Experimental use of hedonics for new cars in the Danish HICP

25 March 2011

Author: Martin Birger Larsen

STATISTICS DENMARK

Table of contents

1. Introduction

Statistics Denmark has run a project during 2010 with the purpose to produce experimental price index series for new cars quality adjusted using hedonic regressions. The results have been compared with the official Danish HICP for new cars and knowledge have been acquired on whether hedonic regressions could be used in practise in the monthly production of the Danish HICP for cars going forward. The project has been co-financed by Eurostat.

2. The current method for new cars in Statistics Denmark

Regarding quality adjustment standards for cars Statistics Denmark are currently using either bridged overlap or direct comparison on a case-by-case basis. The sample consists of around 46 monthly price observations and the price observations are collected using paper questionnaires send to a representative sample of car dealerships. Each car dealership is asked to state one of their most sold models and the corresponding sales price.

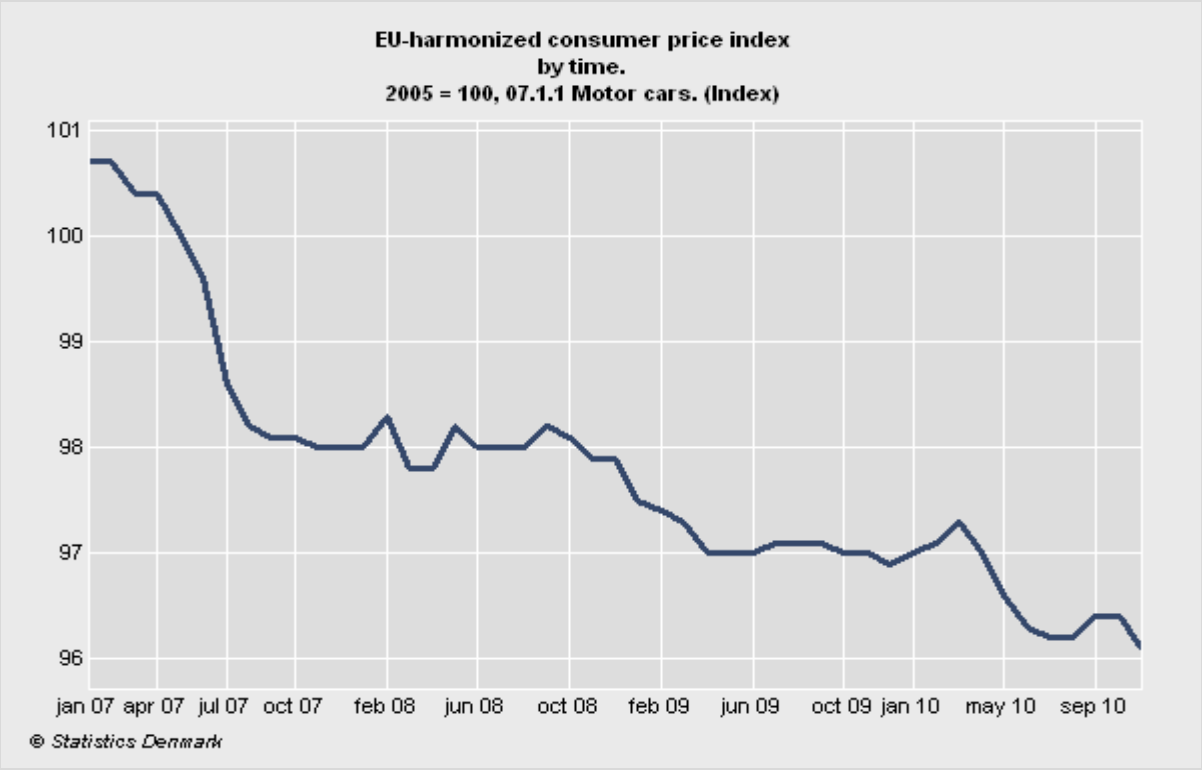
If in case of a product change the products are deemed to be of similar quality the method direct comparison is applied. If on the other hand the products are deemed to be of different quality the method bridged overlap is applied. The later is for instance the case when there is a change in engine volume or model (e.g. from a VW GOLF to a VW POLO).

Currently new cars are split across several elementary aggregates according to brands, see the table below:

COICOP	CAR
071101	VW & Audi
071102	Opel & BMW
071103	Ford
071104	Volvo
071105	Citroën, Peugeot & Renault
071106	Fiat
071107	Nissan, Toyota & Suzuki
071108	Mazda, Mitsubishi & Honda
071109	Skoda
071110	Hyundai

An unweighted geometric mean is applied across the price observations in each elementary aggregate and the Laspeyres formula is used to get a total price index for new cars. The resulting HICP for cars (07.1.1) is shown in the figure below.

Car prices decreased a lot in the middle of 2007 and have decreased steadily up to the end of 2010. In June 2007 there was a major change in the rules for the calculation of registration taxes. Registration taxes makes up between half and two-thirds of the consumer price of a car in Denmark so changes in the tax matters a lot. The major effect was that the registration taxes began to depend on how far cars could drive per litre of gasoline/diesel. The overall effect was that small cars became cheaper while big luxury cars became more expensive. Since Danes tend to buy small cars the overall effect was a price decrease.



3. Data from the Danish Ministry of Taxes and Statistics Denmark's Car Sales Register

To be able to compile indices for new cars with the use of hedonics two different car registers have been put in use, a car register from the Ministry of Taxation containing (list) prices for all cars sold in Denmark and a register from Statistics Denmark containing information on the number of cars actually sold.

Data from the Ministry of Taxation

When a car importer offers a new car for sale, they have to fill in a form specifying the characteristics of the car together with the price to the Ministry of Taxation. This is used to calculate the tax and fees that will be put on the car. This means that the Ministry of Taxation has a register of all the car models that are offered at any moment in time. The register is updated on a daily basis. If a car importer wishes to introduce either a new model or a new price on an existing model, they will fill in a new form and deliver it to the Ministry of Taxation who will update the register after calculating the tax and fee to be applied to the new entry.

In practise datasets from 15th of each month from January 2007 until May 2010 have been used in the calculations in this paper.

The dataset from the Ministry of Taxation contains the following variables:

Brand of the vehicle
Model of the vehicle
Variant of the vehicle
Type of the vehicle (car, van or motorcycle)
Weight of the vehicle
Model year
Engine volume
Engine performance (horse power)
Fuel type (gasoline / diesel)
Transmission type
Number of airbags
ABS – Anti-lock breaking system
ESP – Electronic stability control.
Fuel Economy (km/litre)
Price of the car
Taxes and fees on the vehicle.

Only prices for cars and not vans or motorcycles have been used, so the latter two categories have been filtered away using the type of vehicle variable.

Data from Statistics Denmark's Transport Division

Statistics Denmark publishes the number of new registrations of motor vehicles by type of vehicle as a monthly statistic and so a dataset from Statistics Denmark's Transport Division was available, which contains data for the period of January 2006 to December 2009 with all the newly registered vehicles. For every newly registered vehicle the following variables are available:

Chassis Number
Brand of the vehicle
Model of the vehicle

Variant of the vehicle
Weight of the vehicle
Engine volume
Fuel type
Owner Type (Private or Business)
Registration date

This dataset was used in order to create weights for the prices found in the dataset from the Ministry of Taxation. All vehicles bought by business were dropped from the dataset and the data was summed up on a yearly basis in order to create yearly weights.

Merging the two datasets

The data from Statistics Denmark's Transport Division and the data from the Ministry of Taxation were not directly comparable.

Example on a record from the Transport Division's dataset:

Brand: VOLKSWAGEN
Model: GOLF
Variant: 1.6 TSI

Example on a record from the Ministry of Taxation:

Brand: VOLKSWAGEN
Model: 5K1-43M GOLF /W4B
Variant: /VWP 1K

Since manually merging data would be very time consuming, it was not a valid option. Instead a search for the simple model name from Statistics Denmark's register (e.g. GOLF) was carried out in the dataset from the Ministry of Taxation to make a key between the two datasets. A lot of manual work still had to be put into the process to correct minor errors. Examples of these were problems with the dataset sometimes naming the Audi A4 as "A4" and sometimes "A 4".

Engine volume was also used in merging the two datasets. The two datasets did not necessarily show the same values, but these where close (an example would be the engine volume showing up with 1397 in one dataset and 1396 in the other) and by rounding to the nearest 100, this was easily fixed. All in all the two datasets were merged by brand name, model name and engine volume.

In order to produce weights on the dataset from the Ministry of Taxation, the number of registrations stratified by brand, model and engine volume was summed up and then evenly divided on all price records containing this brand, model and engine volume. Yearly weights based on number of

registrations in year t-1 have been used for the regressions in year t (e.g. data for the number of registrations in 2006 have been used as weights for the 2007 prices).

Finally the dataset were checked for outliers and errors and a few outliers were removed from the dataset due to this process.

4. Hedonics used

The work in this paper has relied heavily on the recommendations on hedonics giving by the CENEX in their handbook regarding quality adjustments (*Destatis, 2009*). The work has focused on two different methods, the time-dummy method that directly gives a price index and the hedonic re-pricing method that is most similar to the current practise of compiling the price index for new cars in the Danish HICP.

The hedonic time dummy variable method

The hedonic time dummy variable method computes the price index as a transformed regression coefficient from analysis on pooled data from more than one time-period.

As an example we can look at a model with three characteristics variables, $z1$, $z2$, $z3$. In the case of two considered periods, the reference period $t = 0$ and the comparison period $t = 1$, the hedonic regression equation may have this form,

$$\ln P = b_0 + b_1 * z_1 + b_2 * z_2 + b_3 * z_3 + y * D_{t=1} + e$$

where Dt-1 is a time dummy variable taking the value 1 for the period $t = 1$ and 0 otherwise. e is the error term.

The price index, for period 1 with reference period 0, now appears as the exponentiated value of a regression coefficient in the regression equation.

The computation procedure is that a regression analysis is run on data from both periods $t = 0$ and $t = 1$ simultaneously, and then the index $I(t)$ is obtained from one of the estimated regression coefficients. The same method could be applied directly in a setting with more than two periods, by taking several time dummy variables, one for each period after the reference period.
In practise yearly regressions have been used with pooled data from December year t-1 to December year t and have added twelve dummy variables. This method has been used for all data within a year

pooled together and for clusters stratified according to brand as is currently done in practise in the official Danish HICP for cars. Furthermore monthly regressions have been carried out with pooled data from month t and month t+1. The results are presented in section 5.

The hedonic re-pricing method

The monthly current index calculation is usually straightforward when the method of hedonic re-pricing is used. The hedonic function *h* (.) has been computed earlier, so it is available for use in the index computation. Then the monthly process for index calculation is as follows:

1. For each product-offer with a model replacement since the preceding month, compute a quality adjustment factor as:

g = h(Characteristics of replaced model) / h(Characteristics of replacement model)

where *h*(.) stands for a hedonic function, which should ideally be estimated fairly recently, but the timing of its estimation is less critical with this method than with other methods.

2. For each of the mentioned product-offers, modify the reference price by multiplying it with the quality adjustment factor *g*, computed for that product-offer.

3. Calculate the index by the usual index formula.

In practise hedonic re-pricing has been used on the historical dataset for new cars in the HICP and for all product replacements hedonic re-pricing has been applied instead of the case by case use of direct comparison or bridged overlap. A yearly hedonic model has been specified exactly like the time-dummy model but of course without dummies added. The results are shown in section 5 below.

5. Hedonic indices for new cars – results

Results from the hedonic time-dummy variable method on a pooled yearly dataset

Data from December in year t-1 have been pooled together with data from all 12 months in year t. In this pooled dataset the logarithm to the price observations from the dataset from the Ministry of Tax have been regressed on the variables (without logarithm applied) shown in the table below. The choice of variables in the regression was based on the wish of getting a parsimonious model with high explanatory power and with plausible parameter estimates for all variables. Time-dummies was added

to the last 12 months in the pooled regression. The regressions have been compiled both using weights and without.

All the estimated coefficients have the expected signs. Gasoline is the only negative since that gasoline cars tend to be cheaper than cars running on diesel. Furthermore all the variables are highly significant.

Resulting coefficients using yearly time-dummy regressions:

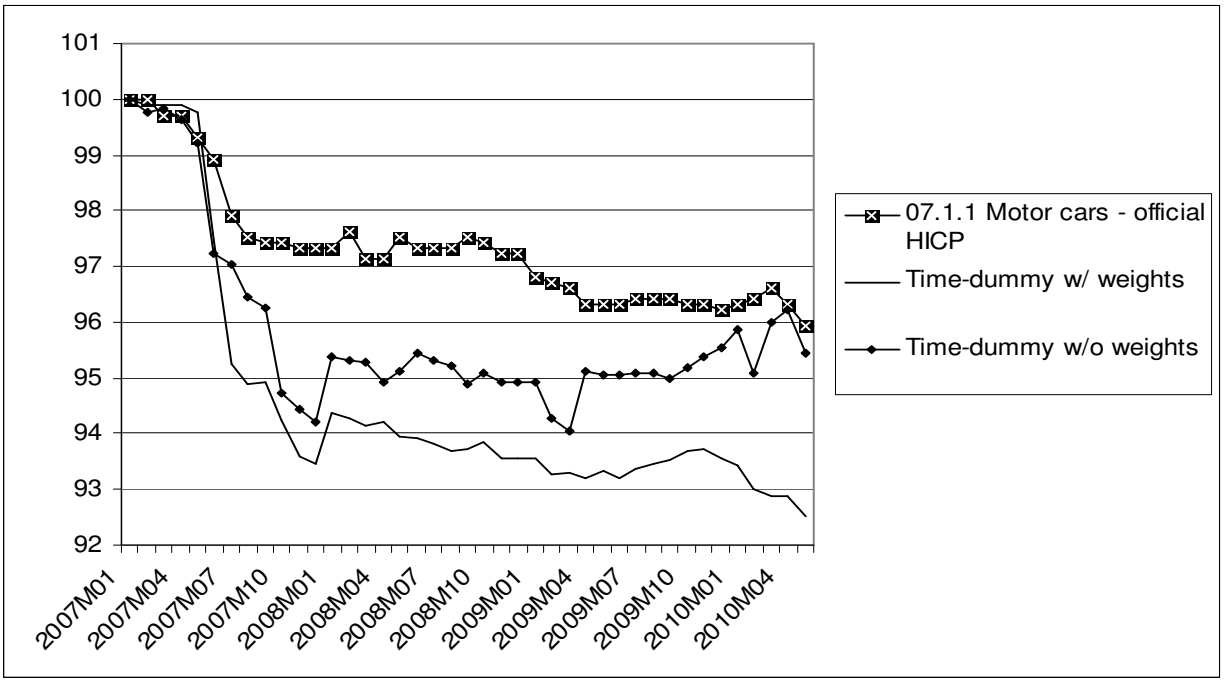
Variable	Parameter estimate 2007	Parameter estimate 2008	Parameter estimate 2009	Parameter estimate 2010
Intercept	1.081.164	1.038.816	1.013.217	1.023.047
Total weight of car	0.00029701	0.00039358	0.00060870	0.00061061
Engine volume	0.00027067	0.00035105	0.00023273	0.00012404
Horsepower	0.00509	0.00421	0.00450	0.00562
Gasoline	-0.14049	-0.07129	-0.03980	-0.10125
No. of Airbags	0.00924	0.03214	0.04284	0.04179
ESP	0.15832	0.10048	0.04987	0.03502

When it comes to explanatory power the above regressions are also performing quite well. The R-square as well as the adjusted R-square are shown below. They equal each other since the sample size is very big. In both cases the R-square is above 0,86 which must be said to be a satisfactory result.

Full time-dummy w/ weights	2007	2008	2009	2010
R-Square	0.8659	0.8726	0.8917	0.8935
Adj R-Sq	0.8659	0.8726	0.8917	0.8935

In the below figure the resulting price indices for the time-dummy regressions with and without weights are compared to the official HICP index for new cars. All three diverge and the major difference is around June 2007 at the same time as the changes in the registration tax rules took place. The time-dummy regressions are both falling much more than the official index suggesting that the official Danish HICP might have understated the price falls in connection with the tax changes.

Official HICP for cars and time-dummy indices with and without weights:



Looking only at 2008 and ahead the developments in the official index and the time-dummy regression with weights are to a high extent similar while the time-dummy regression without weights is increasing instead of decreasing. Since the R-square is bigger when using weights in the time-dummy regressions the conclusion must be that weights should be included. This is also suggested by the CENEX handbook for time-dummy regressions. Hence weights have been used in the regressions in the rest of the time-dummy regressions shown in this paper.

Results from the time-dummy variable method on a stratified dataset

To compare the results of the time-dummy regressions more closely with the current practise weighted time-dummy regressions have been calculated stratified according to car brands as in the Danish HICP for new cars:

COICOP	CAR
071101	VW & Audi
071102	Opel & BMW
071103	Ford
071104	Volvo
071105	Citroën, Peugeot & Renault
071106	Fiat
071107	Nissan, Toyota & Suzuki
071108	Mazda, Mitsubishi & Honda
071109	Skoda
071110	Hyundai

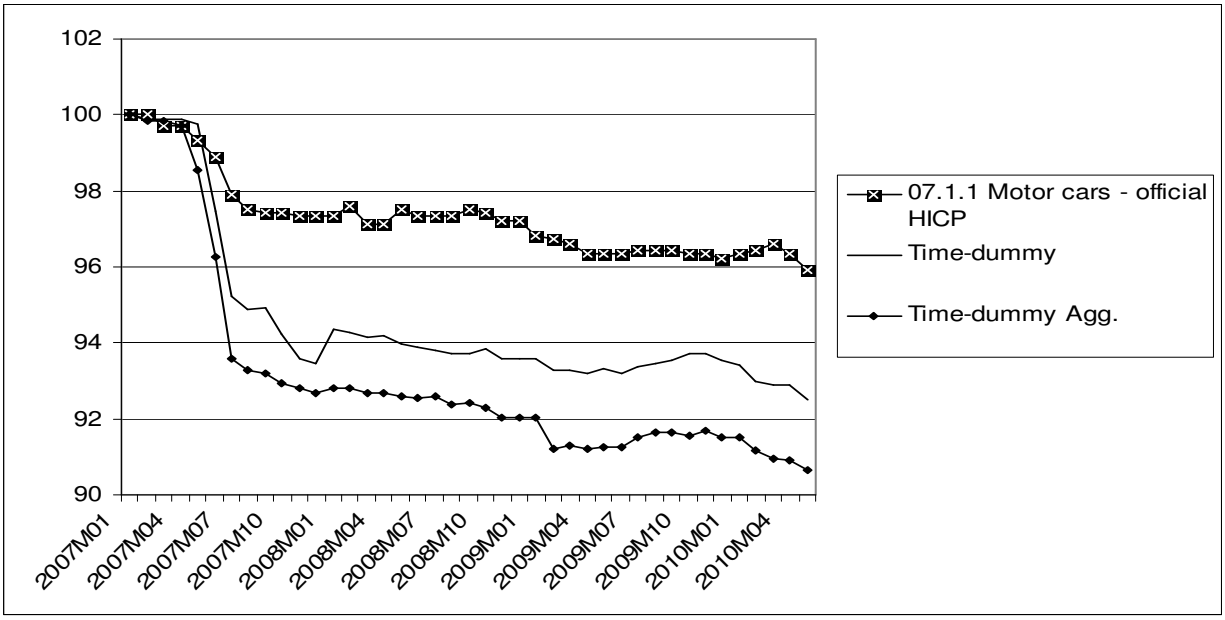
Otherwise the stratified regressions are similar to the regressions shown above (yearly pooled data and same variables). In almost all of the stratified regressions the signs on the coefficients are as expected and they are also almost always significant. The R-square is between 0,85 and 0,95 for the different stratified regressions.

Finally the detailed stratified time-dummy indices have been aggregated using the same weights that have been applied at the official HICP (see table below). This makes it possible to compare with the time-dummy model that was not stratified.

COICOP	CAR	2006 - weights	2009 - weights
071101	VW & Audi	0,7241	1,1861
071102	Opel & BMW	0,3919	0,6064
071103	Ford	0,3157	0,4973
071104	Volvo	0,1033	0,1831
071105	Citroën, Peugeot & Renault	0,7666	1,3647
071106	Fiat	0,1947	0,2893
071107	Nissan, Toyota & Suzuki	0,6317	0,9701
071108	Mazda, Mitsubishi & Honda	0,3097	0,4343
071109	Skoda	0,1454	0,2895
071110	Hyundai	0,2188	0,3663
	Total	3,8019	6,1872

The aggregated (stratified) time-dummy index is compared with the time-dummy index without stratification and the official HICP price index for cars in the figure below. From 2008 and onwards all three are showing similar trends. In 2007 the aggregated stratified time-dummy index is decreasing the most whereas the pooled time-dummy index is decreasing slightly less. The official index is decreasing even less in 2007. Comparing the two time-dummy indices it is seen that the use of stratification as in the official index contributes to a downward movement in the time-dummy index at the time of the changes in the registration taxes. In other words the weights applied in the official index seems to put a greater weight on car brands where prices are falling compared with the weights applied at the pooled time-dummy regression.

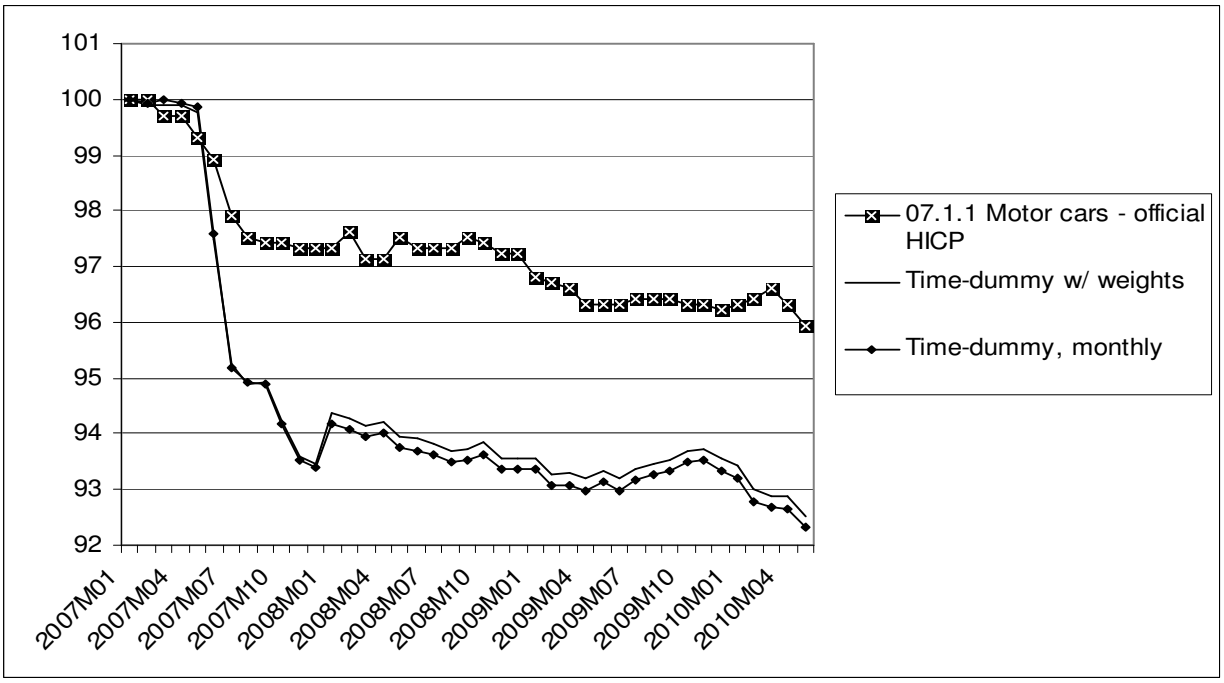
Official HICP index for cars and time-dummy indices with and without stratification:



Results from the monthly time-dummy variable method

If a hedonic time-dummy index should be used in practise in the monthly production of the HICP a yearly pooled time-dummy index could not be used. Hence a chained monthly time-dummy index has been calculated and compared with the yearly index in the figure below. The two time-dummy indices are showing almost exactly the same results. It can therefore be concluded that it only matters a little whether data are pooled across a full year or only on a monthly basis.

Official HICP index for cars and monthly and yearly pooled time-dummy indices:



Results from the hedonic re-pricing method

To be able to calculate hedonic re-pricing indices, regressions have been carried out on yearly pooled data. The regressions are exactly similar to the time-dummy regressions carried out above. The only difference is that time dummies naturally have not been added in the regression equation. Weights have been applied in the hedonic re-pricing indices even though the CENEX recommends otherwise due to an implicit reduction of the sample size. The use of weights increases the R-square and gives all coefficients the expected signs which are not the case without the use of weights in the regressions. The table below shows the coefficients for regressions on pooled data. The coefficients are almost exactly identical to the coefficients from the time-dummy regression which is reassuring. The adding of time dummies does not significantly alter the model.

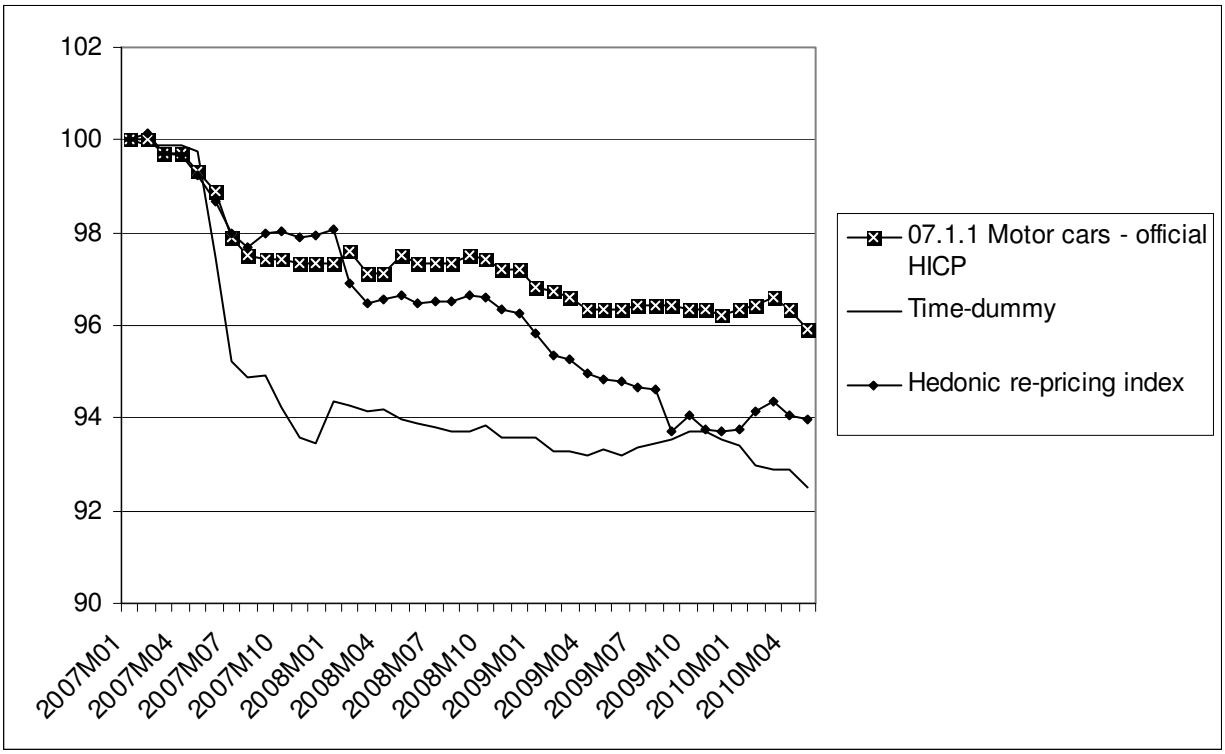
Parameter coefficients for hedonic re-pricing:

Variable	Parameter estimate 2007	Parameter estimate 2008	Parameter estimate 2009	Parameter estimate 2010
Intercept	1.079.641	1.039.289	1.013.093	1.022.533
Total weight of car	0.00029463	0.00039353	0.00060869	0.00061055
Engine volume	0.00027373	0.00035116	0.00023269	0.00012401
Horsepower	0.00509	0.00421	0.00450	0.00562
Gasoline	-0.14233	-0.07132	-0.03980	-0.10128
No. of airbags	0.00659	0.03214	0.04282	0.04177
ESP	0.15727	0.10006	0.04994	0.03480

Similar hedonic re-pricing regressions has been stratified and run according to brands. The coefficients from the stratified regressions had been used in practise when calculating the hedonic re-pricing index.

The sample of models and price observations from the official HICP index has been taken as a starting point. Whenever a product change was present the (most) corresponding new and old models in the register from the Danish Ministry of taxes have been found. The reason is that the information collected for the official index has not been detailed enough to carry out hedonic re-pricing. For instance information on total weight of the car, horse power or number of airbags is not known. The information on the variables in the register has been transferred to the models in the official HICP sample and the calculated coefficients have been applied to the price of the model leaving the sample together with knowledge regarding changes in the variable values. A hedonic re-pricing index has been calculated for each cluster and finally they have been aggregated using the same weights as in the official index. The figure below compares the resulting aggregated hedonic re-pricing index with the official HICP index and the pooled yearly time-dummy index.

The official HICP for cars, the pooled time-dummy index and the hedonic re-pricing index:



The hedonic re-pricing index falls much more than the official HICP for cars. Hence the use of hedonics for model changes in the sample is causing the index to fall more over time. The hedonic re-pricing index is not showing the same drastic fall as the time-dummy index in connection with the tax changes in the summer 2007. But the index level of the time-dummy index is reached by the hedonic re-pricing index in the end of 2009. The hedonic re-pricing index is more or less stable from here onwards.

Analysis of the effect changes in registration taxes in 2007 on the hedonic models

To see the effect of the changes in the car registration taxes in 2007 monthly hedonic models throughout 2007 have been run. The resulting coefficients are shown in the table below. Most of the coefficients are relatively stable. Gasoline however becomes considerably less negative in July 2007. This is a plausible effect since diesel cars usually run longer per litre of fuel than gasoline cars. Hence diesel cars became relatively cheaper compared with gasoline cars as a result of the tax changes. The coefficients for gasoline are for some reason however back to "normal" at the end of 2007. The coefficient on no. of airbags increases considerably in July 2010. This is probably due to the fact that the first and second airbag in a car no longer gave reductions to the registration tax after the tax changes.

The table suggests that the hedonic model should be updated whenever major changes take effect as for instance a major overhaul of the registration tax system. Looking at the relatively stable monthly coefficients below in the first and last months of 2007 this also suggest that it in normal circumstances it will be an acceptable choice to use regressions that are updated once a year only.

	jan-07	feb-07	mar-07	apr-07	maj-07	jun-07	jul-07	aug-07	sep-07	okt-07	nov-07	dec-07
Intercept	1.070.915	1.067.351	1.067.564	1.067.986	1.086.733	1.084.785	1.027.900	1.025.153	1.027.990	1.073.421	1.070.178	1.073.585
Total weight of car	0,0005	0,0006	0,0006	0,0006	0,0003	0,0003	0,0006	0,0007	0,0007	0,0002	0,0002	0,0001
Engine size	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0001	0,0001	0,0001	0,0004	0,0004	0,0004
Horsepower	0,0049	0,0047	0,0048	0,0049	0,0051	0,0051	0,0048	0,0047	0,0049	0,0050	0,0048	0,0049
Gasolin	-0,1275	-0,1186	-0,1211	-0,1225	-0,1554	-0,1533	-0,0613	-0,0689	-0,0782	-0,1386	-0,1316	-0,1419
No. Airbags	0,0031	0,0025	0,0020	0,0020	0,0072	0,0072	0,0295	0,0317	0,0307	0,0250	0,0234	0,0279
ESP	0,1450	0,1474	0,1459	0,1509	0,1561	0,1420	0,1251	0,1150	0,1157	0,1631	0,1429	0,1344

6. Possibilities of regular production using hedonics

In conclusion the possibilities of using the register from the Danish Ministry of Taxes in combination with hedonics seem very promising. The calculated hedonic models have high explanatory powers and the parameter estimates seems plausible and are significant. The calculated hedonic time-dummy indices show similar trends as the official HICP for cars when looking apart from 2007, where a major change in the rules for the calculation of registration taxes took place. Taking into account the relatively small sample for cars in the official HICP, the hedonic time-dummy indices probably gives a better picture of the price falls that actually took place during the summer of 2007. The hedonic re-pricing index is showing a different trend in the observed period but this is likely due to the fact that there is some time lag in "catching up" with the new price levels shown by the time-dummy indices.

Regarding the timing of the data this will probably not pose problems in a monthly production since the datasets from the Ministry of Taxation are automatically sent and received in the middle of the month containing current prices at that particular point in time. Regarding weights for the regressions these should only be compiled once a year. It is not certain whether it will be possible to get access to the necessary data for year t-1 before the end of January in year t but in that case the new weights could probably be implemented in the regressions with a time lag of one month.

A drawback with the use of the register from the Danish Ministry of taxes is that it contains list prices and not actual sales prices. The car dealerships in the official HICP for cars are asked to state actual transaction prices but in practise most of them give list prices. Most of the prices stated by the car dealerships could be found with the exact amount in the register.

Given the promising results Statistics Denmark plans on running a monthly production of both hedonic time-dummy indices and a hedonic re-pricing index during 2011 along with the current practise. This should make it possible to conclude whether it will be recommendable to change the data sources and methods for new cars in practise in the running monthly production. If everything goes without trouble then the new data sources and some hedonic index could be implemented for new cars in the HICP starting with the index for January 2012. In this process Statistics Denmark will also have to decide between using a time-dummy index or a hedonic re-pricing index. The time-dummy index is based on a very big sample whereas the hedonic re-pricing index is much more transparent. Even if it turns out that the use of hedonics is not possible in practise, Statistics Denmark will most likely turn to the register from the Danish Ministry of Taxes as a data source. This would enable Statistics Denmark to cost-effectively increase the current sample size.

An advantage with waiting to implement the new data sources and possibly hedonics to 2012 is that a possible new COICOP/HICP classification will probably have a different stratification than according to brands as applied currently. Waiting will give us time to test the hedonics on different stratifications. In any case it will be easier to shift to a new stratification using data from the Danish Ministry of Taxes than by changing the questionnaires in the current price collection. Finally the Ministry of Taxes is expected to merge its different car registers with a deadline in October 2011. This new harmonized register might contain a (list) price for each individual registered sale. This will make it unnecessary to use the method to apply weights to the register as has been done for the experimental hedonic indices presented in this paper.

7. Conclusion

Experimental hedonic time-dummy indices as well as a hedonic re-pricing index for new cars have been calculated based on an official register containing car (list) prices. The results looks promising as the explanatory power is high (R-square between 0,85 and 0,95) and the parameter estimates are looking plausible and have the expected signs.

The calculated time-dummy indices and the hedonic re-pricing index differ somewhat from the official Danish HICP for new cars. The time-dummy indices decrease much more in connection with a major change in the registration taxes in June 2007 than the official index. From 2008 to 2010 the time-dummy indices and the official index are however showing similar trends. The hedonic re-pricing index is showing a different trend and is falling more than the official index until it is ”catching up” with the time-dummy index in the end of 2009 after which it is more or less stable.

Given the promising results and the timely reception of data from the register it will most likely be possible to implement the new data source as well as hedonics in the monthly production of the HICP for new cars. To be 100 percent sure Statistics Denmark will calculate a time-dummy index and a hedonic re-pricing index on a monthly basis along with our current method throughout 2011. The new methods could then be implemented in January 2012. In any case Statistics Denmark will most likely change data source to the official register.

8. References

Destatis: *CENEX: Handbook on the application of quality adjustment methods in the Harmonized Index of Consumer Prices, 2009.*