Hedonic Methods of Price Measurement for Used Cars

A hedonic technique for calculating the consumer price index for second-hand cars was introduced in May 2003. The hedonic technique adjusts for quality by using regression analysis to measure the impact of product features on sale price. In this manner price changes that are due to quality shifts in specific features can be separated mathematically from “pure” price changes and purged. The following essay describes the methods underlying the hedonic price index for used cars.

For further details, please contact:
Verena Dexheimer, tel: +49(0) 611 75-2824, e-mail: verena.dexheimer@destatis.de

1. The principles and use of hedonic price measurement

1.1 Calculating a hedonic price index

The objective of official price statistics is to measure what we call “pure” price changes, purged of the adulterating influence of changes in consumption patterns, types of goods and quality features. This essentially reflects the Laspeyres Principle of once defining a basket of goods and keeping it as constant as possible over a defined period of time.1

The price of an item at two separate times can only be usefully compared if the quality of the item remains constant. If this is not the case – for instance due to technological progress – quality adjustment is undertaken in order to introduce the monetary value of an item’s quality change into price observation.2

Hedonic methods constitute a specific quality adjustment technique. The hedonic method uses regression analysis to measure the influence of product features on the sale price. Thus price changes due to qualitative improvements in certain features can be distinguished mathematically and purged from the pure price change which the price index is actually called upon to measure.3

---

1.2 Implementation of hedonic techniques of price measurement at the Federal Statistical Office

In 2002 the Federal Statistical Office in Germany began an extensive programme for introducing hedonic techniques of quality adjustment. Figure 1 provides an overview of the stages in this schedule.

**Figure 1: The Federal Statistical Office programme for implementing hedonic methods**

<table>
<thead>
<tr>
<th>Index position</th>
<th>Project status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer prices for home computers</td>
<td>Hedonic index introduced in June 2002</td>
</tr>
<tr>
<td>Consumer prices for new cars</td>
<td>Evaluation completed in January 2003. Hedonic methods are not used to calculate</td>
</tr>
<tr>
<td></td>
<td>this price index as no systematic deviations were observed between the techniques.</td>
</tr>
<tr>
<td>Consumer prices for used cars</td>
<td>Hedonic price index implemented in May 2003</td>
</tr>
<tr>
<td>Producer, import and export price indexes</td>
<td>Work in progress</td>
</tr>
<tr>
<td>for electronic data processing equipment</td>
<td></td>
</tr>
<tr>
<td>Consumer prices for electrical home</td>
<td>Work in progress</td>
</tr>
<tr>
<td>appliances and consumer electronics</td>
<td></td>
</tr>
<tr>
<td>Consumer prices for owner occupied housing</td>
<td>Work in progress</td>
</tr>
</tbody>
</table>

As a first step the hedonic method was introduced in June 2002 to the regular monitoring of prices for home computers.4

The second step for the German Federal Statistical Office was to evaluate the quality adjustment procedure hitherto applied to the consumer price index for motor vehicles. A hedonic price index was calculated parallel to the conventional price index for new cars. Analysis demonstrated that, for new cars sold in Germany, quality changes due to technological progress are adequately indicated by the conventional method of quality adjustment. No systematic deviations between the two indices were observed. The Federal Statistical Office has, therefore, not incorporated the hedonic method into its quality adjustment techniques for new cars, retaining instead the well-proven and significantly cheaper conventional approach known as “feature adjustment”.

The third step was to design a hedonic price index for used cars. In May 2003 a used car price index based on the hedonic approach was included in the consumer price index.

Work is currently progressing on hedonic producer, import and export price indexes for selected data processing equipment, on hedonic price indexes for the categories “electrical home appliances” and “consumer electronics” and on hedonic price indexes for owner occupied housing.

---

2. Hedonic consumer price index for used cars

2.1 Data sources

In order to calculate a hedonic price index we need price statistics to supply monthly information on the price (sale price), product quality (features) and rate of sale (frequency) of particular classes of product. Regression analysis is then used to establish a link between the item’s sale price and its quality features. Weighted regression is performed to take into account the frequency with which different versions of these items are sold.

The hedonic price index for used cars draws on data provided by the company DAT GmbH on sale prices and quality features. The weighting factors are founded on information supplied by the Kraftfahrtbundesamt, Germany’s national road vehicle registration authority (KBA).

DAT keeps monthly records of the prices for which second-hand cars change hands (transaction price), gleaned from used car dealers and OEM dealer networks. Additional information is derived from internet portals used by dealers as a virtual market-place. All in all 20,000–25,000 price notifications a month from all over Germany can thus be evaluated. For methodological reasons explained below in more detail, only sales of cars aged up to 10 years, accounting for 95% of market volume, are included in the used car index.

The KBA provides data on how often specific models of car change ownership. In line with the net principle, the consumer price index only considers sales from commercial undertakings to private households. The KBA data does not directly indicate whether the transaction was private or commercial. However, it can generally be assumed that a sale preceded by temporary vehicle deregistration is likely to be commercial. Consequently the number of transfers of ownership following deregistration has been selected as an approximate parameter for the required term: the frequency of sale of specific used models.

The information on sale frequencies derived from KBA data is classified according to what we call primary models, e.g. Opel Corsa, VW Golf, VW Polo. The sale prices and quality features provided by DAT, however, are collated more specifically according to sub-model, e.g. Opel Corsa City, Opel Corsa Swing (see Figure 2).

![Figure 2: Linking sale prices and quality features to sale frequency](image)

In order to ensure a direct correlation, a sale price median is calculated for each of the various sub-models and only the sub-model with that median price is used in downstream calculations.

---

5 Deutsche Automobil Treuhand GmbH, founded in 1931 by the automobile manufacturer and trade associations.
2.2 Price index calculation methods

When calculating hedonic price indexes, a choice must be made between two fundamental
techniques, the time dummy variable method and the imputation method. As the time dummy
variable method is simpler to perform for continuing price index calculations in this context, it
was selected for the used car index. In the time dummy variable method, the prices and quality
features of purchased used cars are summarised for two consecutive months and combined in
regression analysis. The procedure is demonstrated in Figure 3 for the months August and
September.

\[ P = f(x_1, x_2, \ldots, D_{\text{time}}, \varepsilon) \]

In the regression equation the price \( P \) is explained by the car's quality features \( x_1, x_2, \ldots \). Quality
features include age, mileage, deflated original price of the new car and brand (see chapter
2.3.1).

The dummy variable \( D_{\text{time}} \) (time variable) differentiates August from September. Dummy variables
have a value of one if a certain property holds true, and are otherwise set to zero. Thus for data
from August \( D_{\text{time}} = 0 \) applies and for data from September is identified as \( D_{\text{time}} = 1 \).

Finally, the random variable \( \varepsilon \) is used to indicate that not all impacts on a used car's price can be
measured. For example, the overall condition of the car cannot be rated in the data here. It is a
general rule that the quality of goods is only reflected very roughly in quality adjustment for price
statistics. In some cases, quality features may be a fairly inadequate basis for explaining the sale
price of a product. However, the number of products included each month in price measurement
is so large that deviations in the actual quality of an individual item will balance one another out
and the general trend in terms of quality shift will be adequately recorded.

The monthly price index which these operations are designed to observe is obtained from the
influence of the time variable on the price as calculated by regression analysis. The regression
equation serves to express the theoretical change in sale price in response to the time variable
assuming all other quality variables remain constant.

Using the procedure described above, a separate regression is calculated for
September/October, October/November and so on. The price index is calculated from this
sequence of month-on-month quality-adjusted price changes.
2.3  Regression analysis

2.3.1 Quality Features

The choice of quality features refers to the consumer behaviour. For used cars one can think for example of a two-step buying decision like this:

(1) First the consumer chooses his preferred car-type and defines an interval for age and mileage of his preferred car. The choice of the car-type is based on variables like the size of the car, its consumption, its engine power, its design, the technical equipment, the image of its brand and so on, always facing the price-level of this car-type.

(2) If the car-type is chosen, the consumer tries to find a special offer for sale. The chosen model is offered usually in different values for age and mileage. The customer tries to find the car with the best price-quality-level. Particularly for newer cars the consumer will make a discount according to age and mileage of the car, starting from the original price of the new car.

For our hedonic price index, such a two-step decision process is assumed, whereas only the second step is captured. Thus, it is assumed that the customers try to find a special car-type and that their set of alternatives consists only of different age and mileage categories of this chosen car-type. This corresponds to the assumption that there is no substitution among car-types which satisfies more a construction of a fixed-basket-of-goods-index instead of a cost-of-living-index. This approach is chosen according to the procedures used in the Federal Statistical Office for official price statistics.

Further it is assumed that customers are able to distinguish between price changes of new cars in consequence of quality changes and otherwise price changes because of inflation. Therefore the original price of the new car is adjusted by the official price index of new cars and this deflated original price of the new car is included in the regression equation.

The brand of the car is included as explanatory variable. This makes it possible to use the same regression equation for all car-types and at the same time the relationship between age and price can reach different values among different brands. Thus, different discount factors are assumed for different producers. But for different car-types from the same producer a uniform average discount factor is used.

2.3.2 Regression results

A weighted regression was performed with the average annual sale frequencies of the entire previous calendar year as weighting factors. This weighting based on calendar years was selected in order to avoid potential bias due to seasonal variations. A logarithmic function was selected to perform the regression equation:

\[ \ln(\text{SP}) = \alpha + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{kil} + \beta_3 \cdot \ln(\text{NP}) + \gamma_1 \cdot \text{D}_{\text{brand}} + \ldots + \gamma_{15} \cdot \text{D}_{15\text{brand}} + \delta \cdot \text{D}_{\text{time}} + \varepsilon \]

The variables are explained in the following table.
The semi-log link between sale price and age of car reflects the assumption that used cars lose value at a constant percentage for each month of age. The same applies to mileage, although in this case a relative mileage (kilometres travelled per month of age) was applied to bypass the correlation between mileage and age. For the deflated original price of the new car a double-log link to sale price proved more appropriate. Coefficient $\beta_3$ of the deflated original price can be regarded as a partial elasticity factor and denotes the average percentage increase in sale price of a used car for each one percent increase in the deflated original price (assuming all other variables remain constant). Dummy variables for manufacturers’ brands take account of variations due to brand awareness among consumers. For example, a used BMW has an average sale price that is higher by a certain factor than the price for other used cars. The price mark-ups for certain brands measured by the dummy variables are broadly independent of deflated original price.

Finally, the time variable $D_{time}$ is used to calculate the quality adjusted index, i.e. the price increase or decrease compared with the previous month when quality shifts are excluded. In a logarithmic function the quality adjusted percentage of price change over the previous month is calculated from the coefficients of the time dummy variables based on the following formula:

\[
(3) \text{ Quality adjusted price change} = [\exp(\delta) - 1] \cdot 100
\]

The estimations for the August/September 2003 regression are listed in the following table. The coefficient of determination is 0.96. Test statistics and residual plotting indicate that there may be heteroscedasticity. Additional heteroscedastically robust t-values were, therefore, included, but these confirmed the significance of the coefficient estimators.
The coefficients of all continuous variables and all brand dummy variables are significantly unequal to zero. The regression results are generally stable, similar values having been determined for earlier time intervals in the observation period from August 1997 to September 2003. However, the parameter estimators for the brand dummy variables were subject to fluctuations over time.

The parameter estimator for the age variable $\beta_1 = -0.01437$ denotes average monthly loss in value; projected over a year this is equivalent to a mean reduction in value of approx. 16% per year. The influence of mileage on sale price is far smaller at only about 0.1% per average annual mileage. Furthermore there is a close link between deflated original price and sale price: if a new car is one per cent more expensive – compared with other cars of the same age, mileage and brand – it will sell later at a used car price 0.916% higher.

As expected the link between deflated original price and sale price diminishes as the age of the car increases. The older a car, the more significant other factors become, such as the overall condition of the car, specific defects, etc. As these factors are hard to measure, cars older than ten years have been removed from price index calculations. The data demonstrates that

---

**Figure 5: Regression results August/September 2003**

\[ R^2 = 0.96 \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimator</th>
<th>Standard deviation</th>
<th>( t ) value</th>
<th>( P ) Value</th>
<th>Robust ( t ) values</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute term</td>
<td>( \alpha = 0.97948 )</td>
<td>0.0036</td>
<td>272.38</td>
<td>&lt;0.001</td>
<td>172.69</td>
<td>0.0000</td>
</tr>
<tr>
<td>age</td>
<td>( \beta_1 = -0.01437 )</td>
<td>0.000000279</td>
<td>-5154</td>
<td>&lt;0.001</td>
<td>-4133.28</td>
<td>1.22524</td>
</tr>
<tr>
<td>rel. mileage</td>
<td>( \beta_2 = -0.000117 )</td>
<td>0.0000006125</td>
<td>-190.72</td>
<td>&lt;0.001</td>
<td>-153.87</td>
<td>5.20785</td>
</tr>
<tr>
<td>In (original price defl)</td>
<td>( \beta_3 = 0.91569 )</td>
<td>0.00044224</td>
<td>2070.55</td>
<td>&lt;0.001</td>
<td>1346.32</td>
<td>5.16725</td>
</tr>
<tr>
<td>brand-dummies (reference value = Volkswagen)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUDI ( y_1 = 0.09288 )</td>
<td>0.0003721</td>
<td>249.61</td>
<td>&lt;0.001</td>
<td>238.99</td>
<td>1.23295</td>
<td></td>
</tr>
<tr>
<td>BMW ( y_2 = 0.07717 )</td>
<td>0.0003703</td>
<td>208.55</td>
<td>&lt;0.001</td>
<td>174.09</td>
<td>1.35077</td>
<td></td>
</tr>
<tr>
<td>CITROEN ( y_3 = -0.20343 )</td>
<td>0.00064163</td>
<td>-317.06</td>
<td>&lt;0.001</td>
<td>-213.63</td>
<td>1.05533</td>
<td></td>
</tr>
<tr>
<td>NISSAN ( y_4 = -0.03100 )</td>
<td>0.00056686</td>
<td>-54.69</td>
<td>&lt;0.001</td>
<td>-62.42</td>
<td>1.06282</td>
<td></td>
</tr>
<tr>
<td>FIAT ( y_5 = -0.10170 )</td>
<td>0.00044523</td>
<td>-228.41</td>
<td>&lt;0.001</td>
<td>-219.57</td>
<td>1.16885</td>
<td></td>
</tr>
<tr>
<td>FORD ( y_6 = -0.06579 )</td>
<td>0.00035724</td>
<td>-184.17</td>
<td>&lt;0.001</td>
<td>-189.20</td>
<td>1.16449</td>
<td></td>
</tr>
<tr>
<td>HONDA ( y_7 = -0.03123 )</td>
<td>0.00064286</td>
<td>-48.58</td>
<td>&lt;0.001</td>
<td>-68.26</td>
<td>1.04579</td>
<td></td>
</tr>
<tr>
<td>MAZDA ( y_8 = -0.02799 )</td>
<td>0.00068333</td>
<td>-40.96</td>
<td>&lt;0.001</td>
<td>-72.36</td>
<td>1.03888</td>
<td></td>
</tr>
<tr>
<td>BENZ ( y_9 = 0.09663 )</td>
<td>0.00031242</td>
<td>309.29</td>
<td>&lt;0.001</td>
<td>348.21</td>
<td>1.49511</td>
<td></td>
</tr>
<tr>
<td>MITSUBISHI ( y_{10} = 0.03830 )</td>
<td>0.00063516</td>
<td>60.29</td>
<td>&lt;0.001</td>
<td>90.91</td>
<td>1.04407</td>
<td></td>
</tr>
<tr>
<td>OPEL ( y_{11} = -0.09666 )</td>
<td>0.00027435</td>
<td>-352.31</td>
<td>&lt;0.001</td>
<td>-320.88</td>
<td>1.28098</td>
<td></td>
</tr>
<tr>
<td>PEUGEOT ( y_{12} = -0.04159 )</td>
<td>0.00056052</td>
<td>-74.19</td>
<td>&lt;0.001</td>
<td>-80.37</td>
<td>1.06382</td>
<td></td>
</tr>
<tr>
<td>RENAULT ( y_{13} = 0.00953 )</td>
<td>0.00032103</td>
<td>29.7</td>
<td>&lt;0.001</td>
<td>32.88</td>
<td>1.20953</td>
<td></td>
</tr>
<tr>
<td>SEAT ( y_{14} = -0.04029 )</td>
<td>0.00052788</td>
<td>-76.32</td>
<td>&lt;0.001</td>
<td>-91.92</td>
<td>1.07887</td>
<td></td>
</tr>
<tr>
<td>TOYOTA ( y_{15} = 0.00265 )</td>
<td>0.00060993</td>
<td>4.35</td>
<td>&lt;0.001</td>
<td>7.53</td>
<td>1.04862</td>
<td></td>
</tr>
<tr>
<td>Time dummy ( \delta = 0.00282 )</td>
<td>0.00016214</td>
<td>17.37</td>
<td>&lt;0.001</td>
<td>17.39</td>
<td>1.00031</td>
<td></td>
</tr>
</tbody>
</table>
excluding older cars hardly detracts from the representative nature of the price index, as it still covers 95% of all commercial turnover in used cars. Sales of older cars are relatively insignificant in commercial trading. In fact, limiting the used car price index to newer second-hand cars significantly increased its accuracy, as the sale price of these cars can be explained more fully by observable quality features.

2.4 Results

In September 2003 the used car index, adjusted for quality using the techniques described above, demonstrated a price change of +0.8% on the previous year. The price change on the previous month is +0.3%. Month-on-month price changes are shown in the Figure below.

![Figure 6: Used car price index month-on-month](http://www.destatis.de/mve/e/bv4.htm#BV4)

The index demonstrates strong variations in month-on-month changes. These can partially be ascribed to seasonal fluctuations, but may also reflect extraordinary impacts. The seasonal component can be purged from the price series by seasonal adjustment. The BV4 method was implemented for this purpose. This procedure is based on the assumption that the time series is an additive combination of seasonal and trend/business cycle components. The next Figure shows the shape of the average seasonal graph for the period from September 1997 to September 2003. The curve demonstrates how prices would change due solely to seasonal influences if all other influences were ruled out.

---

6 On this procedure see http://www.destatis.de/mve/e/bv4.htm#BV4.
When the used car price index is purged of seasonal influences, the month-on-month price curve becomes more even. Nonetheless, even after seasonal adjustment the price index indicates strong variations.

More pronounced falls in prices in the course of the year 2000 are no doubt primarily due to the introduction and extension of the environmental tax on fossil fuels (April 1999 and then at the start of each subsequent calendar year). Rising petrol prices as a result of this tax and higher prices for crude oil meant that older cars with higher fuel consumption lost value on the market. In early 2001, cars with lower pollution emissions also benefited from a tax break. Apart from these influences, the business cycle plays a major role in affecting price trends on the used car market.

Mid-term price developments emerge more clearly from the hedonic index. Figure 8 demonstrates a trend of falling prices until late 2000, which is reversed in the early months of 2001.
Figure 8:

Hedonic price index for used cars