
9	3										1.69									
10	3							2.59												
13	3														1.79					
16	3													2.09					1.89	
17	3										1.99									
19	3															1.89				
20	3											1.99								
2	4					0.49	0.49		0.49		0.49			0.49	0.49	0.49	0.49			
5	4	0.98	0.98	0.98	0.98					0.98	0.98	0.98	0.98					0.98	0.98	0.98
12	4			1.79		1.99	1.75				1.75			1.79						
4	5										1.69									
6	5														0.49					
14	5																			3.19
15	5			1.49		1.89														1.89

Tabelle 2 (Wdh.): Preistableau für Reis; nach Geschäftstypen sortiert.

## 6 Regionale Preisniveaus für Güterkategorien (Schritt 3)

- In Schritt 3 werden die Vektoren einer Güterkategorie für eine weitere CPD-Regression herangezogen.
- Da für die Geschäftstypen ( $t$ ) Ausgabengewichte ( $w_t$ ) vorliegen, wird eine gewichtete Variante eingesetzt:

$$\sqrt{w_t} \widehat{\ln P_t^r} = \sqrt{w_t} \sum_{s=1}^R \ln P^s \text{region}^s + \sqrt{w_t} \sum_{m=1}^T \ln \pi_m \text{geschtyp}_m + u_t^r, \quad (3)$$

mit  $\sum_{s=1}^R \widehat{\ln P^s} = 0$  und

$$\text{geschtyp}_m = \begin{cases} 1 & \text{wenn } t = m \\ 0 & \text{wenn } t \neq m \end{cases}$$

- Die Schätzung von (3) liefert für die betrachtete Güterkategorie den Vektor  $\widehat{\ln P} = \left( \widehat{\ln P^1} \widehat{\ln P^2} \dots \widehat{\ln P^{402}} \right)$ .

- Für jede der 646 Güterkategorien ( $g$ ) wird ein solcher Vektor ( $\widehat{\ln P}_g$ ) ermittelt:

$$\widehat{\ln P}_1 = \left( \widehat{\ln P}_1^1 \quad \widehat{\ln P}_1^2 \quad \dots \quad \widehat{\ln P}_1^{402} \right)$$

$$\widehat{\ln P}_2 = \left( \widehat{\ln P}_2^1 \quad \widehat{\ln P}_2^2 \quad \dots \quad \widehat{\ln P}_2^{402} \right)$$

$$\vdots$$

$$\widehat{\ln P}_{646} = \left( \widehat{\ln P}_{646}^1 \quad \widehat{\ln P}_{646}^2 \quad \dots \quad \widehat{\ln P}_{646}^{402} \right)$$

- Die meisten dieser Vektoren sind lückenhaft.

## 7 Regionale Preisniveaus (Schritt 4)

- Für jede Güterkategorie ( $g$ ) existiert ein Ausgabengewicht ( $w_g$ ).
- In Schritt 4 werden die regionalen Mietniveaus aus Schritt 1 – der Vektor  $\widehat{\ln P}_{\text{miete}}$  – und die regionalen Preisniveaus der Güterkategorien aus Schritt 3 – die Vektoren  $\widehat{\ln P}_1$  bis  $\widehat{\ln P}_{646}$  – in einer weiteren gewichteten CPD-Regression verknüpft.
- Insgesamt werden also 647 Vektoren verarbeitet.

- Die Regressionsgleichung lautet:

$$\sqrt{w_g} \widehat{\ln P_g^r} = \sqrt{w_g} \sum_{s=1}^R \ln P^s \text{ region}^s + \sqrt{w_g} \sum_{h=1}^H \ln \pi_h \text{ gutkat}_h + u_g^r \quad (4)$$

mit  $\sum_{s=1}^R \widehat{\ln P^s} = 0$  und

$$\text{gutkat}_h = \begin{cases} 1 & \text{wenn } g = h \\ 0 & \text{wenn } g \neq h \end{cases}$$

- Die Schätzung von (4) liefert die logarithmierten regionalen Preisniveaus  $\widehat{\ln P} = (\widehat{\ln P^1} \widehat{\ln P^2} \dots \widehat{\ln P^{402}})$ .
- Auch hier ergeben sich bislang noch einzelne Lücken.

## 8 Ergebnisse

- Für 393 der 402 Regionen konnte das jeweilige Preisniveau ermittelt werden.
- Der regionale Preisniveauindex (Karte "All") reicht von 0,79 bis 1,27.
- Der Wohnkostenindex (Karte "Housing") reicht von 0,65 bis 1,55.
- Der Nicht-Wohnkostenindex (Karte "Non-Housing") reicht von 0,95 bis 1,04.

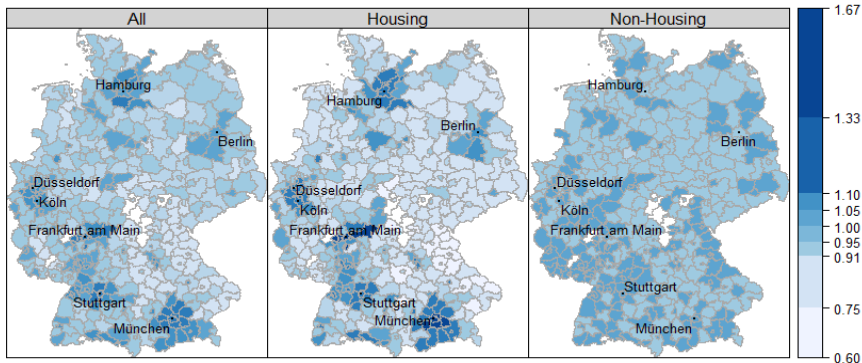


Abbildung 1: Regionale Preisniveaus in Deutschland

## 9 Der neue Regionale Preisindex (RPI)

- Kein Land der Welt hat bislang einen *flächendeckenden offiziellen* regionalen Preisindex (RPI) auf *NUTS-3 Ebene*.
- Versetzen wir uns ins Jahr 2020. Was könnten dann



gemeinsam haben?



## 10 Fazit und Ausblick

- Eine Berechnung regionaler Preisniveaus ist mit den ohnehin erhobenen Preisdaten für ganz Deutschland möglich.
- Datenanonymisierung und -aufbereitung machen den Großteil der Arbeit aus.
- Für den Vergleich kann ein repräsentativer Monat ausgewählt werden.
- Es muss nicht jedes Jahr eine vollständige Berechnung durchgeführt werden.
- Zentral für die Verlässlichkeit der Ergebnisse ist der Wohnkostenindex.
- Destatis könnte sich durch den weltweit besten regionalen Preisindex (RPI) profilieren.