Federal Statistical Office of Germany 08.05.2017

Timm Behrmann, Bernhard Goldhammer

15th Meeting of the Ottawa Group on Price Indices

New developments in the field of house and rental price indices in German price statistics in the light of the hedonic method

Abstract

In recent years, the German price statistics have seen important developments in the field of house and rental price indices.
House price indices have been developed in German official statistics in the European context. In Germany data about transactions and prices of residential property is gathered electronically from regional Expert Committees for Property Valuation. These committees are official bodies that survey developments of the real estate market and collect data about every single transaction of real estate in Germany. As the purchase of a single dwelling is a rare event and as a single dwelling is unique, a special methodology is used in order to facilitate price comparisons between two periods: the double imputation hedonic method. The article highlights the main challenges regarding the setting up of this new statistic and describes the hedonic model used in detail.
Rents for dwellings are closely related with purchases of residential property in economic terms. The German rental price index as included in the German consumer price statistics has also undergone major revisions during the recent years: A new sample was introduced, being stratified by type of dwelling, region, and type of landlord; an online survey was set up for the first time; the dwelling characteristics that are surveyed were revised and many more issues were touched. One major topic in this endeavour was the introduction of hedonic quality adjustment in situations of replacement and modernization of a dwelling. The article points out the highlights of the new rental price statistic, describes the used hedonic method and discusses some challenges we had to address when employing the hedonic method, like the inclusion of variables affecting price but not quality or the problem of re-transforming a double logarithmic regression result.

Keywords: hedonic methods, quality adjustment, double imputation, residential property, house price index, consumer price index, rents

A Developments in hedonic price adjustment in German House Price Index

A1 Hedonic methods in German price statistics

“Hedonic methods have come to play a growing role in measuring price trends and economic growth in recent years. Germany’s Federal Statistical Office has formulated a step-by-step plan for a detailed examination of the effects and practicability of applying hedonic methods.”[1]

These are the first statements of an article which describes the introduction of hedonic quality adjustment in German price statistics in 2002: applying hedonic methods for desktop PCs in German CPI.

After the use of hedonic methods had been rejected in official statistics of several states at the beginning of the 1990s because of difficulties in practical application, the Boskin report (1996) resulted in a new discussion on quality adjustment and hence on hedonic methods.

Boskin et al. assumed that a quality change bias occurs especially for products with improvements in the product quality because of inaccurate measurement. This report had impact on the use of hedonic methods; e. g. in the US CPI the application of hedonic quality adjustment was extended to Desktop PCs (1998) and TV-Sets (1999) shortly after.

Since these days of extension of hedonic quality adjustment in official price statistics, hedonic methods have been investigated in detail and different guidelines were written by international organizations, e. g.:

- Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes (2006), OECD.
- Handbook on the application of quality adjustment methods in the Harmonised Index of Consumer Prices (2009), Destatis.
- Handbook on Residential Property Prices Indices (2013), Eurostat, ILO, IMF, OECD, UNECE, World Bank.

In the course of analysing hedonic methods, it became apparent that this method is not only appropriate for products which continually improve due to technical advancement but also for products with high complexity of the characteristics and for those with many product features. Eurostat stated hedonic methods as “A-method” of quality adjustment e. g. in the Handbook on Residential Property Prices Indices (2013), page 159:

“Subject to data being available on salient housing characteristics, the hedonic regression method is generally the best technique for constructing a constant quality residential property price index. The imputations approach to hedonic quality (mix) adjustment has advantages over the time dummy approach. Stratified hedonic indices are preferred over a straightforward application of hedonic regression to the whole data set.”

In German price statistics the application of hedonic methods was expanded step by step, too. While at the beginning of the 21st century hedonic methods have been adopted primarily for IT products, meanwhile these methods have been improved and adopted also for mobile phones, rents and residential property.

---

[1] See Linz, Eckert (2002).

Price index	Products with Hedonic quality adjustment[2]
Consumer Price Index	Desktop PCs (2002) Notebooks (2003) Used cars (2003) Tablet PCs (2013) Rents (2016)
Index of Import Prices Index of Wholesale Prices Index of Producer Prices Index of Export Prices	Desktop PCs (2004) Notebooks (2004) Servers (2004) Processors (2004) RAM (2004) HDD (2004) Printers (2004) Tablet PCs (2013) Mobile phones (2013)
Residential Property Price Indices	House Price Index (2004) Owner-Occupied Housing Price Index (2010)

Table A1: Application of hedonic methods in German price statistics

In German price statistics, hedonic methods have become a well-established quality adjustment method especially for complex products and for products with high technical progress. In these cases hedonic methods have provided high quality results in monthly or quarterly index calculation – under certain conditions.

The application of hedonic methods in statistical agencies requires well qualified staff with a detailed knowledge of regression analysis and knowledge of the product and its market conditions. Furthermore, comprehensive and reliable information on prices and product features is required. And a sufficient number of observations is needed. Within the framework of IT products we have made the experience that a minimum number amounting to 80-100 observations for each product per period seems to be reasonable – having in mind that the data situation for IT products is excellent because all required product features are available on the internet. The situation for residential property is significantly worse.

These are not negligible burdens. However, it must be taken into account that any quality adjustment for complex products is complex itself. And without detailed information on product features it is hard to calculate any quality adjustment method.

In the following the application of hedonic methods in German price statistics will be illustrated using the example of residential property.

A2 House Price Index: requirements of the EU

To improve measurement of price development for residential property on European level Eurostat took concrete steps since the turn of the millennium. Pilot projects with financial assistance of Eurostat started in 2002. The main goal was to build up a system of harmonized data on residential property price indices on EU level. Price indices for owner occupied housing (OOHPI) should close the gap in the coverage of residential living: in the Harmonized Consumer

[2] For more details, see Linz, Eckert (2002) / Linz, Dexheimer, Kathe (2003) / Linz, Behrmann, Becker (2004) / Behrmann, Kathe (2004) / Goldhammer (2016).

Price Index rents are covered, but costs of owner occupied housing are not. OOHPI measures the changes in the transaction prices of dwellings new to the household sector and of other goods and services that households acquire in their role as owner-occupiers.[3] Indices are available since 2010 at least, for few states since 2005, see Eurostat's Homepage for details.

With the financial crisis in 2007/2008 another aspect of housing became important: the price changes of residential properties purchased by households. As a consequence the House Price Index (HPI) was developed and published independently of OOHPI. The HPI developed to an important indicator for monetary policy on European level (EZB), macro-prudential monitoring (ESRB) and within the macroeconomic imbalance procedure (MIP). The HPI measures the changes in the transaction prices of dwellings purchased by households.[4]

Since 2013, there is a legal obligation to calculate house price indices and owner occupied housing price indices, see Commission Regulation (EU) No 93/2013. In addition, price indices for residential property are part of the regulation on Harmonized Indices of Consumer Price: Regulation (EU) 2016/792. A general overview of the conceptual and practical issues related to the compilation of an HPI is given in the 'Handbook on Residential Property Price Indices (RPPIs)'. The methodological manual Commission Regulation (EC) No 93/2013 refers to is revised at the moment. It covers the most important principles and aspects of the methodological framework for OOHPI and HPI.

Some of the methodological and organizational requirements are shown below:

Methodological requirements

- Calculation of a price index
- Application of the net acquisition approach
- Use of transaction prices, no offer prices or appraisals
- Prices include not only the building but also the site area
- Recommended quality adjustment method: hedonic methods

Organizational requirements

- Member states compile national HPIs
- Indices for newly built and existing properties
- Eurostat calculates HPIs on EA and EU level as weighted averages of the national HPIs
- Indices are calculated quarterly with a delivery deadline of t+85
- Yearly update of the weighting scheme
- Data published for 27 member states of the EU (without Greece) + Norway + Iceland
- Data published since 2008 at least, for 14 states since 2005

A3 German circumstances

In Germany, the change of ownership of real estate has to be certified by a notary. The notary is obliged to send copies of the selling/buying contract to the local tax office and to the local Expert Committees for Property Valuation (ECPV).

[3] See Commission Regulation (EU) No 93/2013 Article 2.1.
[4] See Commission Regulation (EU) No 93/2013 Article 2.2.

ECPVs are obliged by law to register and analyse data on transactions of buildings, flats and plots of land in their respective area of responsibility in order to increase market transparency. The main tasks of these Expert Committees consist on the one hand in collecting data on the purchase of buildings and dwellings and on the other hand in estimating current market values of dwellings and land. The committees are also responsible for the calculation of so called Standard Land Values for land and properties. Standard Land Values are estimates of local average land prices per m² provided by the ECPVs. Moreover, these committees maintain a database on prices regarding the purchase and sale of buildings and plots of land (sales price collection). Normally, annual reports are supplied on developed and undeveloped property as well as on the land market in general.

The ECPVs are supplied a copy of each contract connected to a dwelling transaction by the notaries, and enrich these data with further information. Questionnaires are sent to the seller to be filled out on a voluntary basis. Another way to enrich the data is to contact the cadastral land register and other administrations to receive additional data. The ECPVs then store the transaction data in their sales price collection.

For HPI calculation the administrative data from the Expert Committees for Property Valuation is used and surveyed quarterly. But it has to be considered that all of these ECPVs are independent and based on federal (not statistical) law. Furthermore, until the adoption of Commission Regulation (EU) No 93/2013, they were not obliged to provide statistical data to the Federal Statistical Office on the purchase of buildings.

Unfortunately there are differences in data availability between regions and federal states due to the federal construction of Germany. ECPVs are mainly organised on NUTS-3 level (district area level), partly even on LAU1/LAU2-level (municipality level). Thus in one federal state a full geographical coverage seems not to be possible because the ECPVs are mainly organised on LAU1/LAU 2 level so that nearly 1,000 of a total of 1,380 German ECPVs are installed in this one federal state. To handle this, for HPI calculation only 100 out of nearly 1,000 ECPVs have been selected by cut-off sampling using the amount of 20,000 inhabitants as a minimum level. For the other 15 federal states a full geographical coverage (all ECPVs within these federal states provide data) is aimed at. Based on this approach, the HPI for entire Germany is based on about 460 ECPVs and achieves a regional coverage of 95%. With Commission Regulation entered into force 2013 it was possible to collect data of all of these 460 ECPVs.

In addition in 2012 the system of data collection has been improved. Until then, one of the main problems resulted from the fact that the Expert Committees did not follow harmonised classifications for their data collection. Since then the data have been collected by a standardised IT-System (.core) to make the data collection more efficient and to ensure a safer way of data transfer. For the data transfer, the ECPVs have to follow a data record description. This description specifies the characteristics that have to be reported, defines characteristics values and the formal structure of the transmitted data. An overview of the survey characteristics is given in chart A1. In consequence, all data that are transmitted have the same structure and are available in a standardised file-format.

Chart A1: Overview of survey characteristics

A4 Hedonic methods

As is well known, price indices are supposed to measure pure price changes. Changes in quality and quantity should be excluded. So in case of product replacements quality adjustment is used by determining the monetary value of the change in quality. Besides “conventional” methods hedonic methods are one way of quality adjustment. Nowadays these methods are well acknowledged and used in various countries.

Houses and flats are products which are well suited for applying hedonic methods: Houses are unique objects with very heterogeneous characteristics and houses are sold rarely, so that a direct price comparison of identical properties is not possible. Hence “normal” price statistics by observing prices of the same product in different points of time is not possible. In this case hedonic methods can help to calculate meaningful price indices. However, real estate markets are not very transparent and it is very difficult to receive information on property characteristics. This is a problem for all quality adjustment methods, also for hedonic calculations.

To calculate price indices quality adjusted with hedonic methods two steps are necessary. First, a functional relationship between the price and the product features has to be found and modelled in a regression. Secondly, regression results have to be used for index calculation. There are different index calculation methods: the hedonic imputation method and the hedonic time dummy variable method are mentioned by the “Handbook on Residential Property Prices Indices”. The “Handbook on the application of quality adjustment methods in the HICP” mentions additionally the hedonic characteristics price method and the re-pricing method. The German HPI is calculated with the double imputation method, a special case of the hedonic imputation method.

A4.1 Regression analysis

As a first step a functional relationship between the price and the product features has to be derived. According to standard practice and as recommended, we use standard OLS (ordinary least square) regression instead of WLS (weighted least square) regression. This seems reasonable because our data base is almost constructed as a complete survey. Furthermore it is recommended to stratify the sample, to use hedonic methods within each strata and to weight explicitly afterwards.[5] For calculation of German HPI we stratify the sample into four strata and we calculate a regression function for each stratum, in each case for the whole of Germany:

- Existing houses,
- Existing flats,
- Newly built houses,
- Newly built flats.

In addition to the choice of the regression method several questions have to be answered: the functional form, the set of exogenous variables and the set of criteria to evaluate the goodness of fit. Below we show which decisions we made and why we did so.

A4.2 Functional form

A **log-log function** is used to calculate the regression equation. This is a more appropriate form than the linear or the simple log function, because non-linear relationships between the price and specific features as well as relationships between specific features can be mapped more accurate.

As can been seen in a simple graphical analysis the relationship between the price and the living space of existing houses is not linear. The use of OLS is more accurate if the price and the living space have been transformed by taking their natural logarithms.

Chart A2: Scatter plot, existing houses 2016Q4

Considering the scatter plots more precisely, it is obvious that in the case of not transforming the continuous variable by taking the natural logarithms problems with **heteroskedasticity** can occur. Such problems occur oftentimes within the scope of hedonic regression. Using a log-

[5] The recommendations are based on the "Handbook on Residential Property Prices Indices" (2013).

transformation these problems could be improved significantly. The same holds true considering the assumption of **normal distribution**.

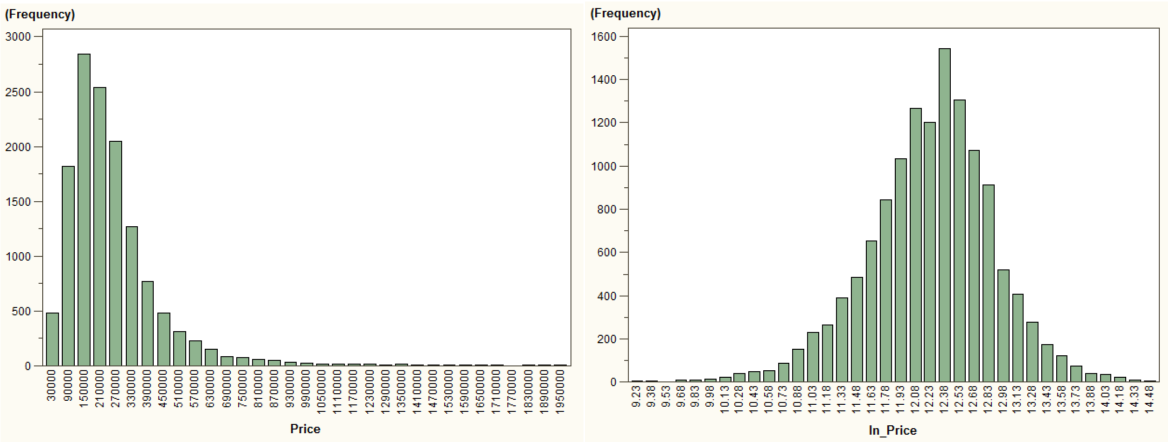

Chart A3: Distribution of prices, existing houses, 2016Q4

A4.3 Set of variables

An important issue of regression analysis is the choice of the set of explanatory variables. To decide exogenous variables should enter the regression function we consider different criteria. First (using a stepwise regression) we take a look at the **level of significance** of the different variables. Derived from the Final Prediction Error we use a significance level amounting to 15.7%. But this criterion doesn't become effective because all variables which are considered in the regression function are significant on a level of 0.1%.

A measure to evaluate the whole model as well as the impact of one single variable is the **adjusted R^2** and the **partial R^2**, respectively. In general, we use a spare parameterization so that only variables which increase the adjusted R^2 remarkably (1-2 percentage points) enter the regression model. Although variables like the legal form of purchaser and seller or the type of building are partly statistically significant there aren't considered in the regression function because the additional explanatory power is too small.

Unfortunately, **partial non-response** is present in many data sets. That is, the information on one or more characteristics may be missing for a part of the sample. So there is a trade-off: either all possible variables are used for regression analysis and the observations with missing information get lost or all observation could be used and some variables can't enter the regression function. Even in these cases we use a spare parameterization and we prefer many observations instead of many regression variables. Variables like cellar, parking space or type of building area don't enter the regression function because of partial non-response. However, the explanatory power of these variables doesn't exceed one percentage point.

Even **multicollinearity** is a well-known problem in hedonic regressions. A high correlation between some of the included variables increases the standard errors of the regression coefficients; the coefficients become unstable.[6] The Variance Inflation Factor (VIF) provides information concerning multicollinearity. But similar to the significance level there are different hints in scientific literature but no agreement about the critical value of VIF. Based on our experience using hedonic regression for IT products, we use an upper limit amounting to 3. But in

[6] RPPI Handbook (2013) p. 51

context of German HPI no problems concerning multicollinearity occur in the log-log regression function: VIF doesn't exceed the value two.

Following this approach the resulting regression function for existing houses is shown below. The price and the continuous exogenous variables have been transformed by taking their natural logarithms. Additionally, dummy variables for the Standard Land Value (SLV) enter the regression function which classifies the quality of the location. The standard land value is converted into dummy variables (up to 10) to avoid the inclusion of prices in the exogenous variables. Although the exogenous variables represent only the basic features of a residential property this model explains 70-80% of the variation in prices. The R^2 the four types of residential property (house/flat, existing/new) amounts between 70-80%.

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

Formula A1: German HPI hedonic regression function for existing houses

Chart A4 shows the result of OLS regression for existing houses 2016Q4. The adjusted R^2 accounts for about 73%. Furthermore the regression coefficients (Parameterschätzer) are highly significant and no problems with multicollinearity occur.

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

Chart A4: Regression analysis for existing houses 2016Q4

Another quality issue could only be considered over time. The coefficients of the variables should be plausible and stable over time. That this is true for German HPI regression analysis as can be seen in chart A5.

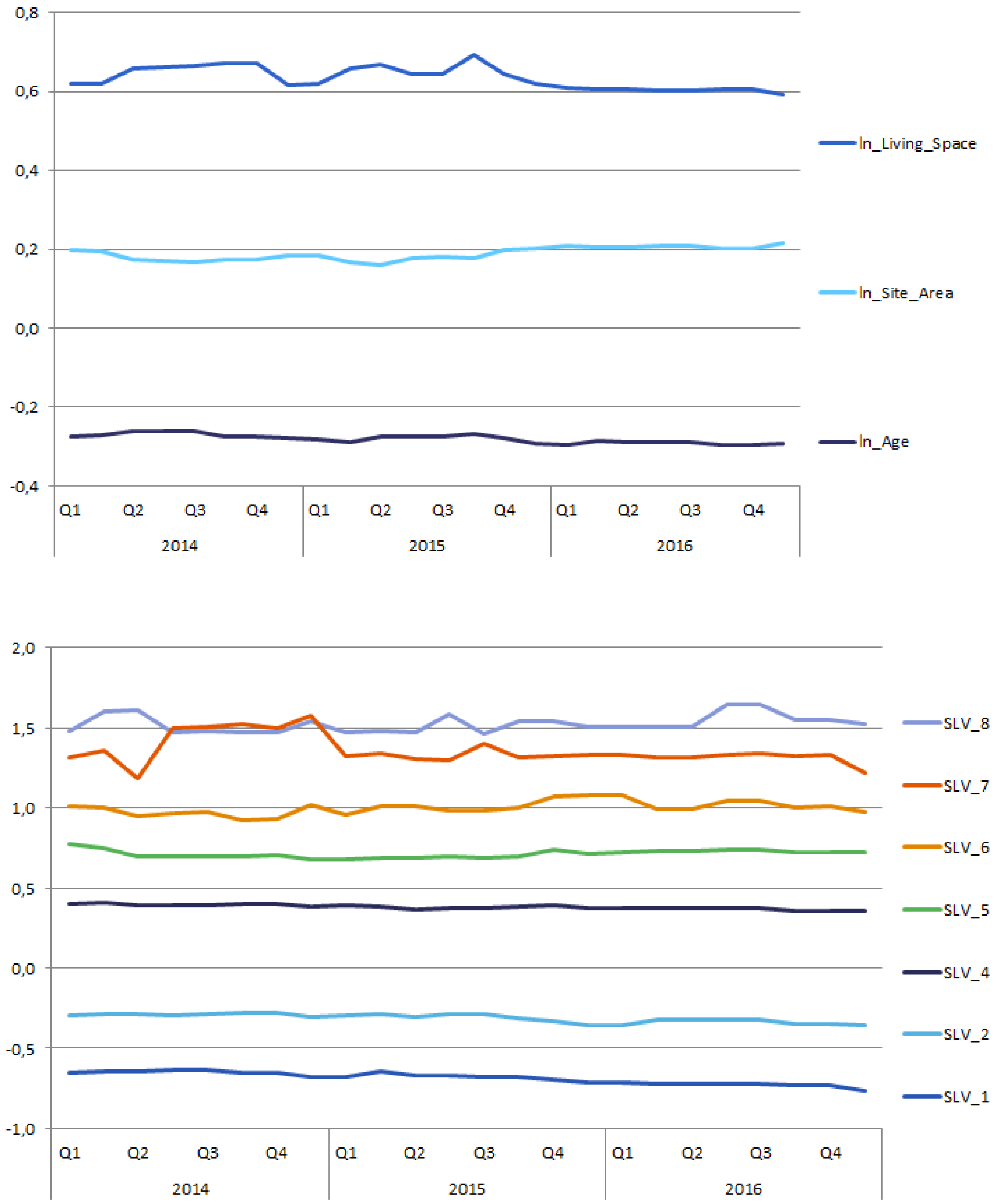

Chart A5: Regression coefficients for existing houses 2014-2016

A4.4 Index calculation

Before calculating the hedonic price index, the quarterly data on the prices and features of the chosen products is divided into two sample subsets A and B.

Sample subset	Base period t=0	…	Current period t=T
A: Old products	p_0^A	…	-/-
B: New products	-/-	…	p_T^B

Table A2: Sample subsets

The "old" products were included in the sample in the base period, but they cannot be observed in the current period. A price is imputed for these products for the current period: this is the average that would have to be paid for a product of the old quality under current market conditions, estimated by regression analysis for the current period.

The "new" products were not available in the market during the base period, but have been observed in the current period. A base price is imputed to these new products: this is the price that would have been paid for a product of the current quality under the market conditions prevailing in the base period, estimated by regression analysis for the base period.

Sample subset	Base period t=0	…	Current period t=T
A: Old products	p_0^A	…	$\widehat{p_T^A} = f(X^A, \widehat{\beta_T})$
B: New products	$\widehat{p_0^B} = f(X^B, \widehat{\beta_0})$	…	p_T^B

Table A3: Single Imputation Method

Regression analysis is used to impute current prices for old products and base prices for new products (Single Imputation Method). However, when it comes to practical application, there are advantages to comparing the imputed, estimated price with a price that has similarly been estimated. To do this, the regression is also performed for the period for which we have observed a price. This estimated price replaces the observed price (Double Imputation Method).

Sample subset	Base period t=0	…	Current period t=T
A: Old products	$\widehat{p_0^A} = f(X^A, \widehat{\beta_0})$	…	$\widehat{p_T^A} = f(X^A, \widehat{\beta_T})$
B: New products	$\widehat{p_0^B} = f(X^B, \widehat{\beta_0})$	…	$\widehat{p_T^B} = f(X^B, \widehat{\beta_T})$

Table A4: Double Imputation Method

The German HPI is calculated using the Double Imputation Method. Currently we use an adjacent periods approach so that the current price development is calculated related to the previous quarter. Although this approach is not without controversy because of a possible index drift we have decided to do so. In particular at the time regulation 93/2013 came into force every quarter some ECPVs provided data for the first time. To include these "new" ECPVs as soon as possible in index calculation we decided to apply the adjacent periods approach. To avoid a bias index calculation is done only with ECPVs which have provided data for both quarters: the current and the previous quarter.

The overall price index is calculated as the average price movement for old and new products together. At the first stage of aggregation we use the unweighted geometric mean or Jevons index to aggregate individual price quotations into an index.

$$I_{0,T}^{A} = \frac{\sqrt[n]{\prod \widehat{p_T^A}}}{\sqrt[n]{\prod \widehat{p_0^A}}} = \sqrt[n]{\prod_{n} \frac{\widehat{p_T^A}}{\widehat{p_0^A}}}$$

Formula A2: Calculation of an index for the old products

$$I_{0,T}^{B} = \frac{\sqrt[m]{\prod \widehat{p_T^B}}}{\sqrt[m]{\prod \widehat{p_0^B}}} = \sqrt[m]{\prod_{m} \frac{\widehat{p_T^B}}{\widehat{p_0^B}}}$$

Formula A3: Calculation of an index for the new products

$$I_{0,T} = \sqrt{I_{0,T}^{A} \cdot I_{0,T}^{B}}$$

Formula A4: Calculation of an overall index

A5 Results

At the end we want to show some results for the House Price Index of Germany and other European countries. The price development for residential property is very different in Europe depending among other things on the effects of the financial crisis in 2007/2008. Only looking at the five great countries there are significant different price developments from 2010 to 2016Q4. While prices in Germany rose by 29.6% and in UK by 26.5%, they remained nearly unchanged in France (+2.2%) and decreased in Italy (-14.5%) and Spain (-21.4%). The difference of the development of HPI between Germany, Spain and the EU is shown in Chart A6.

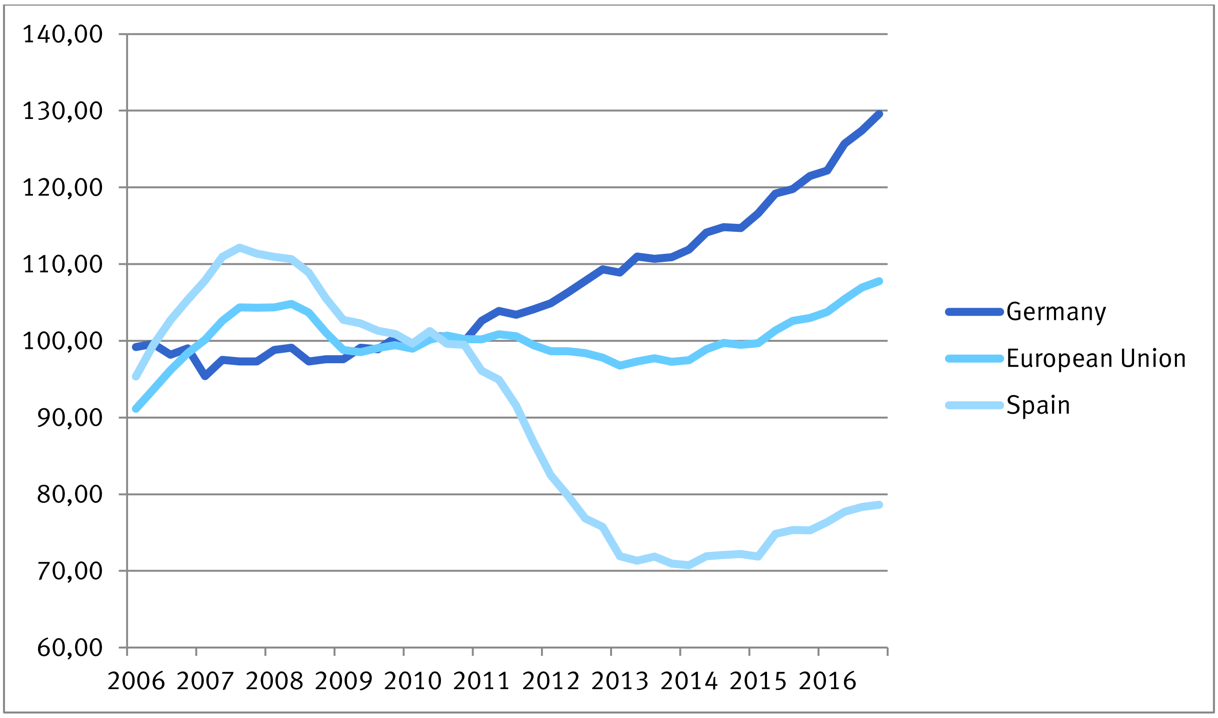

Chart A6: House Price Indices of EU, Germany and Spain

A References

Behrmann, T. / Kathe, A. (2004): Zur Anwendung hedonischer Methoden beim Häuserpreisindex, Wirtschaft und Statistik 5/2004, English version: Using Hedonic Pricing for the German House Price Index, accessed 08/05/2017: https://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/Prices/HedonicHousePrice.pdf?__blob=publicationFile

De Haan, J. / Diewert, W. E. et al. (2013): Handbook on Residential Property Prices Indices, Eurostat.

Goldhammer, B. (2016): Die neue Mietenstichprobe in der Verbraucherpreisstatistik, Wirtschaft und Statistik, 5/2016.

Linz, S. / Eckert, G. (2002): Zur Einführung hedonischer Methoden in die Preisstatistik, Wirtschaft und Statistik 10/2002, English version: Introducing hedonic methods in price statistics, accessed 08/05/2017: https://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/Prices/HedonicPC.pdf?__blob=publicationFile.

Linz, S. / Dexheimer, V. / Kathe, A. (2003): Hedonische Preismessung bei Gebrauchtwagen, Wirtschaft und Statistik 6/2003, English version: Hedonic Methods of Price Measurement for Used Cars, accessed 08/05/2017: https://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/Prices/HedonicUsedCars.pdf%3F__blob%3DpublicationFile

Linz, S. / Behrmann, T. / Becker, U. (2004): Hedonische Preismessung bei EDV-Investitionsgütern, Wirtschaft und Statistik 6/2004, English version: Hedonic Price Measurements for IT Products, accessed 08/05/2017: https://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/Prices/HedonicIT.pdf?__blob=publicationFile.

Linz, S. et al. (2009): Handbook on the application of quality adjustment methods in the Harmonised Index of Consumer Prices, Destatis.

Triplett, J. (2006): Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes (2006), OECD.

B Developments in hedonic price adjustment in German rental index

With a weight of 21% in the German CPI and 10.4% in the HICP, rentals for dwellings have a large importance on the overall development of German consumer price indices. A proper methodology is therefore crucial for a meeting the high-quality standards of the indices. Unlike traditional fields of the application of hedonic techniques, the characteristics of dwellings change only at a slow pace; but the variety of dwellings is enormous, which makes it difficult to compare two different dwellings in terms of prices. Here, hedonics can make their stance. In the recommendation on the treatment of actual rentals in the HICP (Eurostat 2014), hedonic quality adjustment is considered an A-method if quality adjustment is needed. This situation can arise, when the dwelling is modernized or when a dwelling in the sample needs to be replaced by another dwelling not yet surveyed.

B1 Need for a hedonic approach

So far, in the few cases of replacement, the new dwelling entered the sample in a way neutral to the index. If a modernization was carried out, the price of the quality difference was calculated by using data on the investment necessary for the modernization. With the new recommendation of Eurostat coming into force that defines for the first time for rents when quality adjustment should be used, it was decided to develop a hedonic approach, as the procedure applied in replacement situations does not meet the methodological requirements anymore.

B2 The German rental sample and the dwelling characteristics surveyed

The German sample on rentals for consumer price statistics covers 20,000 dwellings which where sampled randomly from a register that was built up for the census on buildings and dwellings in the year 2011 (for an in-depth description of the sample and the process of setting it up, see Goldhammer 2016). The sample is stratified according to three dimensions:

- Type of dwelling (seven characteristics):
  - Appartments with less than 70 m², built until 1948
  - Appartments with more than 70 m², built until 1948
  - Appartments with less than 70 m², built from 1948 on, privately financed
  - Appartments with less than 70 m², built from 1948 on, publicly financed (social housing)
  - Appartments with more than 70 m², built from 1948 on, privately financed
  - Appartments with more than 70 m², built from 1948 on, publicly financed (social housing)
  - Single-family houses
- Type of landlord
  - Private persons
  - Publicly-owned or cooperative housing company
  - Private sector housing company
- Regional dimension: types of settlement structure within regions

Additionally, eleven characteristics are surveyed for each dwelling. The characteristics can be classified according to their purpose as identifying, modifying, differentiating and quality characteristics. While identifying characteristics are necessary for data processing only, the other characteristics have an impact on the price. Differentiating characteristics are used for the stratification of the sample. Quality characteristics refer to the perceived quality of the dwelling. A modifying characteristics impacts the price only, not the quality. They are displayed together with their suggested price relevance in Table *B1* B1.

Characteristic	Purpose	Price relevance
Address	Identification	No
Dwelling number	Identification	No
Location of Dwelling in the building	Identification	No
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 (m²)	Differentiation/Quality	Yes

Table B1: Surveyed characteristics of dwellings.

Together with the rental price, this is the data that is available for the hedonic regression analysis. For setting up the hedonic equation, decisions have to be taken how to operationalize the variables. As the variables are of continuous, cardinal, and ordinal nature, different approaches are pursued.

B2.1 Variables for type of dwelling

The type of dwelling is delineated by the size of the dwelling, its year of construction, the type of financing and whether it is an apartment or a single-family house. Except the last, which is not a value by itself and which may be seen as a matter of substitution especially in urban areas, all delineation matters can be observed as characteristics of the dwelling as displayed in Table B11. Hence, no own variables were created for type of dwelling.

B2.2 Variables for type of landlord

Type of landlord is a cardinal measure without natural order. For this purpose, dummy variables are common for data modelling (Wooldridge 2009, pp. 225). As type of landlord is an exclusive variable, two variables are sufficient for its modelling. We presume that dwellings owned by private persons have the lowest level of rents compared to those of the other types of landlord; hence, two dummy variables are chosen for representation in the model:

$$\mathrm{v_{ij}}, \text{for dwelling i and landlord type j; j} \in \{2{,}3\}\text{:}$$

$$\mathrm{v_{i2}} = \begin{cases} 1 \text{ if i is owned by a publicly} - \text{owned or cooperative housing company} \\ 0 \text{ otherwise} \end{cases}$$

$$\mathrm{v_{i3}} = \begin{cases} 1 \text{ if i is owned by a private sector housing company} \\ 0 \text{ otherwise} \end{cases}$$

B2.3 Variables for the regional dimension

For administrative purposes, Germany is divided into 401 districts (Kreise), which are grouped into 96 planning regions (Raumordnungsregionen)[7]. For the districts, a classification according to type of settlement structure was developed by the Federal Institute for Construction, Urban and Spatial Research (Bundesinstitut für Bau-, Stadt- und Raumforschung; BBSR 2015). It comprises four types of district:

- Type 1: Metropolitan cities: Cities with a population of more than 100.000 that form an own district.
- Type 2: Urban districts: Districts where at least 50% of the population live in large- or medium-sized cities and with a population density of at least 150 persons per km², as well as districts without large- or medium-sized cities, but a population density of at least 150 persons per km².
- Type 3: Rural districts with higher population density: Districts where at least 50% of the population live in large- or medium-sized cities, but have a population density of less than 150 persons per km², as well as districts where less than 50% of the population live in large- or medium-sized cities and a population density without large- and medium-sized cities of at least 100 persons per km².
- Type 4: Rural districts with lower population density: Districts where less than 50% of the population lives in large- or medium-sized cities and where the population density without large- and medium-sized cities is less than 100 persons per km².

As we expect that both region and type of district have separate effects on the rental price, three dummy variables are used for the district types (district type 4 is seen as the base level) and 95 dummy variables indicating the planning region (the base level is set for region 1402, Oberlausitz-Niederschlesien):

$$\mathrm{d_{ik}}, \text{for dwelling i and district type k; k} \in \{1,2,3\}:$$

$$\mathrm{d_{ik}} = \begin{cases} 1 \text{ if i is located in a district of district type k} \\ 0 \text{ otherwise} \end{cases}$$

$$\mathrm{r_{il}}, \text{for dwelling i and region l; l} \in \{\text{all planning regions in Germany}\}:$$

$$\mathrm{r_{il}} = \begin{cases} 1 \text{ if i is located in region l} \\ 0 \text{ otherwise} \end{cases}$$

It can be expected that rentals located in a district of type 1 are higher than in districts of type 2; rentals for type 2 higher than for a type 3 district; and type 3 may indicate, all other things being equal, higher rents than type 4. The assumptions for the coefficients are henceforth:

$$\beta_{d1} > \beta_{d2} > \beta_{d3} > 0$$

Regarding the regions, the expectations about coefficient values are not so clear-cut, but large coefficients are expected in regions with high rents like 910 (Munich), 201 (Hamburg), 604 (Rhein-Main – the region around Frankfurt) and 810 (Stuttgart),[8] while the coefficients will be of

[7] An overview of the regions can be accessed on https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis/NichtAdministrativ/Aktuell/31Raumordnungsregionen.html (accessed 08/05/17).
[8] The coding of the region consists of the code of the respective *Bundesland* followed by a two-digit number for the region.

comparably small size for 911 (Oberfranken-Ost), 914 (Oberpfalz-Nord –both regions belong to the economically rather weak North-East of Bavaria), and several regions of former Eastern Germany.

B2.4 Variables for the quality of the dwelling

B2.4.1 Energy consumption/need

This characteristic represents the energetic quality of the building in which the dwelling is located. It is measured in kWh/(m²*a) and tells the energy used per m² within a year for heating purposes. It can either be calculated from actual consumption values (energy consumption approach) or based on theory following a standardized engineering approach (energy need approach). Since 2014, it is mandatory in Germany to calculate this figure and show it to tenants in form of a so-called "Energieausweis" (energy performance certificate), with some exceptions e.g. for listed buildings. It is a continuous variable. A negative coefficient seems to be plausible, because higher energetic buildings are of less perceived quality to the user, who has to pay larger energy bills if living in such a building. The variable representing energy consumption/need is called en_i.

It was also investigated whether the type of energy performance certificate has an influence on the rental price. Without making any judgement on the coefficient, a dummy variable indicating the use of "energy consumption" for the certificate was tested, referring to as ec_i.

B2.4.2 Quality of residential area

This indicator describes the quality of the area where the dwelling is located, as perceived by the users. It is also used in the so-called "Mietspiegel", a municipality-based inventory displaying the average rent charged for dwellings with a certain age, furnishing and size in areas with a certain quality. The distinction of the quality levels varies between municipalities; however, four levels as used in this approach are very common, ranging from simple, normal, good to very good. In the rental sample, these qualities are coded with 1, 2, 3, and 4. As the quality of the residential area is an ordinal variable, it is operationalized with three dummy variables, indicating levels 2, 3, and 4:

$$\mathrm{qa_{im}}, \text{for dwelling i and quality of residential area m}; \mathrm{m} \in \{2,3,4\}:$$

$$\mathrm{qa_{im}} = \begin{cases} 1 \text{ if i is located in an area of quality m} \\ 0 \text{ otherwise} \end{cases}$$

As dwellings in an area of high residential quality should have a higher rent than those in areas of lower quality, the following assumptions can be made regarding the coefficients:

$$\beta_{qa4} > \beta_{qa3} > \beta_{qa2} > 0$$

B2.4.3 Quality of furnishing of dwelling

The quality of furnishing of the dwelling refers to the quality of the dwelling itself, not the surroundings. For example, a dwelling with a second bathroom, an elevator, parquet floor and triple glazing has a higher quality of furnishing than one with just one bathroom, no elevator, carpet, and single glazing. Like the quality of the residential area, it is used in the "Mietspiegel" – again with individual operationalization in different municipalities. The approach used for the rental price index is derived from an official guideline for the value estimation of real estate that is issued by the Federal Department of Construction (BMVBS 2012, Tabelle 2). As in this source, three levels are distinguished: simple, normal, and good. They are coded with 1, 2, and 3. As the quality of furnishing of a dwelling is an ordinal variable, it is operationalized as two dummy variables, indicating levels 2 and 3:

$$qf_{in}, \text{for dwelling i and quality of furnishing n; } n \in \{2,3\}:$$

$$qf_{in} = \begin{cases} 1 \text{ if i is a dwelling with furnishing quality n} \\ 0 \text{ otherwise} \end{cases}$$

As dwellings with furnishing of good quality should have a higher rent than those with simple furnishing, the following assumption can be made regarding the coefficients:

$$\beta_{qf3} > \beta_{qf2} > 0$$

B2.4.4 Other furnishings: balcony, garage, fitted kitchen

There are some other furnishings, which may have an impact on the rental price as tenants have an additional benefit of it, but which are not reflected in the quality of furnishing pointed out in the previous chapter. Three points are recorded in this characteristic:

- Balcony: It provides an additional recreational benefit to the tenant. Similar features are also subsumed under the term “balcony”, like gardens, terrace, or loggia.
- Garage or parking space: a shelter for a car – or at least a place where it can be parked without paying extra fee and without exhausting search – is another extra benefit to the tenant. No differentiation is made regarding the kind of parking space.
- Fitted kitchen: As a fitted kitchen with including electrical appliances is provided by the landlord, the tenant does not need to invest in one by him-/herself.

As all three devices add benefits to the tenant, the willingness to pay more should be prevalent. Hence, the model comprises three dummy variables, one for each point mentioned:

$$bc_i = \begin{cases} 1 \text{ if i is a dwelling with a balcony or a similar device} \\ 0 \text{ otherwise} \end{cases}$$

$$gar_i = \begin{cases} 1 \text{ if i is a dwelling with a garage or a similar device} \\ 0 \text{ otherwise} \end{cases}$$

$$fk_i = \begin{cases} 1 \text{ if i is a dwelling with a fitted kitchen} \\ 0 \text{ otherwise} \end{cases}$$

It is expected that, if significant, all three variables have positive coefficients.

B2.4.5 Year of construction

For the construction of dwellings, every age had its special legal requirements and fashions. For example, many dwellings built in the second part of the 19th century and early 20th century have ceilings as high as 3.50 m, while dwellings built after World War II only reach 2.50 m; buildings from the 50th had to be erected very fast to give shelter to people that were either bombed in World War II or refugees from former Eastern parts of Germany, resulting in quite low construction quality; energetic regulations on buildings were only to be introduced by 1978 and have become largely stricter since then. These are some reasons for the assumption that the year of construction is a price-determining quality characteristic of a dwelling. It is also used for that purpose in the Mietspiegel. Hence, this characteristic is also considered for the hedonic model.

As it is not exactly the year of construction ycons_i, but rather the age that is supposed to determine the price, in a first step the age of the building is calculated:

$$\text{age}_i = 2017 - \text{ycons}_i$$

The variable age_i is entered in the hedonic regression. A negative coefficient is expected.

Additionally, it should be mentioned that the year of construction is not only a quality characteristic, but also used for differentiation, because the sample is structured in two classes according to the year of construction being before or after 1948. However, this is not reflected in the hedonic approach.

B2.4.6 Living space

The living space in m² is without a doubt the most important factor determining the height of the rent. For example, in a Mietspiegel, the values given as usual rents for a certain category of dwellings (according to age, residential area quality, furnishing quality) are expressed in €/m². Hence, the continuous variable ls_i is used in the regression model, expecting to have a strongly positive value. Like year of construction, living space is also a differentiation characteristic dividing the sample in dwellings up to and larger than 70 m². However, this is not reflected in the hedonic approach either.

B2.5 Other variables influencing the rental price

There are two more variables that are collected and are supposed to have an influence on the level of the rent as well as its development: the first is the date of the rental contract; the second is the type of funding.

The date of the rental contract states when the current rental agreement has begun. Research has proven that long-lasting rental contracts have a lower rental level than recently closed contracts (Hoffmann/Kurz 2002). This phenomenon is called the "sitting discount". Hence, we expect the rent for apartments with longer-lasting contracts to be lower than that for a recently signed contract. For the hedonic regression, the date of the rental contract is operationalized by the variable dc_i, which is the duration of the contract in days:

$$\text{dc}_i = \text{today} - \text{dateContract}_i$$

The type of financing of the building determines the price mechanism: for an apartment that has been financed without public funding (called "free-financed"), the rent is determined by the market forces (of course, within the limitations of legislation); for publicly funded dwellings, a cost approach that is prescribed by law determines the rent, which has to be applied during the funding period.[9] As a market-determined rent that includes a profit margin to the landlord should be higher than a cost-based rent, we suppose publicly funded dwellings to have lower rents. Hence, a dummy variable ff_i is considered for the inclusion in the hedonic model that indicates free-financed dwellings, for which we expect a positive coefficient.

We would like to stress that the variables dc_i and ff_i are of a special nature, as they determine the dwelling's rental price, but do not alter the quality of the dwelling as it is perceived by the consumer. This has consequences for hedonic quality adjustment upon which we touch later.

---

[9] The end of the funding period is also provided as a characteristic, but has no relevance for the rental level. It just determines the point in time when the price mechanism changes.

B3 The hedonic approach and results

Now, with the hedonic method, we try to combine all price-relevant categories into one regression model. A double-logarithmic model, as very common to price statistics, is used for modelling the relationship:

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

For the parameter estimation, OLS regression is used and the stepwise procedure applied. The regression covers only those observations for which all characteristics were available; incomplete observations were deleted from the data set.[10] The estimation was calculated for the first time in 2015 on a test basis with the December 2014 data for only two out of 16 Bundesländer. Only for those, the new sample had already been implemented and over 60% of the observations could be used, as their characteristics were fully provided. Only one regression equation was estimated; the influence of different Bundesländer is to be loaded on the regional and district type dummies. The exercise was repeated in 2016 (now for 7 Bundesländer) and 2017 (for 12 Bundesländer). The increasing number of covered Länder illustrates the progress in the implementation of the new sample and in surveying all characteristics. In the following, the results for the 2017 estimation are presented. Table B2 contains the overall results of the regression as well as the coefficients of the variables that turned significant at the 5% level. For ease of illustration, out of the regional dummy variables, only the two with the highest and the two with the lowest coefficient are displayed; the other regional dummy variables and their coefficients are displayed in Annex I.

Results of 2017 hedonic regression model for rents

Dependent variable	$\ln p_i$
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 (σϑ	0.187

[10] No rule without exception: as an energy performance certificate is not mandatory for listed buildings, missing values for buildings with a year of construction until 1930 were imputed with the German average on energetic quality for the year 2013, 133.6 kWh/(m²*a). This figure is provided by the real estate association GdW, see http://web.gdw.de/uploads/pdf/infografiken/15.10.2014/Energieverbraeuche.pdf (accessed 08/05/17).

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
…[11]		
$r_{i,911}$	-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
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

Table B2: Regression results for rents.

[11] Out of 96 regions with 95 dummy variables, 56 turned significant at the 5% level. For the other regions, the rent level does not differ significantly from such of region 1402 (Oberlausitz-Niederschlesien), everything else being equal.

For almost every variable, the results are in-line with the expectations. Some exceptional results are:

- Regarding type of landlord, the lowest rental level is not observed for private-owned dwellings, but for dwellings with public and cooperative ownership.
- Regarding type of district, no difference was observed for the level of rent between the two types of rural areas. However, as hypothesized, urban areas and especially major cities have a higher rental level.
- The dummy variable indicating the use of "energy consumption" for the energy performance certificate ec_i turned not significant. Hence, the two ways of calculation the energetic quality of the building seem to lead to comparable results.
- The only disturbing finding was the non-significant coefficient for the garage (gar_i), which was even slightly negative. As, in the 2016 exercise, gar_i showed a clearly positive and significant coefficient in line with the expectations, we assume that the result depends on wrong coding of the variable gar_i in the CPI production environment for dwellings which were not included in the calculation in 2016. This issue is subject to further investigation. As, however, some hedonic coefficient is necessary for quality adjustment (and the inclusion of a garage in the rental agreement is an improvement in quality without a doubt), the coefficient calculated in 2016 was taken over: $\beta(gar_i) = 0.033$.

B4 From double-logarithmic regression equation to quality adjustment

B4.1 Transforming double-logarithmic estimation results into level results

So far, we have estimated only the natural logarithm of the rental price. Hence, it needs to be re-transformed to get the price on a linear scale. For this task, it is important to remember that the model for which the parameters are estimated includes the residual factor $\mathrm{u_i}$, which is supposed to be distributed $\mathrm{N}(0, \sigma^2)$. Therefore, to obtain $\mathrm{p_i}$ from the estimated value $\ln \mathrm{p_i}$, it is necessary to recall that the expected value of $\exp(\mathrm{u_i})$ is not 0, but $\exp(\frac{\sigma^2}{2})$ (Wooldridge2009, p. 211). That means, for the prediction of $\mathrm{p_i}$, the inclusion of an appropriate adjustment factor is necessary:

$$\hat{\mathrm{p}}_\mathrm{i} = \exp(\frac{\sigma^2}{2}) \cdot \exp(\widehat{\ln \mathrm{p_i}})$$

The adjustment factor $\exp(\frac{\sigma^2}{2})$ is also referred to as α_0. Of course, σ^2 is unknown. Hence, for the determination of α_0, Wooldridge (2009, p. 211-212) proposes three different approaches:

1. $\tilde{\alpha}_0 = exp(\frac{\hat{\sigma}^2}{2})$

   where $\hat{\sigma}$ standard error of the regression (also known as root mean square error)
   $\widetilde{\alpha_0}$ will correct the underestimation in the right manner, if $\hat{\sigma}^2$ is an unbiased estimator of σ^2 (which is a property of OLS, if the Gauss-Markov assumptions hold, see Wooldrigde 2009, 102) and if, for the underlying sample, $\hat{\sigma}^2 = \sigma^2$. This estimator of α_0 "is not unbiased, but consistent… However, it does rely on the normality of the error term, *u*." (Wooldrigde, 2009, p. 211) The presence of non-normality distributions of *u* is the motivation for alternative estimators of α_0 which are robust to non-normality.

2. $\hat{\alpha}_0 = \frac{1}{n}\sum_{i=1}^{n} exp(\hat{u}_i{}')$

where $\hat{u}_i{}'$ estimated residual of the log-log regression
Here, the factor α_0 is calculated keeping in mind that it should account for the variation in the error terms u_i assuming that $\hat{u}_i{}'$ is the unbiased estimator of u_i (which is not necessarily the case). However, it is a consistent estimator and has the desired property of values always larger than 1 (see Wooldrigde 2009, 211–212 for a discussion).

3. $\check{\alpha}_0 = \left(\sum_{i=1}^{n} exp(\widehat{\ln p_i})^2\right)^{-1} \cdot \left(\sum_{i=1}^{n} exp(\widehat{\ln p_i}) \cdot p_i\right)$

where $\widehat{\ln p_i}$ estimated logarithmized rental price of dwelling i (without adjustment factor α_0)
p_i actual rental price of dwelling *i*
For this approach, developed by Wooldrigde (2009, p. 212), $\check{\alpha}_0$ is the result of a minimization of the sum of squared residuals of the estimation. Hence, for the data set on which the regression is based, the use of $\check{\alpha}_0$ always minimizes the residuals. A drawback is that α_0, calculated in this way, can lead to values smaller than 1.

We calculated the estimates for α_0 in each of the three ways. The resulting values are presented in Table *B3* B3.

Calculation of $\tilde{\alpha}_0$		Calculation of $\hat{\alpha}_0$		Calculation of $\check{\alpha}_0$	
$\hat{\sigma}^2$	0.18652	$\sum_{i=1}^{n} \exp(\hat{u}_i{}')$	14557.2	$\sum_{i=1}^{n} exp(\widehat{\ln p_i})^2$	2181073645
				$\sum_{i=1}^{n} exp(\widehat{\ln p_i}) \cdot p_i$	2249454886
$\tilde{\alpha}_0$	1.0175	$\hat{\alpha}_0$	1.0170	$\check{\alpha}_0$	1.0313

Table B3: Calculation of correction factor α_0.

For the quality adjustment calculation, $\tilde{\alpha}_0$ was chosen, because the value lies “in-between” the two other estimations.

Now, we are ready for the estimation of a price of a dwelling with the hedonic regression including the adjustment factor α_0. Before a quality adjustment can be done, two more questions have to be answered: how to treat factors that are not relevant for quality, but for the price level; and how a comparison between two estimated prices should be treated mathematically.

B4.2 Treatment of variables relevant to the price but not to the quality

As said before, the variables dc_i (duration of rental contract) and ff_i (type of financing) are relevant to the rental price but not to the quality. If we want to adjust the price for changes in the quality of the dwellings, we have to take this into account. As changes in these two characteristics are not regarded as quality change, they should be treated as price change; hence, both variables should be kept constant, because only in this case, they have no influence on the estimation of the value of the change in quality. But, if we want to determine this quality-related price change ΔQ, which value for the characteristic should be taken in a replacement situation, the one of the old dwelling to be replaced or the one of the replacement dwelling? The same question arises for modernizations: The value of the characteristic before or after the modernization, if it has changed?

In our view, the old value of the characteristic should be retained for the quality adjustment. Quality adjustment always uses a ceteris paribus assumption. That is: what would the price of the good or service be, if all the circumstances remained constant, but the quality features change? In our setting: what would the rental price of the dwelling be, if all non-quality characteristics remained constant, but the quality characteristics change?

The justification for keeping the old value of the characteristic is easy for the type of financing. A change in the type of financing just means a pure price change. Hence, it is obvious that first, based on the type of financing of the old dwelling, the price of the quality difference is estimated; the pure price change is then the difference between quality adjusted rental price of the old dwelling and the rental price of the new dwelling.

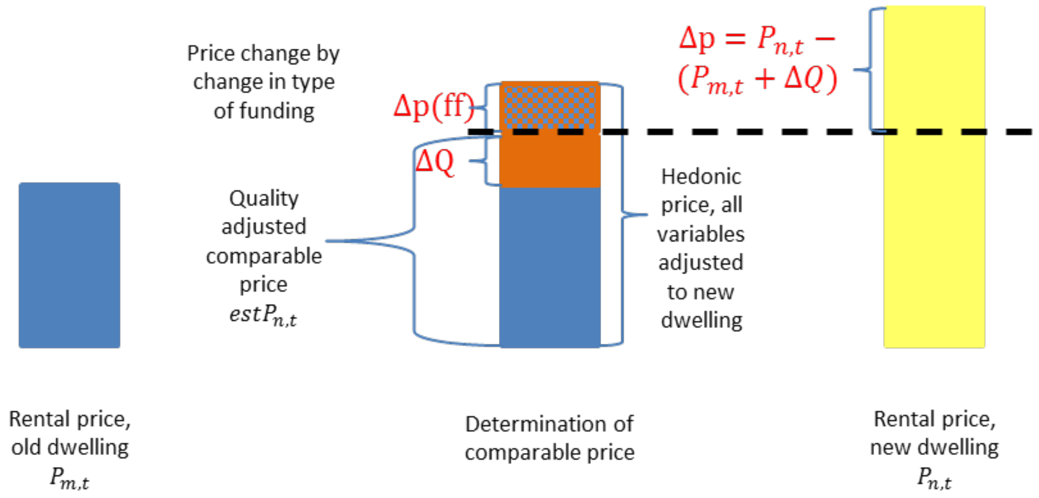

Chart B1: Change in type of funding as a price change.

The same reasoning can be applied for the duration of the rental contract; however, it can be interpreted in an interesting, different way: Not only the rental price, but also the value of the quality difference in rental terms also depends on the duration of the contract, as can be seen from the regression equation. The longer the contract lasts, the lower the rental price is compared to market development and the smaller will be the value of the quality difference. This means that the price change increases vice versa. It can be interpreted that the sitting discount provides a window into the past: It reveals what the price difference in terms of quality means to that particular contract closed in the past, not to a contract that is closed today. The quality-adjusted price difference can then be divided into two components: the change in price between the dwellings (in a replacement situation) or after a modernization was carried out, and the change in price of the quality difference occurring between the settlements of the rental contracts of the old and the new/modernized dwelling (see the following chart).

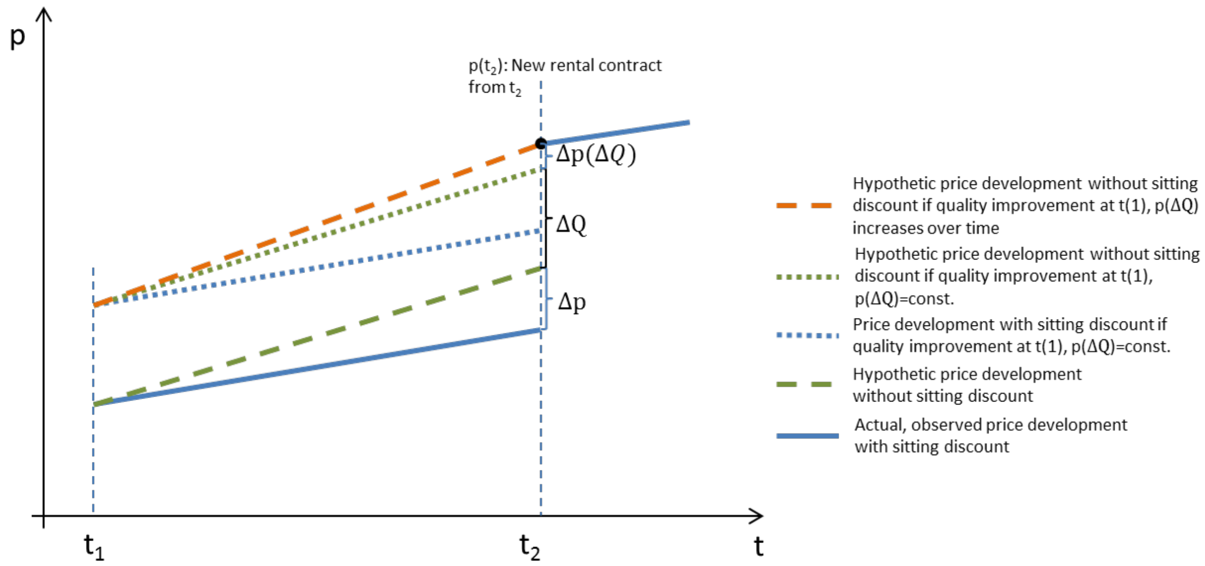

Chart B2: Sitting discount and price change.

B4.3 Meaning of dummy variables for the price estimation

So far, not much attention has been devoted to the case when a double logarithmic hedonic regression equation includes one or more dummy variables. With the many dummy variables as in our case, this is of special importance. Hence, it should be clear what their coefficients mean, in order to get a proper interpretation. The main question is what happens during the re-transformation. We start with a small double-logarithmic model, with one continuous and one dummy variable (and without intercept and error term; this does not limit the evidence in this case):

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

where x_i is the continuous variable and y_i is the dummy variable. Applying the exponential function e^x yields

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

Hence, if the circumstance that is indicated with the dummy variable y_i is present, this will lead to a multiplicative adjustment of the price estimated with the continuous variables by the factor e^{β_2}. This is justifiable for most dummy variables like quality of residential area or type of funding; however, regarding the presence of a garage, it is hard to explain why the additional fee should depend on the “pure rental price” of the dwelling as it does with the multiplicative factors stemming from the dummy variables. We do not find a solution for that, but we find it important to be aware of the limitations of the method.

B4.4 Quality adjustment - estimating the quality-related price difference

If quality adjustment is done with hedonics, the hedonic function estimates the price of a good at old and new quality (or, in replacement situations, the price of both goods according to their quality) not regarding the actual price. The difference is seen as the price of the quality difference. However, it has to be discussed how this quality difference can be calculated. We propose two approaches and discuss them. n indicates the new good, m the old one.

Approach 1: Difference of estimated prices

A first possibility is to simply use the difference between the estimated prices $\mathrm{estP_{n,t}}$ and $\mathrm{estP_{m,t}}$ ($\mathrm{x_{i,n}}$ and $\mathrm{x_{i,m}}$ are the continuous characteristics of the products, $\mathrm{d_{j,n}}$ and $\mathrm{d_{j,m}}$ the respective dummy variables, β_i and γ_j the regression coefficients, α_0 the correction factor):

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

Advantage: Increases and decreases of the quality by the same degree lead to the same value of the quality difference; the value is independent of the price level of the old good.

Disadvantage: The value depends also on variables not relevant to quality, but to the price, which leads to differences in the value estimation already if only these variables have different values. In order not to disturb the quality value estimation, they have to be constant for both estimations (see chapter B4.2).

Approach 2: Ratio of estimated prices[12]

A factor is calculated with which the price of the old good is multiplied in order to calculate a price for the quality difference for the actual period.[13]

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

Advantage: if a variable is included that is only relevant to the price, but not to quality, and this variable does not change its value or is kept constant (as explained in chapter B4.2), it has no influence on the value of the quality difference.

Disadvantage: The value of the quality difference depends on a real price. Hence, for an expensive good, it will be larger than for a cheaper product, even if both have the same quality according to their characteristics. This can lead to a downward bias of the index. Additionally, increases in quality have a larger impact than decreases in quality, due to percentage calculation. An example: estimated price of the new good is 120, of the old good 80: the factor is

---

[12] Triplett (2004) refers to this approach as „Hedonic quality adjustment method“. Regarding his formulation, some adaptation is needed to get from his equations to approach 2. His equation 3.13 (Triplett 2004, p. 79) estimates the price for the new good with the actual price of the old good and an adjustment factor:

$$estP_{n,t} = P_{m,t} * A(h)$$

This means for the quality difference between old and new good:

$$\Delta Q = estP_{n,t} - P_{m,t} = P_{m,t} * (A(h) - 1)$$

From his equation 3.12 regarding the quality adjustment factor A(h)), which seems to be the ratio of the estimated prices $\frac{estP_{n,t}}{estP_{m,t}}$, the formula can be developed further:

$$\Delta Q = P_{m,t} * (A(h) - 1) = P_{m,t} * (e^{\sum \beta_i \ln(\frac{x_{i,n}}{x_{i,m}})} - 1)$$

Triplett provides for $A(h)$ only an example that would lead to a formula like $A(h) = e^{\sum \beta_i \frac{x_{i,n}}{x_{i,m}}}$, not including any dummy variables. Obviously, in a double logarithmic model, $\ln x_{i,n}$ should be used instead of $x_{i,n}$. If, in the calculation of $A(h)$, $x_{i,n}$ was simply replaced by $\ln x_{i,n}$, this would be mathematically incorrect, as $\frac{lnx}{lny} \neq \ln(\frac{x}{y})$.The inclusion of dummy variables in the formula for ΔQ leads to approach 2.

[13] If the price $P_{m,t}$ is not available for period t, then it may be imputed with the price for the previous period $P_{m,t-1}$ and the common price development for the consumption segment.

0.5. Hence, if the actual price of the old good was 100, the quality adjusted price is calculated to 150, $\Delta Q = 50$. If the situation is exactly the other way round (estimated price of the new good 80, of the old go

od 120), the factor is -0.333, hence the quality adjusted price of the old good is 53.3, $\Delta Q = -26{,}7$. But quality increases and decreases by the same change in characteristics should have the same value.

Moving Ahead

For quality adjustment in the German rental sample, we follow approach 1: Only with this approach, the value of the quality difference is independent from the price of the old good and increases and decreases of quality of the same degree get attached the same value. The persisting problem of price dependence on variables relevant only for the price, but not for the quality, is tackled with the approach presented in chapter B4.2.

B5 Discussion and application in German price statistics

The hedonic model presented for the rental sample of German consumer price statistics is a quite straight-forward example of using hedonics in Triplett's sense of a "hedonic quality adjustment method" used in situations of replacement as well as quality change of a good. On the methodological side, its development clarified some minor, but important issues, like the use of an adjustment factor for the residuals α_0 for the re-transformation of the double-logarithmic price estimation, the treatment of variables determining only the price level, but not the quality of the good, the meaning of dummy variables and, especially important, the calculation of the value of the quality difference in a consistent manner regarding quality increases and decreases of the same degree. Only the difference of estimated prices gives a consistent result, not the ratio. Hence, price statisticians should use the differences.

The hedonic quality adjustment according to the method presented has been introduced in German consumer price statistics for modernizations of dwellings in 2016 (only for seven Bundesländer) and 2017 (for five additional Bundesländer). The remaining Bundesländer will follow in 2018 after a new calculation of the regression, as well as the use of the method in replacement situations. This is now in-line with the recommendations of Eurostat regarding quality adjustment in replacement and modernization situations. It will also account for the price change that occurs when newly constructed dwellings enter the market, which will be systematically included in the index from 2018 on. In conclusion, the introduction of the hedonic method as quality adjustment method for the German rental price statistics is an important step forward in maintaining and improving the quality and reliability of German consumer price statistics.

B References

Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR, 2015): Laufende Raumbeobachtung – Raumabgrenzungen: Siedlungsstrukturelle Kreistypen. http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Kreistypen4/kreistypen.html, accessed 08/05/2017.

Bundesministerium für Verkehr, Bau und Stadtentwicklung (BMVBS, 2012): Richtlinie zur Ermittlung des Sachwerts. BMVBS, Berlin. https://www.bundesanzeiger-verlag.de/fileadmin/BIV-Portal/Dokumente/PDF/Sachwertrichtlinie.pdf, accessed 08/05/2017.

Eurostat (2014): Harmonised Index on Consumer Prices – Recommendations: Treatment of actual rentals in the HICP. Version of August 2014, TFQI09.2014-02. Eurostat, Luxemburg.

Goldhammer, B. (2016): Die neue Mietenstichprobe in der Verbraucherpreisstatistik. In: Wirtschaft und Statistik, No. 5/2016, p. 86-101.

Hoffmann, J./Kurz, C. (2002): Rent Indices for Housing in West Germany 1985-1998. Working Paper No. 116, European Central Bank, Frankfurt am Main.

Wooldridge, J. (2009): Introductory Econometrics – a Modern Approach. 4th edition, South-Western, Mason, Ohio.

B Annex I

Variable	Name of region	Coefficient	Standard Error
$r_{i,201}$	Hamburg	0.279	0.012
$r_{i,301}$	Braunschweig	0.044	0.010
$r_{i,306}$	Hamburg-Umland-Süd	0.162	0.033
$r_{i,307}$	Hannover	0.107	0.010
$r_{i,309}$	Lüneburg	0.169	0.029
$r_{i,310}$	Oldenburg	0.077	0.024
$r_{i,504}$	Bochum/Hagen	-0.053	0.011
$r_{i,505}$	Bonn	0.128	0.019
$r_{i,506}$	Dortmund	-0.066	0.012
$r_{i,508}$	Düsseldorf	0.125	0.010
$r_{i,509}$	Emscher-Lippe	-0.053	0.013
$r_{i,510}$	Köln	0.194	0.011
$r_{i,511}$	Münster	0.053	0.019
$r_{i,601}$	Mittelhessen	0.119	0.015
$r_{i,602}$	Nordhessen	-0.090	0.014
$r_{i,603}$	Osthessen	0.062	0.024
$r_{i,604}$	Rhein-Main	0.280	0.009
$r_{i,605}$	Starkenburg	0.178	0.021
$r_{i,801}$	Bodensee-Oberschwaben	0.127	0.027
$r_{i,802}$	Donau-Iller (BW)	0.110	0.024
$r_{i,803}$	Heilbronn-Franken	0.058	0.025
$r_{i,804}$	Hochrhein-Bodensee	0.168	0.019
$r_{i,805}$	Mittlerer Oberrhein	0.134	0.016
$r_{i,806}$	Neckar-Alb	0.211	0.027
$r_{i,807}$	Nordschwarzwald	0.109	0.024
$r_{i,808}$	Ostwürttemberg	0.091	0.028
$r_{i,810}$	Stuttgart	0.297	0.010
$r_{i,811}$	Südlicher Oberrhein	0.169	0.017
$r_{i,812}$	Rhein-Neckar	0.160	0.014
$r_{i,901}$	Allgäu	0.112	0.018
$r_{i,902}$	Augsburg	0.034	0.016
$r_{i,905}$	Donau-Wald	-0.099	0.021
$r_{i,906}$	Industrieregion Mittelfranken	0.096	0.011
$r_{i,907}$	Ingolstadt	0.060	0.028
$r_{i,908}$	Landshut	0.067	0.026
$r_{i,909}$	Main-Rhön	-0.077	0.024
$r_{i,910}$	München	0.424	0.007
$r_{i,911}$	Oberfranken-Ost	-0.149	0.019
$r_{i,913}$	Oberland	0.253	0.023
$r_{i,914}$	Oberpfalz-Nord	-0.220	0.023
$r_{i,915}$	Regensburg	0.076	0.020
$r_{i,916}$	Südostoberbayern	0.114	0.017
$r_{i,917}$	Westmittelfranken	-0.072	0.027
$r_{i,918}$	Würzburg	0.055	0.021
$r_{i,1101}$	Berlin	0.104	0.008
$r_{i,1201}$	Havelland-Fläming	0.067	0.016

Variable	Name of region	Coefficient	Standard Error
$r_{i,1302}$	Mittleres Mecklenburg/Rostock	0.034	0.016
$r_{i,1403}$	Südsachsen	-0.103	0.009
$r_{i,1404}$	Westsachsen	-0.082	0.011
$r_{i,1502}$	Anhalt-Bitterfeld-Wittenberg	-0.031	0.014
$r_{i,1503}$	Halle/S.	-0.032	0.011
$r_{i,1504}$	Magdeburg	-0.024	0.010
$r_{i,1601}$	Mittelthüringen	-0.080	0.014
$r_{i,1602}$	Nordthüringen	-0.079	0.019
$r_{i,1603}$	Ostthüringen	-0.076	0.012
$r_{i,1604}$	Südthüringen	-0.056	0.022

Coefficients of all regional dummy variables that turned significant in order of size of coefficient: