BANCO DE PORTUGAL

Economic Research Department

Hedonic Prices Indexes for New Passenger Cars in Portugal (1997-2001)

Hugo J. Reis
J.M.C. Santos Silva

WP 10-02 September 2002

Please address correspondence to João Santos Silva, ISEG, Rua do Quelhas 6, 1200-781 Lisboa, Fax: #351-213922781, e-mail:jmcss@iseg.utl.pt or Hugo J. Reis, Economic Research Department, Banco de Portugal, Av. Almirante Reis nº 71, 1150-012 Lisboa, Portugal, Tel.#351-213130392; Fax#351-213107804; e-mail: hfreis@bportugal.pt; available in *www.bportugal.pt*.

Hedonic Prices Indexes for New Passenger Cars in Portugal (1997-2001)*

Hugo J. Reis
DEE/Banco de Portugal

J.M.C. Santos Silva
ISEG/Universidade Técnica de Lisboa

Abstract

In this paper we study the effects of quality change on the price index for new passenger cars in Portugal for the years 1997-2001. Hedonic regression models are studied, giving particular emphasis to the relation between the form of the price index and the specification of the hedonic equation and estimation method used. The results indicate that the CPI component corresponding to this item may have been overestimated by as much as 2.2 percentage points per year. This corresponds to an overestimation of the overall CPI by about 0.15 percentage points per year. As a by-product of this analysis it is also possible to conclude that the quality of new cars sold in Portugal increased on average 4.8 percent per year during this period.

JEL classification code: C21; C43; E31; L62.

Key words: CPI bias; Heteroskedasticity; Oaxaca decomposition.

*We are indebted to Carlos Coimbra, Pedro Duarte Neves, Rui Evangelista, José Machado, Paulo Parente, Maximiano Pinheiro, Pedro Portugal and Daniel Santos for helpful comments and suggestions. The computational assistance of João Barrambana was invaluable and Adelino Correia Martins from Marketing Systems GmbH provided many helpful insights on the data used. The usual disclaimer applies. Santos Silva gratefully acknowledges the partial financial support from Fundação para a Ciência e a Tecnologia, program POCTI, partially funded by FEDER, and the hospitality, working conditions and financial support provided by Banco de Portugal, which made this work possible.

Address for correspondence: João Santos Silva, ISEG, R. do Quelhas 6, 1200-781 Lisboa, Portugal. Fax: +351 213922781. E-mail: jmcss@iseg.utl.pt.

1. INTRODUCTION

Although the study of the effect of quality changes on prices of products is 75 years old, this topic is perhaps more interesting now than ever before. In fact, the Boskin Report (Boskin, Dulberger, Gordon, Griliches and Jorgenson, 1996) suggested that mismeasuring the effects of quality change in a period of fast technical progress can lead the CPI to substantially overestimate inflation. This conclusion brought the study of the effects of quality changes on price indexes to the forefront of the political debate and consequently to the agenda of many researchers.

As a consequence of this renewed interest, the literature on the relation between quality change and prices has grown very rapidly. Recent studies in this area have addressed, for example, the effect of quality change on the prices of computers (e.g., Moch, 2001, and Pakes, 2002), housing (Bover and Velilla, 2002, and Hoffmann and Kurz, 2002), and passenger cars (e.g., Bode and Van Dalen, 2001, and Izqierdo, Licandro and Maydeu, 2001). An up-to-date summary of the main contributions to this literature can be found in Triplett (2000).

Building upon the work of Santos and Coimbra (1995), this paper studies the effects of quality change on the price index for new passenger cars in Portugal for the years 1997-2001. The specification and estimation of hedonic regression models are studied and particular attention is given to the relation between the form of the price index and the specification of the hedonic equation and estimation method used. The results indicate that the choice of method used to construct the quality adjusted price index is an important issue since the various methods lead to somewhat different results. The present study departs from the work of Santos and Coimbra (1995) in a number of aspects. Besides the distinct time period considered, we have at our disposal a richer database and benefit from all the developments in the literature on this subject over the more recent years. Of course, there are also some differences which result purely from the different preferences of the authors.

The remainder of the paper is organized as follows. Section 2 discusses some issues on the specification of hedonic regressions and the compatibility between the functional form of the hedonic regression model and the form of the price index that is considered. Section 3 describes the data used in the empirical analysis. Section 4 presents details on the specification and estimation of the hedonic regression models and discusses the main results. Finally, section 5 concludes.

2. ECONOMETRIC SPECIFICATIONS AND PRICE INDEXES

The econometrics of hedonic regression models has been the subject of many studies. A standard reference in this area is Berndt (1991, Chapter 4), and a state-of-the-art review of the subject can be found in the excellent survey by Triplett (2000). This section addresses some econometric aspects of the construction of hedonic price indexes, focusing on the relation between the specification of the hedonic equation and the form of the price index.

The relation between the form of the price index to be corrected and the form of the hedonic regression has been pointed out before (e.g., see Triplett, 2000, Chapter III), but there seems to be some confusion in the literature about this issue. For example, following Triplett (1989), Brown (2000) corrects an index based on arithmetic means using imputed prices constructed from a hedonic regression in which the dependent variable is the logarithm of the price. However, although this is often ignored, strictly speaking it is not adequate to construct indexes based on arithmetic means if the dependent variable in the hedonic equation is the logarithm of the price.

To see why this is so, consider the case in which the researcher estimates

$$\ln(P_{it}) = x_{it}\beta_t + u_{it}, \qquad t = 1, \ldots, T, \qquad i = 1, \ldots, n_t \tag{1}$$

where T denotes the number of time periods used, n_t is the number of observations in period t, P_{it} is the price, x_{it} is a set of product characteristics, β_t is a coefficient measuring the value of the characteristics, and u_{it} is a random variable defined by $u_{it} = \ln(P_{it}) - E[\ln(P_{it})|x_{it}]$. This model is designed to estimate $E[\ln(P_{it})|x_{it}] = x_{it}\beta_t$, but in general it is not possible to recover $E[P_{it}|x_{it}]$ from $x_{it}\beta_t$. In fact, from (1) it is possible to write

$$P_{it} = \exp(x_{it}\beta_t)\exp(u_{it}),$$

which leads to

$$E[P_{it}|x_{it}] = \exp(x_{it}\beta_t)\,E[\exp(u_{it})|x_{it}].$$

Unless u_{it} is statistically independent of x_{it}, $E[\exp(u_{it})|x_{it}]$ is an unknown function of x_{it} and therefore it is not generally possible to estimate the conditional expectation of $E[P_{it}|x_{it}]$ from $x_{it}\hat{\beta}_t$.

In the hedonic regressions literature, it is often assumed that the errors u_{it} are normal and homoskedastic, in which case $E[P_{it}|x_{it}] = \exp(x_{it}\beta_t)\exp(0.5\sigma_u^2)$, where σ_u^2 denotes the variance of u_{it}, which can be estimated as a by-product of the OLS regression of $\ln(P_{it})$ on x_{it}. The problem with this approach is that the assumption of heteroskedasticity is untenable in hedonic regressions since it would imply that individuals have homogeneous tastes. That is, it would imply that the characteristics would have the same value for all individuals, something which is incompatible with the economic theory underlying the hedonic regression (see Triplett, 2000, p. 77).

The parameters in a hedonic function can be interpreted as the conditional expectation of the individual tastes, which vary across individuals. This random parameter variation induces a form of additive heteroskedasticity which cannot be eliminated by transforming the dependent variable. Therefore, no matter what function of P_{it} is used as the dependent variable, the model is expected to be heteroskedastic, precluding the estimation of $E[P_{it}|x_{it}]$ from $x_{it}\hat{\beta}_t$. Naturally, the same kind of problem arises if the researcher estimates a model where the dependent variable is the price and wants to correct an index based on geometric averages. Although these problems have been largely ignored in the hedonic regression literature, they are well known in other fields (e.g., see Mullahy, 1998, and the references therein).

Therefore, except under very strong assumptions, the choice of the functional form of the hedonic regression is indissolubly related to the form of the price index that is used.

In particular, if the index is based on a geometric mean of prices, the regressand has to be the logarithm of the price, whereas if the index is constructed using arithmetic means the regressor has to be the price itself. Notice that the restriction that is imposed on the dependent variable of the hedonic function by the form of the price index is not as strong as it may appear since the right hand side of the function can be as non-linear as necessary. When the purpose of the regression is just to construct a hedonic price index and not the correction of a specific index for quality change, the researcher can choose the functional form of the regression model that he thinks is more suitable, but has to construct the price index in a way that is compatible with the model adopted.

By far, the method that is most often used to construct hedonic price indexes is the so-called dummy variable method, which consists in using data for several periods to estimate a model where the dependent variable is the natural logarithm of the price and which includes as regressors period specific dummy variables. In this case the model has the form

$$\ln\left(P_{it}\right) = x_{it}\beta + \delta_t + u_{it},$$

where δ_t are period specific dummies. Throughout, it is assumed that x_{it} contains an intercept and therefore $\delta_t \equiv 0$ for the base period.

Given the algebraic properties of the ordinary least squares (OLS) estimates, it is clear that $\frac{1}{n_s}\sum_{i=1}^{n_s}\ln\left(P_{is}\right) = \frac{1}{n_s}\sum_{i=1}^{n_s}x_{is}\hat{\beta} + \hat{\delta}_s$. Therefore, defining $t = 0$ as the base period, a price index of the form

$$G_s = \frac{\prod_{i=1}^{n_s} P_{is}^{\frac{1}{n_s}}}{\prod_{i=1}^{n_0} P_{i0}^{\frac{1}{n_0}}}$$

can be written as

$$G_s = \frac{\exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s}x_{is}\hat{\beta} + \hat{\delta}_s\right)}{\exp\left(\frac{1}{n_0}\sum_{i=1}^{n_0}x_{i0}\hat{\beta}\right)} = G_s^{*}G_s^{Q}$$

where $G_s^{*} = \exp\left(\hat{\delta}_s\right)$ is a consistent estimate of the quality adjusted price index and $G_s^{Q} = \exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s}x_{is}\hat{\beta}\right)\Big/\exp\left(\frac{1}{n_0}\sum_{i=1}^{n_0}x_{i0}\hat{\beta}\right)$ is a consistent estimate of the part of the price change that is explained by changes in the characteristics of the products.

In the construction of G_s^* it is implicit that the regression coefficients are constant for all the time periods considered. This assumption is quite strong because it implies that the value of each characteristic remains constant as time passes and technology evolves (see Pakes, 2002). Even if only two adjacent periods are used in the estimation, this assumption can be violated if the product is changing rapidly and yearly data is used.

Fortunately, it is not hard to generalize the dummy variables method to relax the coefficient stability restriction. In fact, estimating the hedonic regression model for the base year, the estimated parameters can be used to evaluate the goods sold in period s at the prices of the base year. In this case the regression model has the form

$$\ln(P_{it}) = x_{it}\beta_t + u_{it}$$

and the estimates of β_t can be used to correct indexes that, like G_s, are based on geometric means. Indeed, using again the fact that OLS residuals have zero mean, it is possible to write $\frac{1}{n_t}\sum_{i=1}^{n_t}\ln(P_{it}) = \frac{1}{n_t}\sum_{i=1}^{n_t} x_{it}\hat{\beta}_t$, and therefore G_s can be written as

$$G_s = \frac{\exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_s\right)}{\exp\left(\frac{1}{n_0}\sum_{i=1}^{n_0} x_{i0}\hat{\beta}_0\right)},$$

which can then be decomposed into

$$G_s = G_s^{\dagger} G_s^{q}, \tag{2}$$

where $G_s^{\dagger} = \exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_s\right) \Big/ \exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_0\right)$ is a consistent estimate of the quality adjusted price index and $G_s^{q} = \exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_0\right) \Big/ \exp\left(\frac{1}{n_0}\sum_{i=1}^{n_0} x_{i0}\hat{\beta}_0\right)$ measures the change in prices that is attributable to changes in the characteristics[1]. In case

[1] Alternatively, if the hedonic model is specified as

$$P_{it} = x_{it}\beta_t + \varepsilon_{it}$$

the estimates of β_t can be used to quality correct price indexes based on arithmetic averages, which can be decomposes as

$$A_s = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s} P_{is}}{\frac{1}{n_0}\sum_{i=1}^{n_0} P_{i0}} = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_s}{\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_0}\,\frac{\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_0}{\frac{1}{n_0}\sum_{i=1}^{n_0} x_{i0}\hat{\beta}_0}.$$

$\hat{\beta}_0$ and $\hat{\beta}_s$ differ only by the value of the constant, then $G_s^{\dagger}$ equals G_s^*, which makes it clear that this is a generalization of the popular dummy variable method. Notice that in practice $\hat{\beta}_s$ does not have to be estimated since, by construction, $\exp\left(\frac{1}{n_s}\sum_{i=1}^{n_s} x_{is}\hat{\beta}_s\right)$ is equal to $\prod_{i=1}^{n_s} P_{is}^{\frac{1}{n_s}}$. It is also worth noting that the bias correction of Teekens and Koerts (1972) should not be used in this context.

The index $G_s^{\dagger}$ effectively compares the values of the characteristics of the products available in period s using the estimated values of the characteristics for periods s and 0. From this point of view, $G_s^{\dagger}$ is preferable to G_s^* as it takes into account the changes in the regression coefficients between s and the base period. Another minor advantage of this index is that, because it is based on the estimation of a regression using data from a single period, the errors of the model can more easily be assumed to be uncorrelated than when the estimation is performed using a sample that includes observations of the same product in different periods of time. On the other hand, in case the parameters are in fact constant during the T periods considered in the sample, estimation is more efficient using the pooled data.

Although the quality corrected index $G_s^{\dagger}$ has been used, for example, by Bode and Van Dalen (2001), it is not frequently used in practice. In fact, it is not even described in Triplett (2000). This is somewhat surprising, not only because of the attractive characteristics of this method, but also because it is based on (2), which is just the well known Oaxaca (1973) decomposition, so often used in labour market hedonic studies.

So far, the case of indexes based on weighted means has not been addressed. Some simple algebra shows that when weighted averages are used the decompositions presented above are still valid if the hedonic regressions are estimated by weighted least squares (WLS) and all the summations in G_s^Q, $G_s^{\dagger}$ and G_s^q are also weighted. Of course, the weights used should be the same that are used in the index. Naturally, it is also true that if the hedonic regression is estimated using WLS, the price indexes constructed from them should be interpreted as indexes based on weighted means. Therefore, the decision of whether or not to use weights in the estimation of the hedonic regression is closely related to the type of index the researcher wants to construct and hence it is not just an empirical question to be decided according to the results of some heteroskedasticity test (see also Machado and Santos Silva, 2001).

3. DATA

Cars are arguably the most complex consumer goods of today. Therefore, any hedonic study of the price of new cars requires a huge amount of information on the characteristics of the car models available. This section describes the main features of the data used in this study.

The information used here comes from two different sources. The prices and characteristics of the new passenger cars available in the Portuguese market were collected by Marketing Systems GmbH[2], a market research company which collects this sort of information both for the motor industry and for leading specialized magazines. The Marketing Systems files contain detailed information on the price and technical characteristics of about 1500 different versions of passenger cars available in the Portuguese market. The information on sales volumes was kindly provided by ACAP, the Portuguese automobile trade association[3]. The information on the prices and characteristics available for this study corresponds to data collected in the month of October from 1997 to 2001. Using the ACAP database, sales were registered for the months of September to November of the same years. In both cases, cars in the database are distinguished by maker, model and version.

Given that the information used was obtained from two different sources, it was necessary to match the two databases. This was a delicate process and great care had to be taken to ensure the correct matching. Fortunately, the ACAP files, besides identifying the different versions, provide the price and some technical characteristics of each version, and this made the matching somewhat easier. Of course, in some cases it was impossible

[2] More information on this company can be obtained at `http://www.marketingsystems.de`.

[3] For more details, see `http://www.acap.pt`.

to obtain a correct match and these observations had to be dropped from the sample. Despite this limitation, the data available corresponds to a total of 267876 units sold in the period from September to November in the five years. This represents more than 90 per cent of total sales over the specific period considered here. However, about 4 percent of these observations were also lost due to incomplete data on the characteristics used in the final specification of the regression model. Therefore, the final sample used was somewhat smaller. The exact number of different models considered in each year is detailed in the next section. Notice that, because some versions had zero sales during the periods considered, they were also effectively dropped of the sample when estimating weighted regressions.

The full set of characteristics on which information is available in the Marketing Systems database is given in table A1 in the appendix and corresponds to those listed in *Automotor*, one of the leading specialized magazines[4]. Although this is a very extensive set of characteristics, it is clear that it cannot possibly account for all the attributes the consumers consider when making their decisions. Moreover, there are subjective aspects, like the styling of the car, which cannot even be measured. Therefore, in contradistinction with the case of other consumer goods like computers, there are always important factors which cannot be included in the construction of hedonic prices for cars.

A final point is worth mentioning. The data on prices available in this database correspond to prices recommended by the manufacturers and include both the value added tax and a specific tax on the sale of new cars. This tax is a function of engine capacity and is regularly updated. In their pioneering study of hedonic price indexes for new cars in the Portuguese market, Santos and Coimbra (1995) modelled the price of the new cars net of the tax on the sale of new cars, but including the value added tax. In the present study, it was decided to consider the price including all taxes since variations in taxation rules are an important source of price variation in this period. Naturally, it would be preferable

[4]More information on this magazine can be found at `http://www.automotor.pt`.

to have data on true transaction prices, but that would require a very different and much more expensive information gathering method.

4. ESTIMATION AND RESULTS

Using the data set described in the previous section, an estimate of the (geometric) average price of new cars sold in period t can be computed as $\tilde{P}_t = \prod_{i=1}^{n_t} P_{it}^{w_{it}}$, where w_{it} denotes the market share of model i. Notice that because the data available are not a sample of prices of cars actually bought but rather it was constructed using list prices reported by the manufacturers, the market shares have to be used as weights to obtain an estimate of the average price of new cars sold in period t. The values of w_{it} used in the estimation were computed as the total sales for model i during the months of September to November of year t, divided by the total sales of the models considered during the same period. The reason to use a geometric mean to estimate the average price has to do with the form of the hedonic regression that was preferred. In fact, as it is often the case, in this study the best regression results were obtained with models in which the dependent variable is the logarithm of the price and, as explained in section 2, these are only compatible with price indexes based on geometric means.

Using $\tilde{P}_t$, a weighted version of the index G_s described in section 1 can be computed as $\tilde{G}_s = \tilde{P}_s/\tilde{P}_0$. Table 1 presents the percentage variation of the sample values of $\tilde{G}_s$ computed as a chain index, that is, using $s-1$ as the base year. This index can be decomposed into the pure price variation and an estimate of the change in quality of the new passenger cars bought. Here, attention is focused on the pure price variation, but it is trivial to get estimates of the quality change.

Three forms of obtaining quality corrected price indexes were considered: the usual dummy variable method, using both pooled and adjacent years data, and the method

Table 1: Variation of sample average prices in percentage

	$s = 1998$	$s = 1999$	$s = 2000$	$s = 2001$
$(\tilde{G}_s - 1) \times 100$	5.714	7.258	6.780	3.207

based on the Oaxaca (1973) decomposition. Because of the need to use weights in the computation of the price index, the estimate of the quality adjusted price index using the Oaxaca decomposition has the form

$$\tilde{G}_s^\dagger = \frac{\exp\left(\sum_{i=1}^{n_s} w_{is} x_{is} \tilde{\beta}_s\right)}{\exp\left(\sum_{i=1}^{n_s} w_{is} x_{is} \tilde{\beta}_0\right)}, \qquad (3)$$

where $\tilde{\beta}_s$ and $\tilde{\beta}_0$ represent the weighted least squares estimates of the parameters in the hedonic regression for the year s and for the base year, respectively. Since the quality corrected price indexes obtained from the dummy variable method do not depend on the value of the covariates, they can be constructed as usual by taking the exponential of the appropriate dummy variable parameters[5]. However, these parameters should also be estimated using weighted least squares.

To obtain quality corrected indexes it is necessary to choose the set of characteristics to be included in the regression. This is a vital process because the results largely depend on the set of characteristics considered. As described in the previous section, the characteristics on which we have data are those detailed in a specialized magazine and therefore should be reasonably representative of the attributes consumers look for when choosing a new car. Although the set of characteristics at our disposal is relatively complete and provides detailed information on the equipment of the vehicles and on their technical specifications[6], there are important characteristics that are not included in this list. For example, the styling of the car and its built quality are certainly important factors for the consumer that are not available in this study and are in general difficult to measure. In order to partially compensate for the lack of these variables, a set of dummies indicating the country of origin of the manufacturer were added as regressors[7]. This set of dummies

[5]Sometimes, a bias correction suggested by Teekens and Koerts (1972) is used in the computation of these estimates. We do not do that here, not only because the correction is asymptotically negligible, but also because it makes the computation of the standard error of the estimate too cumbersome.

[6]See table A1 of the appendix for details.

[7]As in Santos and Coimbra (1995), a single dummy was used for all Eastern European countries.

may also help to reduce the conditional dependence between observations corresponding to cars produced by the same manufacturer.

The choice of which variables to actually include in the preferred model, and the form in which they are included, was made considering two main criteria. Due to the possible coefficient instability, each variable should be statistically significant at the 1 percent level in at least one of the regressions performed using data from a single year, or statistically significant at the 5 percent level in two or more of these regressions. Moreover, the model should pass a simple RESET (Ramsey, 1969) test. Therefore, the decision of which variables to use is intimately related to the functional form of the regression that is used. Naturally, both the t-tests for the individual significance of each regressor and the t-test on which the RESET tests are based were computed using heteroskedasticity consistent covariance matrices (see White, 1980a)[8]. The same set of explanatory variables was used in the pooled data, adjacent years and year-by-year regressions, and only models that are linear in the parameters were considered.

From the first stages of this research it was clear that, imposing linearity in the parameters, models where the dependent variable is the logarithm of the price are clearly preferable to models where the price itself is modelled. Since this situation is typical of hedonic regression studies, no attempt was made to find a suitable non-linear model hav-

[8] The heteroskedasticity robust RESET test was used as the main specification test to evaluate the adequacy of the models since it is easy to compute and has power against several alternatives of interest. In fact, the results in Godfrey, McAleer and McKenzie (1988) show that it has power against alternatives in which the dependent variable is transformed according to the Box-Cox transformation (Box and Cox, 1964), which is particularly interesting in this context. Moreover, and in contradistinction from other tests against Box-Cox alternatives that are sometimes used in this context, this version of the RESET is robust both to heteroskedasticity and to non-normality. Finally, this test can be interpreted as a general test for the adequate specification of the conditional expectation being estimated, which is also interesting in this case because with such a large number of explanatory variables it is easy to neglect important interactions or use inadequate transformations of the continuous variables included. Of course, the fact that a model passes the RESET test does not ensure that it is perfectly specified, but at least it indicates that the specification is reasonably adequate.

ing the price as the dependent variable. Even for models with the logarithm of the price as the dependent variable it was not easy to find a set of regressors satisfying the criteria described above. In particular, models using more than one year of data often failed the RESET test. This is a signal of parameter instability, an issue which will be addressed in more detail below. In view of these results, the set of regressors to use and the form in which they enter the model was chosen on the basis of the results of the year-by-year regressions. The list of variables included in the final specification, as well as information on the form in which they enter the model, is presented in table A2 in the appendix[9]. Table A3 provides basic descriptive statistics for these variables. The results of the RESET tests based on the selected list of regressors are given in table 2.

As noted above, the results of the RESET test suggest some parameter instability. This is an important matter as the choice between the three quality corrected price indexes estimated here depends on the degree of stability of the regression coefficients over time. This problem is frequently addressed in hedonic regression studies (see, for example, Berndt, Griliches and Rappaport, 1995) and the results of Chow (1960) tests are sometimes reported. However, as it is well known, the validity of the Chow (1960) tests depends on the assumption of homoskedastic errors. Therefore, in this context, heteroskedasticity robust versions of these tests should be performed. Here, the heteroskedasticity robust Chow tests were computed using the procedure introduced by Wooldridge (1991). Table 2 contains the results of this test for the regression using the pooled data (1997-2001) and for each of the four adjacent years regressions. Not surprisingly, the stability of the regression coefficients is massively rejected in the case of the pooled regression. In the case of the adjacent years regressions, the null hypothesis of coefficient stability is accepted for the regression using data for 1999 and 2000, but it is either rejected or marginally accepted at the 5 percent significance level for the other years. The rejection is particularly emphatic for the last pair of years. This pattern of results suggests that the coefficient instability is

[9] Besides the variables included in this table, the final model also includes the products of the engine capacity (CC) by the fuel type dummy (Diesel) and by the combined fuel consumption (Cons).

Table 2: Results of the heteroskedasticity robust specification tests

Sample years		**1997-01**	**1997-98**	**1998-99**	**1999-00**	**2000-01**
RESET	Test Statistic	3.13729	2.67852	2.18494	2.13436	1.24072
	p-value	0.00171	0.00739	0.02889	0.03281	0.21471
Chow	Test Statistic	290.775	57.4506	55.3807	45.5519	76.9735
	p-value	0.00000	0.04552	0.06618	0.28834	0.00057
WFF	Test Statistic	301.510	105.611	114.886	169.430	179.304
	p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Number of observations		4384	1394	1643	1869	2098
Sample years		**1997**	**1998**	**1999**	**2000**	**2001**
RESET	Test Statistic	1.28243	1.78858	1.46088	1.21344	0.32063
	p-value	0.19969	0.07368	0.14405	0.22496	0.74849
WFF	Test Statistic	82.8505	72.9105	81.6256	92.1797	86.2449
	p-value	0.00017	0.00217	0.00024	0.00001	0.00007
Number of observations		643	751	892	977	1121

related to changes in taxation. In fact, the taxation on new passenger cars had small changes every year between 1997 and 2000, and was substantially revised for 2001[10]. However, even considering the logarithm of the price net of taxes as the dependent variable of the models, the hypothesis of coefficient stability over the entire period is massively rejected, whereas it is clearly accepted for the 1999-2000 regression, being rejected or marginally accepted at the 5 percent level for the remaining regressions. These results strongly suggest that the hedonic price index based on the year-by-year regressions should be preferred.

As mentioned above, all regressions were estimated using weighted least squares. However, the need to use weights can be tested using the functional form test proposed by White (1980b), as discussed in Machado and Santos Silva (2001). Because it is likely that the hedonic models are heteroskedastic, an heteroskedasticity robust test has to be used.

Since this test can be interpreted as a test for the omission of certain regressors, it can easily be made robust to heteroskedasticity using the method described in Wooldridge

[10]In 2001 the tax benefits for off-roaders and sports utility vehicles were eliminated and the number of tax brackets was reduced from 3 to 2.

(1991), and that is the approach followed here. The results for the White's functional form test (WFF) are reported in table 2 and clearly indicate that the use of weights is necessary.

Finally, the parameter estimates obtained using the different models are not reported here since they are not of direct interest and are difficult to interpret (see Pakes, 2002).

In view of the results obtained, and specially considering the poor performance of the models estimated using more than one year of data, our preferred quality adjusted price index for new passenger cars is the one given by equation (3). This index was computed as a chain index and therefore in the computation of $\tilde{G}_s^{\dagger}$ the base year is $s-1$.

Despite the specification problems of the regressions on which they are based, we also computed the indexes using the dummy variable method with the pooled data and adjacent years regressions. The variation of the three indexes with respect to their value in the previous year are presented in table 3, together with estimates of the respective heteroskedasticity robust standard errors. As a benchmark, table 3 also includes the variation of the CPI component corresponding to the acquisition of new passenger cars[11]. The bottom row of this table reports the average growth rate of the different indexes computed using a geometric mean.

From a statistical point of view, and in spite of the problems with the indexes based on the dummy variable method, they are not very different from the preferred index obtained using the year-by-year regressions. In fact, considering the magnitude of the standard errors of these estimates, the differences between them are, statistically speaking, relatively small. However, in some years the variation of the indexes based on the dummy variable method are more than twice the variation of the index based on the year-by-year regression. This is naturally reflected by the average growth rates which also show important differences. Therefore, the choice of the method used to construct the index can have important economic consequences.

[11] The data on the CPI is that made available by Instituto Nacional de Estatística.

Table 3: Variation of the price indexes in percentage

	Hedonic indexes*			
	Pooled sample	Adjacent years	Year-by-year	CPI
1998	1.127 (0.919)	1.250 (0.709)	0.944 (1.137)	1.717
1999	1.461 (0.910)	1.680 (0.809)	1.076 (0.653)	1.884
2000	0.864 (0.565)	0.933 (0.541)	0.768 (0.583)	2.420
2001	1.298 (0.619)	1.270 (0.574)	0.595 (0.598)	2.756
Average growth rate	1.187	1.283	0.846	2.193

* Heteroskedasticity consistent standard errors in parenthesis

Another remarkable point is that, despite being based on very different information and computed using a different methodology, the variations of this component of the CPI are relatively close to those of the preferred hedonic price index, especially for the first couple of years. However, for the two latest periods, the CPI clearly overstates the changes in the price index for new cars. Indeed, the difference between the CPI and the hedonic price index based on the year-by-year regressions reaches a maximum of 2.161 in 2001. In any case, and considering the values of $\tilde{G}_s$ reported in table 1, it is clear that the CPI managed to eliminate a substantial part of the impact of quality changes. In fact, considering that the total weight of the price index for new cars in the overall CPI is about 7 percent, an overestimation of the variation of the index for this component by as much as 2.161 percentage points implies an overestimation of the overall CPI of approximately 0.15 percentage points in 2001.

It is interesting to notice that in 1997 the methodology of the construction of the CPI was substantially revised (Instituto Nacional de Estatística, 1998). In particular, since 1997, the component corresponding to the acquisition of new passenger cars is computed using information on about 20 different models divided into three categories according to the engine capacity. In each class the best selling models were considered[12]. Clearly, this methodology worked reasonably well until 1999, but the results for more recent years are less satisfactory. This suggests that with the current methodology it is perhaps necessary to regularly update the set of models whose price is considered in the construction of the CPI in order to account for the changes in the market shares of the different models. In any case, the methodology used in the construction of the CPI is being revised and this problem should soon be eliminated.

As a by-product of this analysis, it is also possible to obtain results on the evolution of the quality of new passenger cars sold in Portugal. In fact, comparing the results of tables 1 and 3 it is possible to conclude that the quality of new passenger cars in the Portuguese market increased on average 4.8 percent a year between 1997 and 2001[13]. Interestingly, and despite the very different methodologies adopted in the two studies, this result is similar to that found by Izqierdo, Licandro and Maydeu (2001) for Spain in the period 1997-2000.

5. CONCLUSIONS

In the construction of a hedonic price index a large number of research decisions has to be made and the validity of the results very much depends on the adequacy of these decisions. However, we believe that our main results are relatively robust and that they are not substantially affected by less correct decisions we may have taken.

The results of this study unequivocally show that average quality of the new passenger cars in the Portuguese market is growing at a fast pace and that this phenomenon has to be taken into consideration in the estimation of price indexes. Naturally, the complexity of a study like the one presented here makes it impractical for statistical agencies to use

[12] We are grateful to Daniel Santos and Rui Evangelista of the Instituto Nacional de Estatística for providing this information.

[13] To give an idea of the importance of using weights in this sort of studies, it is interesting to notice that if the information on the sales volume was ignored, the estimate for the growth rate of the average quality would be just 1.6 percent.

these methods on a regular basis to account for the changes in quality when computing price indexes. Therefore, simpler methods are often preferred.

The results described in the last section suggest that the method adopted by the Instituto Nacional de Estatística in the construction of the CPI was reasonably successful in accounting for the quality change of new passenger cars. However, the growing divergence between the hedonic indexes computed here and the results for the CPI suggests that the method adopted in the construction of this index needs to be regularly updated. Despite this, and considering the weight of this component on the overall CPI, the difference between the hedonic indexes and the CPI can represent an overestimation of the overall CPI of up to about 0.15 percentage points in the more recent years.

It would be interesting to extend this type of study to other goods, namely personal computers and telecommunication equipment. However, given that it is difficult to obtain information on the sales volume for the different models of this type of equipment, the possibility of carrying out those studies depends on the implementation of surveys specifically constructed for this purpose.

APPENDIX

Table A1: Vehicle characteristics

GENERAL	Alarm
Body type	Immobilizer
Number of doors	Metallic paint
Boot capacity	Driver air bag
TECHNICAL	Passenger air bag
Fuel type	Electric front windows
Engine capacity	Electric windows front and rear
Max. power	Electric door mirrors
Max. power at (rpm)	Central door locking
Max. torque	Central door locking with remote control
Max. torque at (rpm)	Driver seat with lumbar adjustment
Transmission drive (front/rear/full)	Driver seat with height adjustment
Maximum speed	Driver seat with electric adjustment
Acceleration	Sport seats
Weight/power ratio	Leather upholstery
Fuel tank capacity	Fog lights
Fuel consumption urban	Alloy wheels
Fuel consumption extra-urban	Manual sunroof
Combined fuel consumption	Electric sunroof
Autonomy	Audio preparation
EQUIPMENT	Radio with cassette player
ABS	Radio with CD player
Steering wheel with height adjustment	Foldable back seats
Steering wheel with telescopic adjustment	On-board computer
Air-conditioning	Power assisted steering

Table A2: Definition of the regressors used

NAME	DEFINITION
CntrlLock	1 if has central door locking
CntrlLockR	1 if has central door locking with remote control
Lumb	1 if driver seat has lumbar adjustment
EltrcSR	1 if has electric sunroof
PwrS	1 if has power assisted steering
Comp	1 if has on-board computer
AirCon	1 if has air-conditioning
Immob	1 if has immobilizer
ABS	1 if has ABS
Leather	1 if has leather upholstery
TlscpcAdj	1 if has steering wheel with telescopic adjustment
EltrcW	0 if no electric windows, 1 if front only, 2 if front and rear
Fog	1 if has fog lights
Cassettes	1 if has radio with cassette player
CD	1 if has radio with CD player
Tank	Logarithm of fuel tank capacity
Cons	Combined fuel consumption
TransF	1 if front transmission drive
CC	Logarithm of engine capacity
Diesel	1 if fuel type is diesel
Pwr	Maximum power
MaxS	Maximum speed
Accel	Acceleration
Boot	Boot capacity
Doors	1 if has four or five doors
BR	1 if break
CO	1 if coupé
CA	1 if cabrio
RO	1 if roadster
OR	1 if off-roader or sports utility vehicle
MPV	1 if multipurpose vehicle
D	1 if German
S	1 if Swedish
J	1 if Japanese
K	1 if Korean
USA	1 if from the United States
E	1 if Spanish
East	1 if from an East European country
UK	1 if from the United Kingdom

Table A3: Descriptive statistics for the regressors used

	Unweighted data				Data weighted by sales volume			
Variable	Mean	S. Dev.	Min.	Max.	Mean	S. Dev.	Min.	Max.
CntrlLock	0.36888	0.48254	0	1	0.41585	0.49292	0	1
CntrlLockR	0.66110	0.47338	0	1	0.58382	0.49298	0	1
Lumb	0.30548	0.46065	0	1	0.22031	0.41450	0	1
EltrcSR	0.14990	0.35700	0	1	0.06612	0.24851	0	1
PwrS	0.95434	0.20876	0	1	0.89786	0.30286	0	1
Comp	0.31099	0.46294	0	1	0.17426	0.37937	0	1
AirCon	0.66626	0.47159	0	1	0.34297	0.47475	0	1
Immob	0.95968	0.19672	0	1	0.93203	0.25172	0	1
ABS	0.65283	0.47611	0	1	0.37960	0.48534	0	1
Leather	0.18918	0.39169	0	1	0.03369	0.18044	0	1
TlscpcAdj	0.32357	0.46788	0	1	0.24628	0.43090	0	1
EltrcW	1.43591	0.58341	0	2	1.17853	0.55099	0	2
Doors	0.72433	0.44689	0	1	0.79611	0.40293	0	1
Cassettes	0.47174	0.49924	0	1	0.41989	0.49360	0	1
CD	0.19400	0.39547	0	1	0.10297	0.30396	0	1
Fog	0.63387	0.48179	0	1	0.46112	0.49854	0	1
Tank	4.05733	0.21251	3.40120	4.73620	3.94562	0.16488	3.40120	4.70048
CC	1946.03	815.409	796	6753	1480.53	412.439	796	6750
Diesel	0.27757	0.44784	0	1	0.28044	0.44926	0	1
Pwr	124.681	64.0459	39	485	85.0166	29.5480	39	442
Boot	5.89453	0.39482	4.09434	7.43956	5.77666	0.33899	4.09434	7.43838
MaxS	190.491	28.5690	130	320	171.941	17.9172	130	305
Accel	11.7216	3.02442	4.1	24.8	13.5755	2.50774	4.2	24.8
Cons	7.85822	2.21870	3.0	22.9	6.61047	1.15228	3.0	17.9
TransF	0.78101	0.41360	0	1	0.91097	0.28482	0	1
D	0.31203	0.46336	0	1	0.31103	0.46297	0	1
S	0.04928	0.21646	0	1	0.01186	0.10828	0	1
J	0.17677	0.38151	0	1	0.11190	0.31528	0	1
K	0.04755	0.21284	0	1	0.03720	0.18928	0	1
USA	0.05531	0.22860	0	1	0.07169	0.25801	0	1
E	0.03825	0.19181	0	1	0.04070	0.19763	0	1
East	0.01516	0.12221	0	1	0.01128	0.10564	0	1
UK	0.04394	0.20497	0	1	0.02802	0.16505	0	1
BR	0.17402	0.37916	0	1	0.12138	0.32660	0	1
CO	0.05893	0.23550	0	1	0.01305	0.11351	0	1
CA	0.02412	0.15344	0	1	0.00314	0.05591	0	1
RO	0.02447	0.15450	0	1	0.00352	0.05923	0	1
OR	0.04394	0.20497	0	1	0.04356	0.20414	0	1
MPV	0.07874	0.26935	0	1	0.06741	0.25076	0	1

REFERENCES

Berndt, E.R. (1991). *The Practice of Econometrics Classic and Contemporary.* Addison-Wesley Publishing Company, Reading (MA).

Berndt, E.R.; Griliches, Z. and Rappaport, N.J. (1995). "Econometric Estimates of Price Indexes for Personal Computers in the 1990's", *Journal of Econometrics*, 68, 243-268.

Bode, B. and van Dalen, J. (2001). *Quality-Corrected Price Indexes of New Passenger Cars in the Netherlands, 1990-1999*, Paper presented at the Sixth Meeting of the International Working Group on Price Indexes (Ottawa Group) in Canberra.

Boskin, M.J.; Dulberger, E.R.; Gordon, R.J.; Griliches, Z. and Jorgenson, D. (1996). *Toward a More Accurate Measure of the Cost of Living*, Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index.

Bover, O. and Velilla, P. (2002). *Hedonic House Prices Without Characteristics: The Case of New Multiunit Housing.* European Central Bank, Working Paper No. 117.

Box, G.E.P. and Cox, D.R. (1964). "An Analysis of Transformations". *Journal of the Royal Statistical Society B*, 26, 211-243.

Brown, K.H. (2000). "Hedonic Price Indexes and the Distribution of Buyers Across the Product Space: An Application to Mainframe Computers". *Applied Economics*, 32, 1801-1808.

Chow, G.C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". *Econometrica*, 52, 211-222.

Godfrey, L.G.; McAleer, M. and McKenzie, C.R. (1988). "Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models". *Review of Economics and Statistics*, 70, 492-503.

Hoffmann, J. and Kurz, C. (2002). *Rent Indices for Housing in West Germany 1985 to 1998*, European Central Bank, Working Paper No. 116.

Instituto Nacional de Estatística (1998). *Índice de Preços no Consumidor (IPC$_{97}$) Nova Série (Base 100:1997)*, Instituto Nacional de Estatística/Departamento de Estatísticas Socio-Económicas, Lisboa.

Izqierdo, M.; Licandro, O. and Maydeu, A. (2001). *Car Quality Improvements and Price Indices in Spain.* Estudios Económicos, nº 72, Banco de España, Madrid.

Machado, J.A.F. and Santos Silva, J.M.C. (2001). *Identification with Averaged Data and Implications for Hedonic Regression Studies*, Banco de Portugal, Economic Research Department, WP10-01, Lisbon.

Moch, D. (2001). *Price Indices for Information and Communication Technology Industries - An Application to the German PC Market*, ZEW Discussion Paper No. 01-20, Mannheim

Mullahy, J. (1998). "Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics", *Journal of Health Economics,* 17, 247-281.

Oaxaca R. (1973). "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, 14, 693-709.

Pakes, A. (2002). *A Reconsideration of Hedonic Price Indices with an Application to PC's.* NBER Working paper series, Working Paper 8715.

Ramsey, J.B. (1969). "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis". *Journal of the Royal Statistical Society B*, 31, 350-371.

Santos, E. and Coimbra, C. (1995). "Automobile Prices and Quality Change. An Approach in Terms of Hedonic Prices", *Banco de Portugal Economic Bulletin*, September, 53-60.

Teekens, R. and Koerts J. (1972). "Some Statistical Implications of the log Transformation of Multiplicative Models". *Econometrica*, 40, 793-819.

Triplett, J.E. (1989). "Price and Technological Change in a Capital Good: A Survey of Research on Computers", *in* Jorgenson, D. and Landau, R., eds., *Technology and Capital Formation*, MIT Press, Cambridge (MA).

Triplett, J. (2000). *Handbook on Quality Adjustment of Price Indexes for Information and Communications Technology Products* (draft). Organisation for Economic Co-operation and Development, Paris.

White, H. (1980a). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica,* 48, 817-838.

White, H. (1980b). "Using Least Squares to Approximate Unknown Regression Functions". *International Economic Review*, 21, 149-170.

Wooldridge, J.M. (1991). "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances". *Journal of Econometrics*, 47, 5-46.

WORKING PAPERS

2000

1/00 UNEMPLOYMENT DURATION: COMPETING AND DEFECTIVE RISKS
— John T. Addison, Pedro Portugal

2/00 THE ESTIMATION OF RISK PREMIUM IMPLICIT IN OIL PRICES
— Jorge Barros Luís

3/00 EVALUATING CORE INFLATION INDICATORS
— Carlos Robalo Marques, Pedro Duarte Neves, Luís Morais Sarmento

4/00 LABOR MARKETS AND KALEIDOSCOPIC COMPARATIVE ADVANTAGE
— Daniel A. Traça

5/00 WHY SHOULD CENTRAL BANKS AVOID THE USE OF THE UNDERLYING INFLATION INDICATOR?
— Carlos Robalo Marques, Pedro Duarte Neves, Afonso Gonçalves da Silva

6/00 USING THE ASYMMETRIC TRIMMED MEAN AS A CORE INFLATION INDICATOR
— Carlos Robalo Marques, João Machado Mota

2001

1/01 THE SURVIVAL OF NEW DOMESTIC AND FOREIGN OWNED FIRMS
— José Mata, Pedro Portugal

2/01 GAPS AND TRIANGLES
— Bernardino Adão, Isabel Correia, Pedro Teles

3/01 A NEW REPRESENTATION FOR THE FOREIGN CURRENCY RISK PREMIUM
— Bernardino Adão, Fátima Silva

4/01 ENTRY MISTAKES WITH STRATEGIC PRICING
— Bernardino Adão

5/01 FINANCING IN THE EUROSYSTEM: FIXED VERSUS VARIABLE RATE TENDERS
— Margarida Catalão-Lopes

6/01 AGGREGATION, PERSISTENCE AND VOLATILITY IN A MACROMODEL
— Karim Abadir, Gabriel Talmain

7/01 SOME FACTS ABOUT THE CYCLICAL CONVERGENCE IN THE EURO ZONE
— Frederico Belo

8/01 TENURE, BUSINESS CYCLE AND THE WAGE-SETTING PROCESS
— Leandro Arozamena, Mário Centeno

9/01 USING THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION INDICATOR
— José Ferreira Machado, Carlos Robalo Marques, Pedro Duarte Neves, Afonso Gonçalves da Silva

10/01 IDENTIFICATION WITH AVERAGED DATA AND IMPLICATIONS FOR HEDONIC REGRESSION STUDIES
— José A.F. Machado, João M.C. Santos Silva

2002

1/02 QUANTILE REGRESSION ANALYSIS OF TRANSITION DATA
— José A.F. Machado, Pedro Portugal

2/02 SHOULD WE DISTINGUISH BETWEEN STATIC AND DYNAMIC LONG RUN EQUILIBRIUM IN ERROR CORRECTION MODELS?
— *Susana Botas, Carlos Robalo Marques*

3/02 MODELLING TAYLOR RULE UNCERTAINTY
— *Fernando Martins, José A. F. Machado, Paulo Soares Esteves*

4/02 PATTERNS OF ENTRY, POST-ENTRY GROWTH AND SURVIVAL: A COMPARISON BETWEEN DOMESTIC AND FOREIGN OWNED FIRMS
— *José Mata, Pedro Portugal*

5/02 BUSINESS CYCLES: CYCLICAL COMOVEMENT WITHIN THE EUROPEAN UNION IN THE PERIOD 1960-1999. A FREQUENCY DOMAIN APPROACH
— *João Valle e Azevedo*

6/02 AN “ART”, NOT A “SCIENCE”? CENTRAL BANK MANAGEMENT IN PORTUGAL UNDER THE GOLD STANDARD, 1854-1891
— *Jaime Reis*

7/02 MERGE OR CONCENTRATE? SOME INSIGHTS FOR ANTITRUST POLICY
— *Margarida Catalão-Lopes*

8/02 DISENTANGLING THE MINIMUM WAGE PUZZLE: ANALYSIS OF WORKER ACCESSIONS AND SEPARATIONS FROM A LONGITUDINAL MATCHED EMPLOYER-EMPLOYEE DATA SET
— *Pedro Portugal, Ana Rute Cardoso*

9/02 THE MATCH QUALITY GAINS FROM UNEMPLOYMENT INSURANCE
— *Mário Centeno*

10/02 HEDONIC PRICES INDEXES FOR NEW PASSENGER CARS IN PORTUGAL (1997-2001)
— *Hugo J. Reis, J.M.C. Santos Silva*