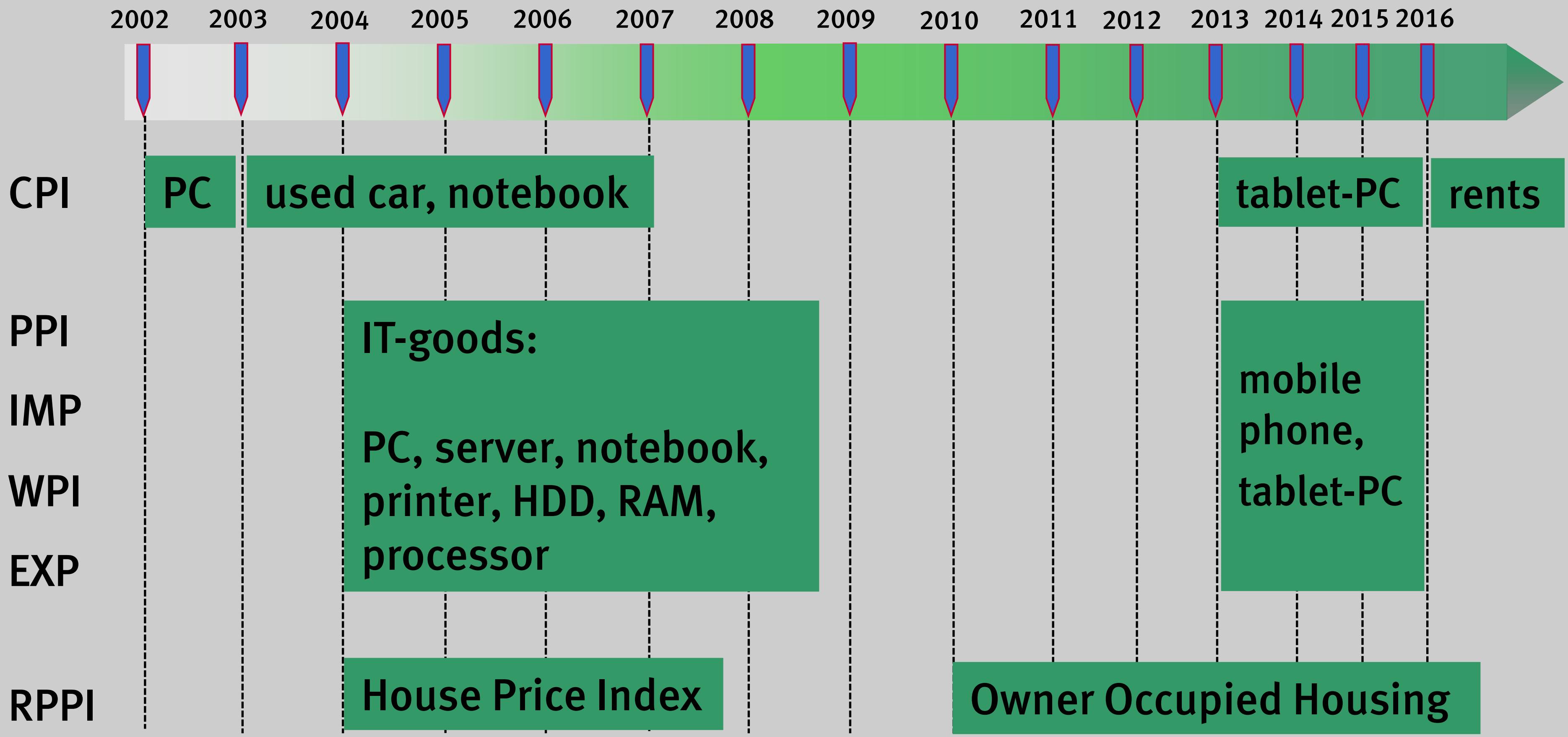
New developments in hedonic quality adjustment in German price statistics

Timm Behrmann, Bernhard Goldhammer

15th Meeting of the Ottawa Group on Price Indices
10. - 12. May 2017 in Eltville am Rhein

Hedonic methods in German prices statistics



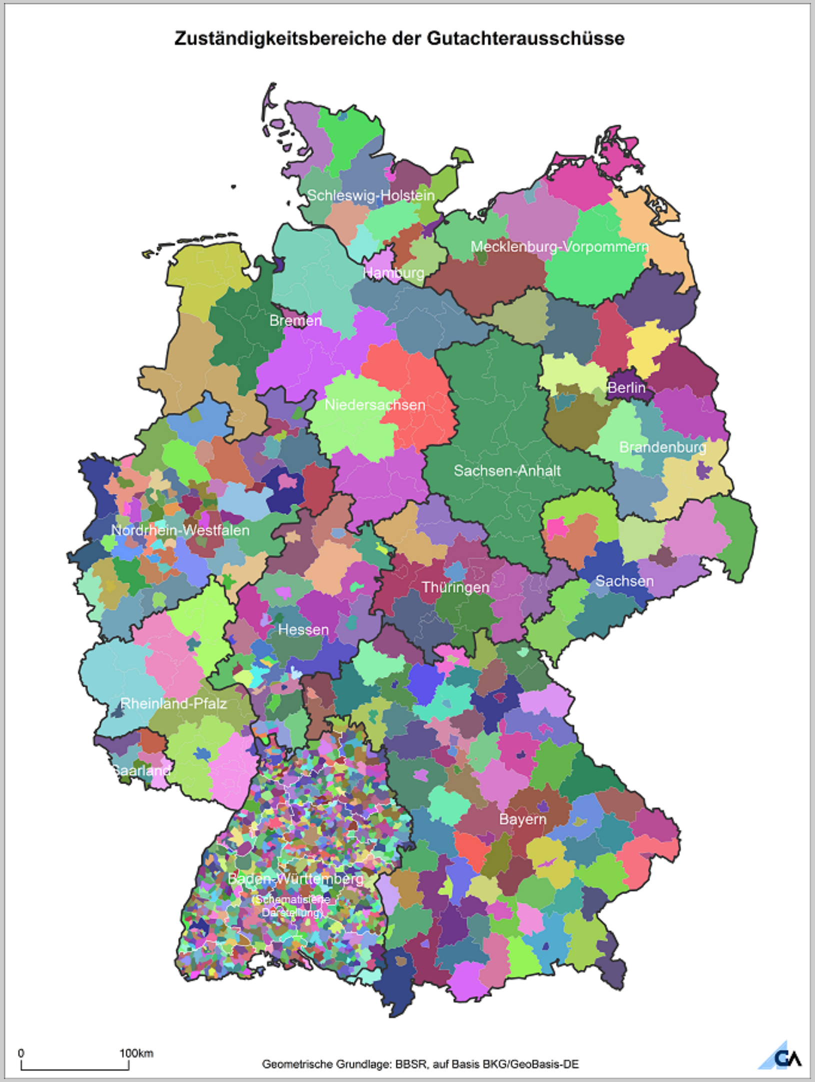
Legal Basis

- Commission regulation (EU) 93/2013
  - Legal obligation to calculate House Price Indices
  - Net acquisition approach
  - Transaction prices (building + property)
  - Quality adjustment: hedonic methods are recommended

  - Indices for newly built and existing properties
  - Indices at least quarterly with a delivery deadline of t+85
  - Yearly updating of the weighting scheme
  - Data published since 2008, for 14 states since 2005

Data Source: ECPVs

- ECPVs
  - Expert Committees for Property Valuation
  - Organised on different regional levels
  - Administrative data
  - All transactions
  - Transaction prices
  - Sales price collection



Characteristics of residential property

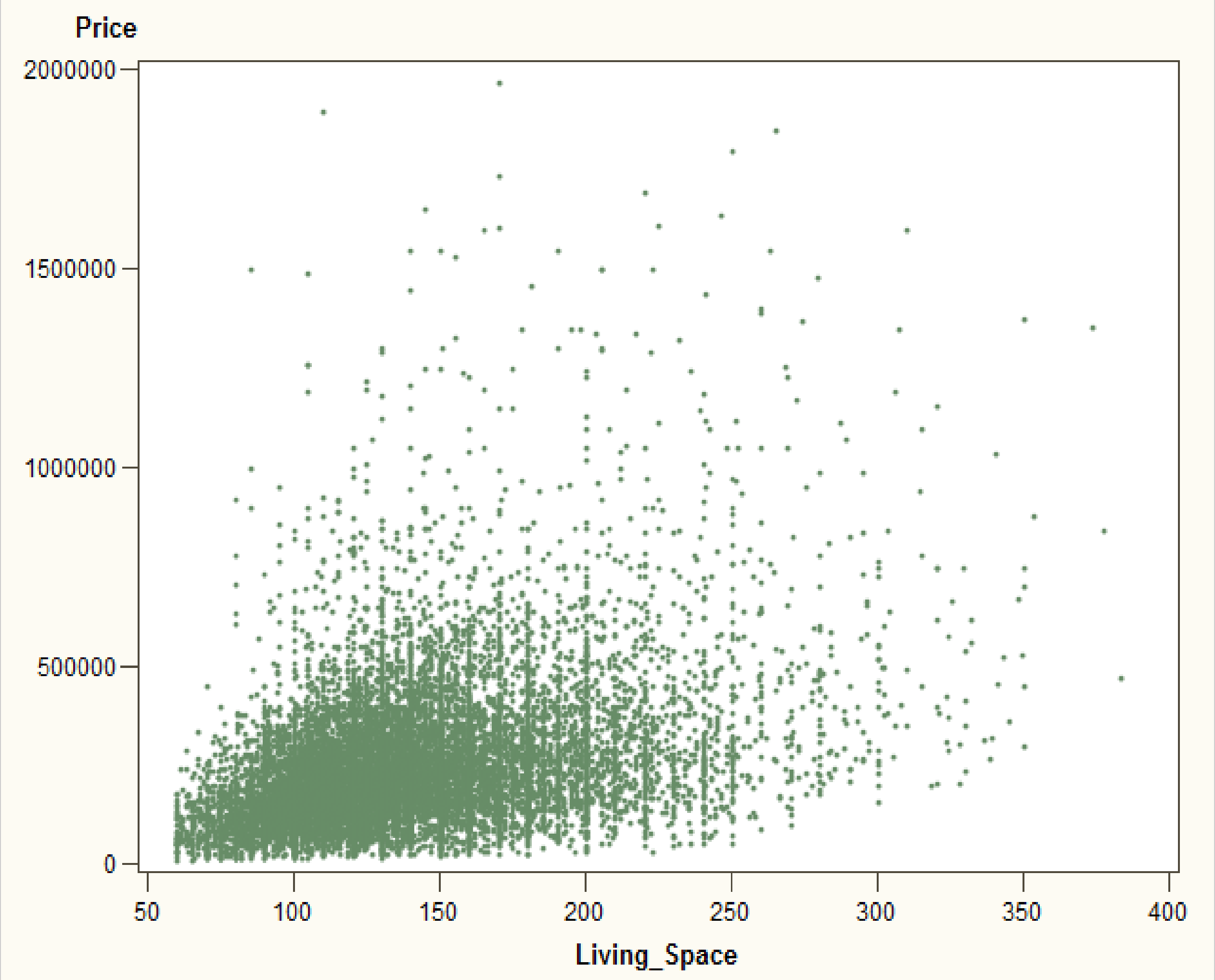
© ArTo - Fotolia.com

Hedonic regression analysis

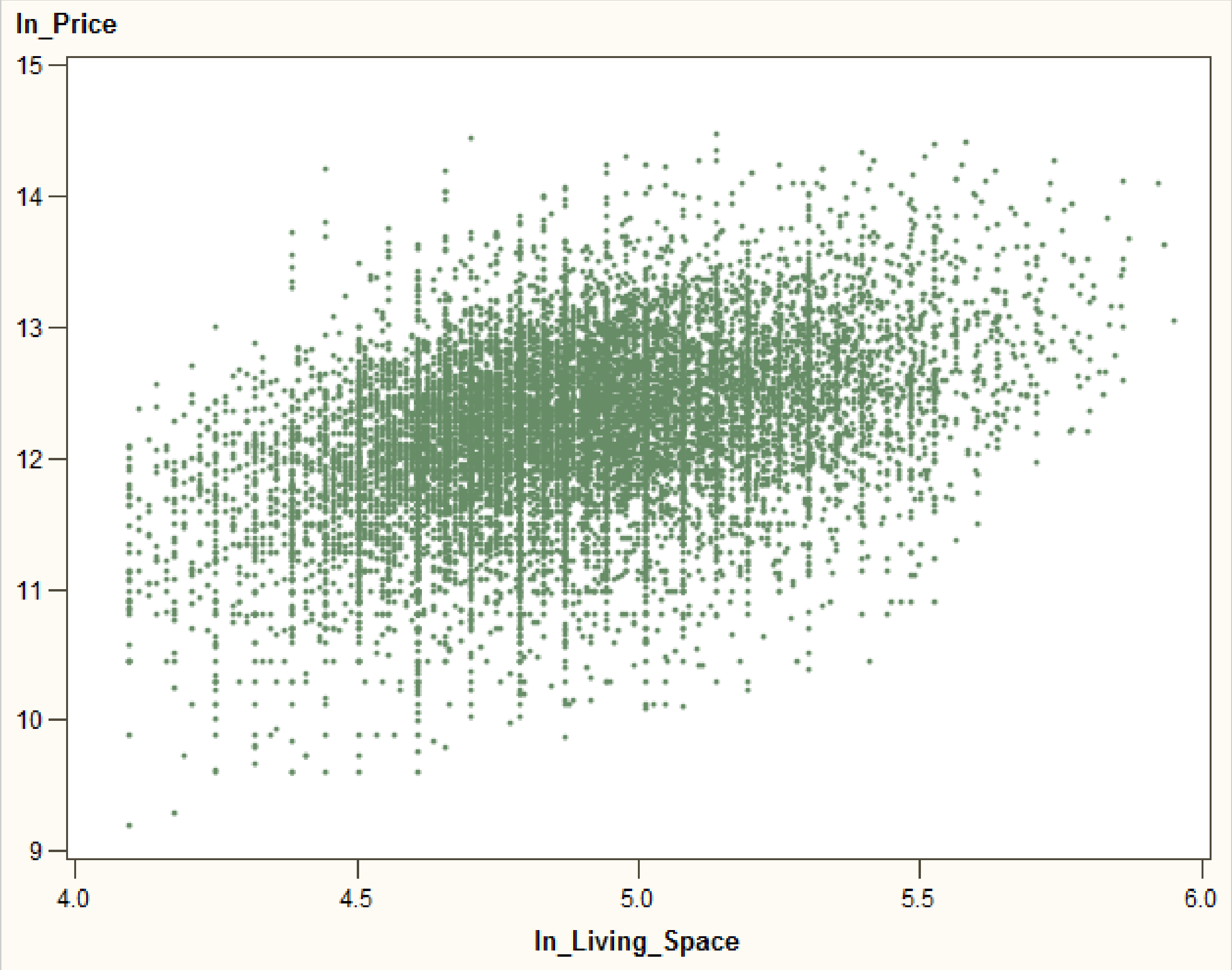
- OLS regression
- Four regression models:
  - Existing + newly built houses
  - Existing + newly built flats
- Log-Log model
- Regression for the whole country
- Formula for existing houses:

$$\ln(p) = \beta_0 + \beta_1 \cdot \ln(living\ space) + \beta_2 \cdot \ln(site\ area) + \beta_3 \cdot \ln(age) + \sum \beta_i \cdot d(SLV_i) + \varepsilon$$

Scatter plot: existing houses 2016Q4



Scatter plot: existing houses 2016Q4



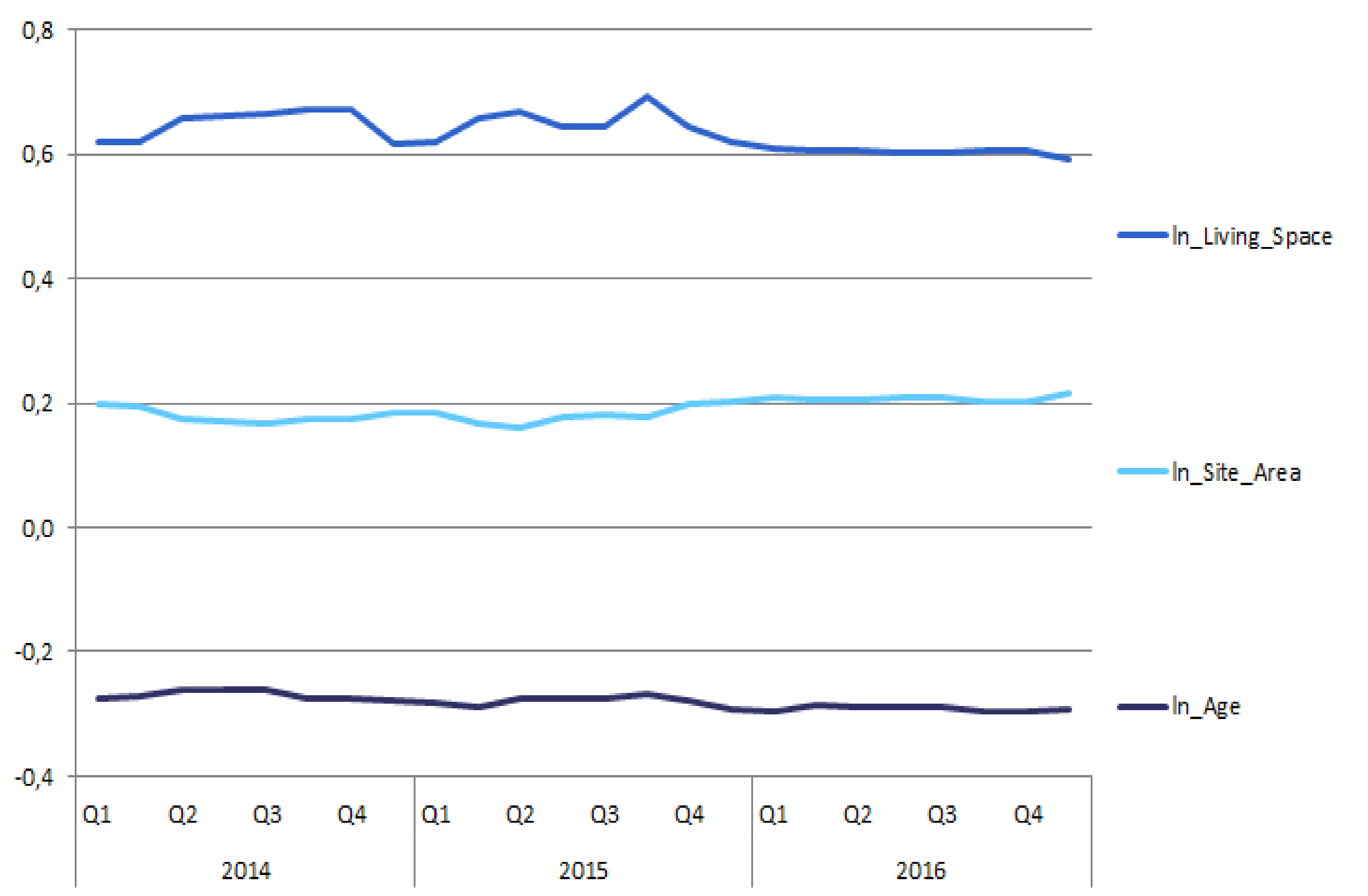
Regression output: existing houses 2016Q4

Varianzanalyse					
Quelle	DF	Summe der Quadrate	Mittleres Quadrat	F-Statistik	Pr > F
Modell	10	4211.37259	421.13726	3634.93	<.0001
Error	13344	1546.01312	0.11586		
Corrected Total	13354	5757.38571			

Root MSE	0.34038	R-Quadrat	0.7315
Dependent Mean	12.23899	Adj R-Sq	0.7313
Coeff Var	2.78111		

Parameter Estimates							
Variable	DF	Parameter-schätzer	Standard-fehler	t-Wert	Pr > |t|	Quadrat;Semi-partiell;Korr. Typ I	Varianz Inflation
Intercept	1	9.09926	0.05417	167.97	<.0001	.	0
ln_living_space	1	0.59305	0.01095	54.17	<.0001	0.16385	1.21095
ln_site_area	1	0.21654	0.00574	37.75	<.0001	0.02601	1.40487
ln_age	1	-0.29297	0.00423	-69.19	<.0001	0.08711	1.04408
SLV_1	1	-0.76400	0.00921	-82.94	<.0001	0.18376	1.40435
SLV_2	1	-0.35602	0.00845	-42.14	<.0001	0.11845	1.36256
SLV_4	1	0.35879	0.00827	43.39	<.0001	0.00717	1.35399
SLV_5	1	0.71910	0.01427	50.38	<.0001	0.03344	1.12216
SLV_6	1	0.97752	0.02098	46.59	<.0001	0.03676	1.05612
SLV_7	1	1.22124	0.03755	32.53	<.0001	0.01939	1.01572
SLV_8	1	1.52485	0.02903	52.53	<.0001	0.05554	1.02871

Regression coefficients: existing houses



Hedonic Imputation Method

© Daniel Berkmann - Fotolia.com

$$\hat{p} = f\left(X_A, \hat{\beta}_{t=0}\right) \longrightarrow \hat{p} = f\left(X_A, \hat{\beta}_{t=1}\right)$$

$$\hat{p} = f\left(X_B, \hat{\beta}_{t=0}\right) \longrightarrow \hat{p} = f\left(X_B, \hat{\beta}_{t=1}\right)$$

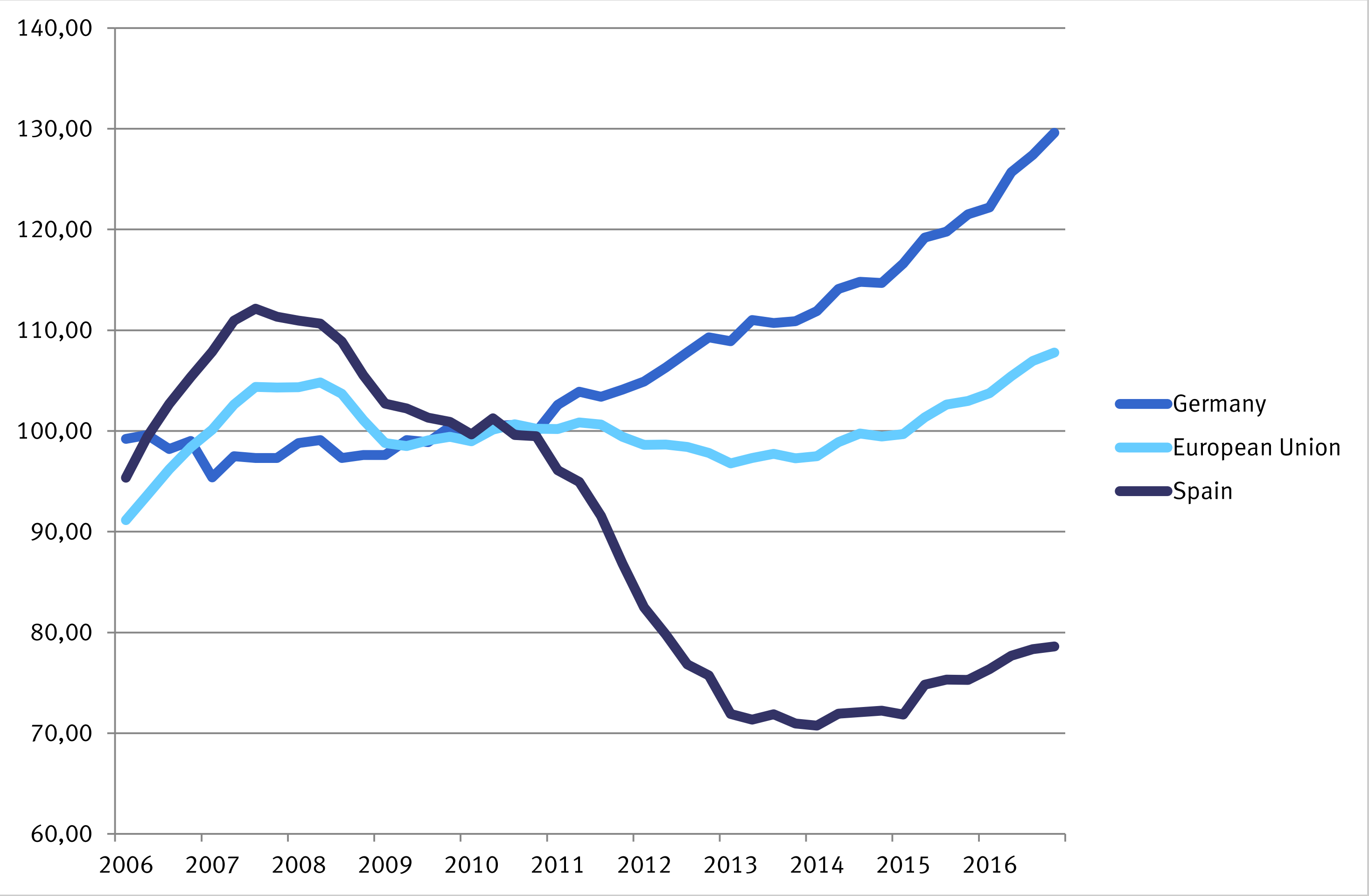
© KB3 - Fotolia.com

Index calculation

- **Aggregation on elementary level**
  - **Jevons-Index, geometric mean**
  - **Equal weighting of observations**

$$I_{t,T}^{Jevons} = \frac{\sqrt[n]{\prod_n p_T}}{\sqrt[n]{\prod_n p_t}} = \sqrt[n]{\prod_n \frac{p_T}{p_t}}$$

House Price Index results



The German rental sample for consumer price statistics

- Weight: 21% (CPI, with rental equivalence approach), 10.4% (HICP)
- Size: 20,000 dwellings, stratified according to
  - Type of dwelling
  - Type of landlord
  - Regional dimension
- New recommendation by Eurostat for Quality adjustment
- Quality adjustment approach so far:

Situation	Approach used	Approach recommended	In-line with recommendation?
Modernization	Supported judgmental	Hedonic (A), Bridged overlap, Supported judgmental (B)	✓
Replacement	Index neutrality	Hedonic (A), Bridged overlap (B)	✗

=> Need for improvement: hedonic quality adjustment

Surveyed characteristics for rented dwellings

Characteristic	Purpose	Price relevance
Address	Identification	No
Dwelling number	Identification	No
Location of Dwelling in the building	Identification	No
Municipality -> Type of district	Differentiation	Yes
Type of landlord	Differentiation	Yes
Energy consumption/need	Quality	Yes
Quality of residential area	Quality	Yes
Quality of furnishings	Quality	Yes
Date of rental contract	Modification	Yes
Other furnishings: balcony, garage, fitted kitchen	Quality	Yes
Type of financing/end of funding period	Differentiation	Yes
Year of construction	Differentiation/Quality	Yes
Living space (m^2)	Differentiation/Quality	Yes

These characteristics should be tested for a price estimation

Hedonic model for rented dwellings

$$\ln p_i = \beta_0 + \sum_{j=2}^{3} \beta_{1j} v_{ij} + \sum_{k=1}^{3} \beta_{2k} d_{ik} + \sum_{l \in PR} \beta_{3l} r_{il} + \beta_4 \ln en_i + \beta_5 ec_i + \sum_{m=2}^{4} \beta_{6m} qa_{im}$$
$$+ \sum_{n=2}^{3} \beta_{7n} qf_{in} + \beta_8 bc_i + \beta_9 gar_i + \beta_{10} fk_i + \beta_{11} \ln age_i + \beta_{12} \ln ls_i$$
$$+ \beta_{14} \ln dc_i + \beta_{15} ff_i + u_i$$

Continuous variables

- en_i energy consumption or need (kWh/m²*a)
- age_i current age of dwelling in years (today-year of construction)
- ls_i living space (size of dwelling) in m²
- dc_i duration of rental contract in days (today-concluding date of contract)

Hedonic model for rented dwellings

$$\ln p_i = \beta_0 + \sum_{j=2}^{3} \beta_{1j} v_{ij} + \sum_{k=1}^{3} \beta_{2k} d_{ik} + \sum_{l \in PR} \beta_{3l} r_{il} + \beta_4 \ln en_i + \beta_5 ec_i + \sum_{m=2}^{4} \beta_{6m} qa_{im}$$
$$+ \sum_{n=2}^{3} \beta_{7n} qf_{in} + \beta_8 bc_i + \beta_9 gar_i + \beta_{10} fk_i + \beta_{11} \ln age_i + \beta_{12} \ln ls_i$$
$$+ \beta_{14} \ln dc_i + \beta_{15} ff_i + u_i$$

Dummy variables

- v_{ij} type of landlord, $j \in \{2,3\}$
- d_{ik} type of district, $\mathrm{k} \in \{1,2,3\}$
- r_{il} region (96 planning regions)
- ec_i type of energy performance certificate (consumption)
- qa_{im} quality of residential area, $\mathrm{m} \in \{2,3,4\}$

Hedonic model for rented dwellings

$$\ln p_i = \beta_0 + \sum_{j=2}^{3} \beta_{1j} v_{ij} + \sum_{k=1}^{3} \beta_{2k} d_{ik} + \sum_{l \in PR} \beta_{3l} r_{il} + \beta_4 \ln en_i + \beta_5 ec_i + \sum_{m=2}^{4} \beta_{6m} qa_{im}$$
$$+ \sum_{n=2}^{3} \beta_{7n} qf_{in} + \beta_8 bc_i + \beta_9 gar_i + \beta_{10} fk_i + \beta_{11} \ln age_i + \beta_{12} \ln ls_i$$
$$+ \beta_{14} \ln dc_i + \beta_{15} ff_i + u_i$$

Dummy variables

- qf_{in} quality of furnishing, $n \in \{2,3\}$
- bc_i, gar_i, fk_i dwelling equipped with balcony, garage, fitted kitchen
- ff_i type of financing: free-financed
- u_i residual

Estimation results for the hedonic model

Variables not significant:

- d_{i3}=>no difference in rents in both types of rural districts
- ec_i=>both ways of calculation of energy performance lead to comparable results
- gar_i=> disturbing finding, probably coding error. Coefficient estimated last year used: $\beta(gar_i)$=0.033

Number of *Länder* included	12
Number of observations	16798
Number of observations used for regression	14314
R^2	0.767
R^2 adjusted	0.766
Sum of squared residuals	495.5
Standard error of the regression (σ)	0.187

Variable	Coefficient	Standard Error
Type of landlord		
v_{i2}	-0.070^{***}	0.006
v_{i3}	0.037^{***}	0.006
Regional dimension: type of district		
d_{i1}	0.137^{***}	0.005
d_{i2}	0.042^{***}	0.042
Regional dimension: planning region		
$r_{i,910}$	0.424^{***}	0.007
$r_{i,810}$	0.297^{***}	0.010
… (52 other regions with significant coefficient)		
$r_{i,911}$	-0.149^{***}	0.019
$r_{i,914}$	-0.220^{***}	0.023
Energy consumption/need		
en_i	-0.043^{***}	0.004
Quality of residential area		
$qa_{i,2}$	0.047^{***}	0.006
$qa_{i,3}$	0.094^{***}	0.006
$qa_{i,4}$	0.161^{***}	0.009

Variable	Coefficient	Standard Error
Quality of furnishing		
$qf_{i,2}$	0.041^{***}	0.004
$qf_{i,3}$	0.130^{***}	0.007
Other furnishings		
bc_i	0.018^{***}	0.004
gar_i	-0.006	0.006
fk_i	0.080^{***}	0.007
Year of construction (age)		
age_i	-0.032^{***}	0.003
Living space		
ls_i	0.868^{***}	0.006
Date of contract (duration)		
dc_i	-0.048^{***}	0.001
Type of financing (free-financed)		
ff_i	0.134^{***}	0.005

From regression equation to quality adjustment (1)

Transformation of double-logarithmic estimation results into level results

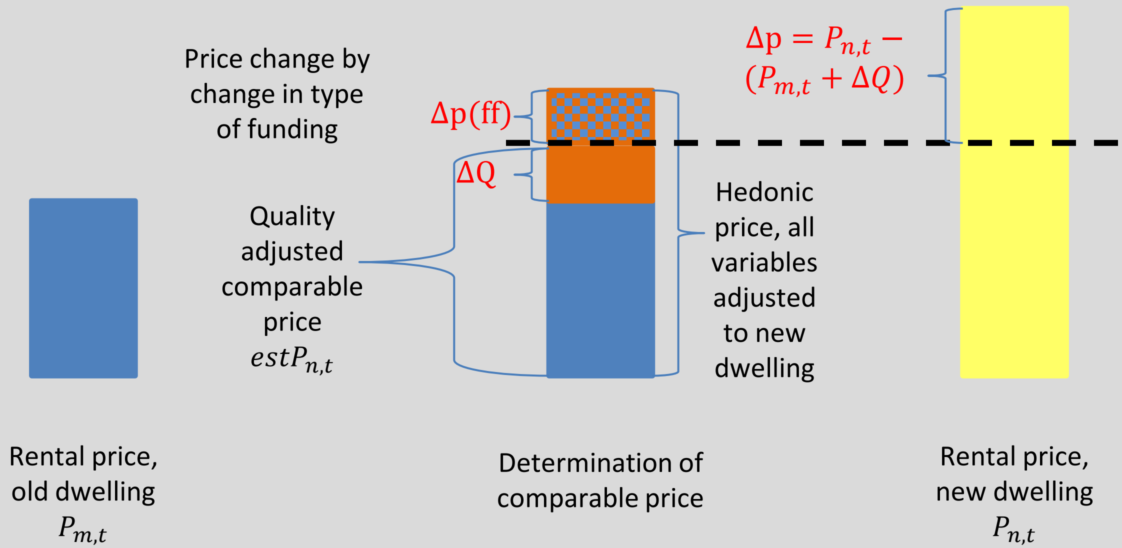
- Remember: $\ln p_i = \beta_0 + \sum_{j=2}^{3} \beta_{1j} v_{ij} + \cdots + u_i$ => $e^{\ln p_i} = e^{\beta_0 + \sum_{j=2}^{3} \beta_{1j} v_{ij} + \cdots + u_i}$
- $E(e^{u_i}) = e^{\frac{\sigma^2}{2}} \neq e^0$ (Wooldridge, 2009, p. 211) => adjustment factor needed
- $\hat{p}_i = \exp(\frac{\sigma^2}{2}) \cdot \exp(\widehat{\ln p_i}) = \alpha_0 \cdot \exp(\widehat{\ln p_i})$
- Three approaches for determination of α_0:
    - Use of estimated standard error of regression: $\tilde{\alpha}_0 = exp(\frac{\hat{\sigma}^2}{2})$
    - Use of estimated residuals: $\hat{\alpha}_0 = \frac{1}{n} \sum_{i=1}^{n} exp(\hat{u}_i{}')$
    - Minimization of sum of squared estimateted residuals: $\check{\alpha}_0 = \left(\sum_{i=1}^{n} exp(\widehat{\ln p_i})^2\right)^{-1} \cdot \left(\sum_{i=1}^{n} exp(\widehat{\ln p_i}) \cdot p_i\right)$

$\Rightarrow$ First approach is used: $\tilde{\alpha}_0 = 1.0175$
$\Rightarrow$ Without $\tilde{\alpha}_0$, underestimation by 1.8%

From regression equation to quality adjustment (2)

Treatment of variables relevant to the price, but not to the quality

- Type of financing, duration of rental contract
- Remember *ceteris paribus* comparison: change the quality, keep all other things constant



⇒ For quality valuation, keep old value constant
⇒ The impact of the change of this variables is seen as price change

From regression equation to quality adjustment (3)

Interpretation of dummy variables in a double-logarithmic estimation

- What happens with dummy variables during re-transformation?
- Easy example, x continuous variable, y dummy variable, no residual:

$$\ln p_i = \beta_1 \ln x_i + \beta_2 y_i$$

- Re-transformation:

$$p_i = e^{\beta_1 \ln x_i + \beta_2 y_i} = e^{\beta_1 \ln x_i} \cdot e^{\beta_2 y_i} = x_i^{\beta_1} \cdot e^{\beta_2 y_i} = \begin{cases} x_i^{\beta_1} & if\ y_i = 0 \\ x_i^{\beta_1} \cdot e^{\beta_2} & if\ y_i = 1 \end{cases}$$

- Dummy variables lead to a multiplicative adjustment of the price estimated with continuous variables
  - Appropriate for characteristics that apply to the dwelling as a whole, e.g. quality of furnishing or quality of area
  - Problematic for characteristics that show a certain feature, e.g. garage

$\Rightarrow$ Awareness of this limitation, but lack of solution

Estimation of ΔQ: difference approach

Value of the quality difference = difference between the estimated prices

- Estimate the price for the old/not modernized dwelling and the new/modernized dwelling with the regression equation
- Value of the quality difference:

$$\Delta Q = estP_{n,t} - estP_{m,t}$$
$$= \alpha_0 * (e^{\beta_0 + \sum \beta_i \ln x_{i,n} + \sum \gamma_j d_{j,n}} - e^{\beta_0 + \sum \beta_i \ln x_{i,m} + \sum \gamma_j d_{j,m}})$$

+ Increases and decreases of the same degree of quality have the same value
+ Value is independent of the price of the good
– Value depends on variables not relevant to quality, but to the price of the good (need to be constant for both estimations)

Estimation of ΔQ: ratio approach

Value of the quality difference expressed by ratio of estimated prices

- Estimate the price for the old/not modernized dwelling and the new/modernized dwelling with the regression equation
- Value of the quality difference:

$$\Delta Q = P_{m,t} * \left(\frac{estP_{n,t}}{estP_{m,t}} - 1\right) = P_{m,t} * \left(e^{\sum \beta_i \ln \frac{x_{i,n}}{x_{i,m}} + \sum \gamma_j (d_{j,n} - d_{j,m})} - 1\right)$$

+ Value depends not on variables not relevant to quality, but to the price of the good (if constant)
– Value depends on the price of the good
– Increases and decreases of the same degree of quality have different values => downward bias of the index

Hedonic for rentals - conclusion

- Difference approach is used
- Introduction of hedonics in German rental statistics:
  - Modernization: between 2016 and 2018
  - Replacement: 2018 => especially important for the inclusion of newly constructed dwellings
- Lessons learned for other projects regarding the „hedonic quality adjustment method"
  - Use of adjustment factor α_0
  - Treatment of variables influencing the price, but not the quality of a product
  - Influence of dummy variables in a double-logarithmic environment
  - Determining the value of the quality difference: use difference of estimated prices, not ratio

THANK YOU FOR YOUR ATTENTION!

Timm Behrmann
timm.behrmann@destatis.de
Bernhard Goldhammer
bernhard.goldhammer@ecb.europa.eu
www.destatis.de