The Effects of Quality Adjustment Methods on Price Indices for Digital Cameras

Kari Manninen*
Statistics Finland
8/2004

Abstract

In this paper we test different hedonic and conventional quality adjustment methods in an uniform, but somewhat unconventional, descriptive framework. The main aim is to address questions on hedonic quality adjustment methods and their robustness in index compilation and to give an empirical example with digital camera prices. We will show how conventional quality adjusting methods may be treated parallel with hedonic ones and how these methods may be evaluated similarly with regression based methods. Contrary to structural models that many hedonic quality adjusted price indices are based on, the hedonic models in this paper are all used as forecast models which, we believe, add to the robustness and practical utility of hedonics as a tool for statistical agencies using quality adjustment.

The empirical part of the paper is based on findings from a quarterly digital camera database including some 1,200 prices from over 250 different digital camera models over the years 1998 to 2002. The main findings indicate that, in an aggregate context, such as price index, relatively simple hedonic models may be sufficient for accurate quality controlling even in high technology products. Further, if compared with a matched model framework, the collection of characteristics data for hedonics may not need to exceed the precision already needed in the matched model. This suggests that it may be feasible to use hedonic indices even in high frequency index compilation. To validate this, we claim that additional cross sectional explanatory power from a set of added quality characteristics in hedonic models have only marginal longitudinal effects in the index series.

* I would like to thank Brent Moulton and Marshall Reinsdorf for research support and for useful comments.

Contents:

I Introduction

Constructing quality adjusted price indices for high technology goods have been discussed for years. In this article we take part in the discussion and use digital cameras as an example.

There is a sound mutual understanding among the academia and statistical agencies that, in volume and piece measures, changes in quality of goods and services have to be accounted for. Various regression based methods, or hedonic methods, have been developed and used in a variety of ways to control for quality in index number calculations. Even though theoretical foundations for use of different methods have been made, most notably Fixler and Zieschang (1992), Feenstra (1995), and Diewert (2001), the spectrum of methods still blurs the answer whether the results are a figment of the technique that is used. Especially with high frequency indices such as the CPIs and PPIs, the theoretical justifications are hard to make use operationally. This paper tests the sensitivity of indices for digital cameras to the choice of these techniques. The main finding is that all reasonable methods give approximately the same answer.

In the second section we introduce, within a unified estimation approach, six seemingly different methods; grand unit-value, class unit-value, matched model, time dummy pooled regression, time dummy 2-period regression, and full hedonic regression. We compare the results of applying these methods to the digital camera data. Somewhat similar comparative studies for different data sets have been made at least by Silver (1999), Moulton et al. (1999), Kokoski et al. (1999) and, Silver et. Heravi (2000). These studies mainly focus on differences between matched model and hedonic methods. Few studies are made on differences that result from applying different hedonic methods.

Next, we discuss the advantages and shortcomings of each method. We discuss the implicit assumptions each method imposes and weight how justified these assumptions may be.

In section four we lay out the data for digital cameras and discuss the sample frame and some reservations. In section five present the results for the six different methods introduced earlier. The presentation is the resulting quality adjusted price indices.

Finally, we give conclusions and propose some additional topics to be investigated.

The aim of this exercise is twofold. First, we want to assess the effect of quality adjustment on the price index and to show how the model specification affects the quality adjustment and the index. Second, we want to study the implied accuracy and robustness of the hedonic index and compare the indices based on different types of models.

All indices are treated as changes between two consecutive periods. The index strategy is to chain the index over the periods. In some cases, an equivalent fixed base period index would be identical and hence give exactly the same results.

II Methods to be tested

Our approach is strictly descriptive and does not make any assumptions about market behavior while allowing all possible market interactions to be reduced into a joint distribution of all quality characteristics. We show how this same theoretical base may be used for apparently different quality correction methods.

A) Set-up

Without formal set-up we state that there is a time dependent joint distribution of price and quality characters of all goods considered. In this case we restrict ourselves to various digital camera models that are on the market at each period. A wider treatment of cross dependencies with other goods and services has been left out. Among other things, the unknown distribution fully describes the relation between the price of and all characters that influence the price. Unfortunately we can measure only a fraction of the characters and we have to leave most factors describing the camera markets out. Surely color, brad, appearance, production technology, market competition, regional differences, second hand markets, etc. affect the asking prices and transaction prices. These are, however, all included in the very complex multidimensional distribution.

We are interested in a particular conditional marginal distribution, or more precisely its' first central moment. This is the expected value of price conditioned on measurable quality characters $\boldsymbol{x}$ at certain time period. The conditioning has first been done over all other characters and we are left with a price function distribution. Next, we want to condition this on measurable characters and have the price as a function of these characters.

We describe the price function behind all types of indices simply by[1]

(1) $\left(h(P) \middle| x, t \right) = g^t\left(x\right) + \varepsilon^t,$

where the transformation function h is not generally known[2]. Because of nice properties of symmetry, summation over time, congruence with geometric means, and interpretation, we use natural logarithm as the transformation function for price in all hedonic regression models. Hence we base the price changes on log-changes. To simplify the notation we denote $\log P = p$. Empirical evidence also supports this choice at least in some cases, such as digital cameras here[3]. Now, we are interested in the systematic part of (1), i.e. the conditional expectation

(2) $E\left(\ln P \middle| x, t\right) = E\left(p \middle| x, t\right) = g^t\left(x\right).$

The function g gives the mean log-price for each combination of quality variables. At this point, nothing is assumed on the time specific functions g^t or the independent variables in x.

[1] One could also argue, as usually is done, that the real relation is an inverse of m which we transform this one just for sake of ease of estimation.

[2] This approach of estimated function follows Vartia and Koskimäki (2001).

[3] See e.g. Diewert (2002) and a summary treatment in Triplett (2002) or IMF (2002).

Individual transformation functions (that may differ from identity or logarithm function) of measurable quality characteristics are applied to end up with vectors $x_i^t = \left(1, x_{1i}^{\,t}, \ldots, x_{Ki}^{\,t}\right)$ for each observation, each period. The *K* characteristics are treated as "measured in transformed form" and the question of model specification is not discussed further[4]. One should note that this representation still allows us to use flexible functional forms such as translog and quadratic models[5]. In all models, we assume the estimated function to be linear with respect to the parameters of the transformed original quality variables. Now, let function *f* be simply an estimate of the unknown relation g^t:

$$(3) \quad f^t(x) \equiv est\left[g^t(x)\right] = est\left[E\left(\log P \middle| x, t\right)\right].$$

For our purposes we always use the following linear (with respect to parameters) functional form:

$$(4) \quad f^t(x) = \hat{\beta}'^t x,$$

where the *(K+1)* β–vector includes a constant term.

For the purposes of this study the estimation method used does not have to be OLS as long as it forces the sum of residuals to zero at each period[6]. Regardless of the estimation, we denote $\hat{p}^t_{x_i^{t-1}} = \hat{\beta}'^t x_i^{t-1} = f^t\left(x_i^{t-1}\right)$ as period *t* estimated price for period *t-1* observation *i,* and similarly $\hat{p}^{t-1}_{x_j^t} = \hat{\beta}'^{t-1} x_j^t = f^{t-1}\left(x_j^t\right)$ as period *t-1* estimated price for observation j in period *t.* Also, we denote the estimated log-price change as $\dot{p}$ for both $\hat{p}^t_{x^{t-1}} - p_i^{t-1} = \dot{p}_i$ and $p_i^t - \hat{p}^{t-1}_{x^t} = \dot{p}_i$. Alternatively, we could have used estimated prices on both periods[7].

We define the logarithmic hedonic price indices as weighted geometric means of estimated log-price changes of either base or reference period observations. The estimated log-prices are derived either using the reference period hedonic model for base period observations or base period model for the reference period observations. We have chosen to include also the weights in the table 1 below, although in the empirical part we will only use equally weighted means. The equally weighted geometric mean is also called Jevons index. We call these weighted indices hedonic log-Laspeyres (1) and log-Palgrave (2) respectively. By taking geometric mean of the two indices, we get the formula for hedonic Törnqvist index[8]. Other

---

[4] For functional forms and economic approach to hedonic indices, see Rosen (1974), Diewert (2001) and Triplett (2002). Here the question of model selection and functional form is not discussed further.

[5] See Diewert (2001).

[6] An example of trivial zero sum residual "estimation" of *g* is a relation *f* that gives the observed value for each observation *i*.

[7] De Haan (2003) proposes the use of what he calls double imputation.

[8] It is rather straightforward to show that the index refers to the Törnqvist formula only if the price relatives in (1) and (2) refer to the same (number of) observations. This is not true in the digital camera data. One cannot directly compute the true Törnqvist index without having the weights for each observation on both periods.

index number formulas based on log changes, such as Sato-Vartia or Walsh formulas are not investigated in this study.

Table 1 Hedonic logarithmic indices

Index	Formula	Weight	Estimated price
Log-Laspeyres (1)	$\log P_{t-1}^{t} = \sum_{i=1}^{N^{t-1}} w_i^{t-1}\left(\hat{p}_{x_i^{t-1}}^{t} - p_i^{t-1}\right)$	Period *t-1*	Period *t* using period *t-1* observations
Log-Palgrave (2)	$\log P_{t-1}^{t} = \sum_{i=1}^{N^{t}} w_i^{t}\left(p_i^{t} - \hat{p}_{x_i^{t}}^{t-1}\right)$	Period *t*	Period *t-1* using period *t* observations
"Törnqvist"	$P_{t-1}^{t} = \sqrt{\exp(1)\times\exp(2)}$	Both	Both

B) Decomposition of hedonic price index

This section provides an illustrative way of showing the quantitative effect of hedonic quality adjustment with geometric indices. The proof of formula (X) is given in appendix 2 which also discusses the implications with patched model index. Some of these decompositions are presented in section seven together with regression results from the data.

It can be show that, regardless of the type of the hedonic model, hedonic log-Laspeyres index can be presented as

$$(15)\ P_{t-1}^{t}(La) = \exp\left[\sum_{i=1}^{N^{t-1}}\left(w_i^{t-1}\left(\hat{p}_i^{t} - p_i^{t-1}\right)\right)\right]$$

$$= \frac{G(P^{t})}{G(P^{t-1})}\times\exp\left(\hat{\beta}^{\prime t}\left(\overline{x}^{t-1} - \overline{x}^{t}\right)\right)\times\exp\left(\operatorname{cov}\left(\frac{w}{\overline{w}}, \dot{p}\right)\right)$$

$$= \frac{G(P^{t})}{G(P^{t-1})}\times\exp\left(\hat{\beta}^{\prime t}\left(\overline{x}^{t-1} - \overline{x}^{t}\right)\right)\times\exp\left[\operatorname{cov}\left(\frac{w}{\overline{w}}, \left(\hat{\beta}^{\prime t} - \hat{\beta}^{\prime t-1}\right)x^{t-1}\right)\right]\times\exp\left[-\operatorname{cov}\left(\frac{w}{\overline{w}}, e^{t-1}\right)\right].$$

This decomposition, though slightly differently, was first presented in Koev and Suopera[9]. It says that the quality adjusted log-Laspeyres index may be presented as a product of relative of geometric average prices at two periods, a separable quality correction term that depends on change in average quality, and a sample covariance term between mean adjusted weights and estimated pure log-price change. The last row of (5) breaks the covariance term into two parts; one that shows the effect of model differences at two periods and a term between the adjusted weight and period *t-1* estimation error. In case of a time indicator model, the first correlation term is zero, because the two models are same. If we used double imputation, also the second part of the covariance term would disappear. Then the hedonic log-Laspeyres index would be simply

[9] Koev and Suoperä (2002).

$$P_{t-1}^{t}(La) = \frac{G(P^t)}{G(P^{t-1})} \times \exp\left(\hat{\beta}^{t}{}'\left(\bar{x}^{t-1} - \bar{x}^{t}\right)\right).$$

Also, in case of a time indicator model the coefficient estimate of β is not time specific and hedonic log-Laspeyres and hedonic Törnqvist indices are identical in formula. As said, the quality correction term may be separated into effects of each quality character and each effect may be presented separately:

$$(17)\ \hat{\beta}'\left(\bar{x}^{t-1} - \bar{x}^{t}\right) = \hat{\beta}_1\left(\bar{x}_1^{t-1} - \bar{x}_1^{t}\right) + \ldots + \hat{\beta}_K\left(\bar{x}_K^{t-1} - \bar{x}_K^{t}\right).$$

What is important here is that we can compare the quality correction factors of different model types to be introduced below. Hence, when comparing different quality adjustment methods we could, instead of the actual indices, compare the overall quality correction factors. If the factors are found to be sensitive to the choice of method (or regression model), the most accurate method should be used. The good new would be that, if the quality correction factor is fairly robust to the choice of method, index compiling agency may choose the method that can most easily be implemented in practice. Further, it is easy to see that with equally weighted indices, as elementary aggregate indices usually are, only changes in average quality matter.[10]

C) Methods

We separate six different types of hedonic models based on the model specification. Regardless of the model type, indices may always be calculated using the above formulas even if it may not be the most efficient way. The different methods we are going to apply in the later sections are the following.

I) Grand unit-value

Quality adjustment is ignored and price index is simply the geometric average of price relatives. The model does not depend on the quality characters and reduces to a constant, but may still be estimated as a regression model:

$$(5)\quad f^{t}(x) = a^{t}.$$

The quality correction factor is simply 1.

II) Class unit-value

In this case some quality differences are taken into account. Class unit-value method is analogous to grouping the goods into subgroups and following the class averages. Using regression model we write

$$(6)\ f^{t}(x) = a^{t} + b_1^{t}D_1 + b_2^{t}D_2 + \cdots + b_C^{t}D_C = a^{t} + b_c^{t} = a_c^{t}$$

[10] This decomposition may of course be extended to any log-based index formula and also mixed methods, such as patched model (see Pakes (2002). However, these are not discussed further in this study.

where D_c^t is a binary indicator variable for qualities in class c. The function *f* returns the mean price a_c^t in class c. The quality correction factor depends on chances in the distribution of observations into different classes. This can be seen from (x), where the change in average quality $\left(\bar{x}^{t-1} - \bar{x}^t\right)$ is now calculated from dummy variables representing the classes.

III) Matched model

As its name suggest, matched model follows the prices of same goods. This can be presented in terms of estimation function *f* returning the observed price for each period *t*. A possible notation of this is

$$(7) \quad f^t(x) = p_i^t \ \ \forall j \in \left(p \middle| x = x_i^t\right).$$

The quality correction factor is again just one, since there is no quality changes. It can be shown that for a patched model the quality correction factor is otherwise identical but also depends on the share of missing/new items and average estimation error between matched and unmatched observations.

IV) Time dummy pooled hedonic regression

This is the first 'real' hedonic model and maybe the most widely used hedonic specification. Quality characters are assumed to affect price similarly across time and the level transitions in time account for price chances. Using the first period as a reference the model may be written as

$$(8) \ f^t(x) = a + b'x + \beta_1 D^2 + \beta_2 D^3 + \cdots + \beta_{T-1} D^T = a + b'x + \beta_t = a^t + b'x.$$

The quality correction for all 'real' hedonic methods is given by (5).

V) Time dummy 2-period hedonic regression

The 2-period hedonic regression method is basically the same as the previous. The first difference is that instead of using all periods in the estimation, we now relax the temporary restrictions on the quality coefficients over longer time period. Another difference is, although rarely considered, that we may change the quality characters included in the regression as well.

VI) Full hedonic imputation

This method lets the price-quality -relation to change freely from period to period. Leaving aside the matched model, all other methods are special cases of full hedonic imputation. The model is simply the same as in (4):

$$(9) \quad f^t(x) = a^t + b^{\prime t} x.$$

These are the methods, or model specifications that we are going to apply to the digital camera data. We expect that rapidly evolving product space and rapid technological change in quality would prove some of the methods clearly unsatisfactory and some clearly better. Before doing this we shortly discuss some general pros and cons of each method.

III Advantages and Disadvantages of the Methods

All six methods may be practical is some cases and some are clearly not. Some methods clearly depend on more information that has to collected.

I Grand unit-value
The grand unit-value index is a good method for homogenous items and it is used as elementary indices in some CPIs and PPIs. Especially, if the goods are truly homogenous and prices volatile the unit-value method may be appropriate. It escapes the difficulty of pricing point in time. However, in elementary aggregate case, the quality is more often controlled by the matched model approach, which would of course give the same result when calculated as a geometric mean.

II) Class unit-value
The Class unit-value method uses only categorical variables. In essence, by grouping the observations into subgroups we eliminate the quality differences between these subgroups, just as in analysis of variance. The quality adjusted price index may be constructed directly from class means implied by the classification (and cross classification).

III) Matched model
The matched model is more of an approach than a quality adjustment method. It is safe to same that majority of national CPIs and PPIs are based on the matched model way of thinking. The quality adjustment is seen as a separate problem only when we are missing the match. The advantage of collecting prices of same finely defined products from same stores are obvious. We can simultaneously control for some immeasurable quality characters without explicitly stating them.

The major drawback of the matched model approach is that the sample may become unrepresentative if not updated frequently enough. It incurs a sample selection bias. Matched model fails to account for new models that are introduced, as well as the old ones for which no matches can be found. If we don't just ignore the non-matches, some quality adjustment method is still needed.

Statistical agencies usually collect replacement items if the initial specification of a good are not found in consecutive periods. Then regression, or hedonic, methods are use to adjust the price of this new (or the missing) item.

IV) Time dummy pooled hedonic regression
This is the "classical" use of regression analysis with price indices and is what people often think when they perceive hedonics.

There are clearly some good properties that make time dummy models useful. First, the calculation is fairly simple and the index may also be constructed solely based on the estimated time indicator coefficients. Especially in research type studies that examine historical data, compared to updating an index series form time to time, Secondly, as Tripplett (2002) proposes, the regression coefficient in the reduced regression models are often relatively stable and time indicator models work fairly well.

However, there are drawbacks as well. One should always confirm that the coefficients indeed are stable. With high frequency indices and rapidly evolving commodity population sudden changes may go unnoticed.

The time indicator model does not have to be Using the formulas in table 1 including the weights, we can derive a "Törnqvist" index for time indicator model as well. It is needless to say that the resulting index is identical to log-Laspeyres and log-Palgrave indices.

V) Time dummy 2-period hedonic regression

Now instead of estimating just one regression model for 20 period data we estimate 19 models. The index could also be deduced directly from the time dummy indicators.

Time dummy pooled and 2-period pooled regression. The quality adjusted price index may be constructed directly from the estimates of time coefficients and, as mentioned before, the quality variables may be binary or continuous.

VI) Full hedonic imputation

We argue that this should be the basic starting point for all hedonic price index studies. From the standpoint of the resulting index all previous methods are special cases or simplifications of full hedonic imputation method. This, of course requires that the weight data is either derived from elsewhere or the data used for the regression model does not use e.g. permutations of "same items".

In addition to these basic methods, one may want to use combinations of them. E.g. it may not be feasible to estimate the whole regression separately each period because data restrictions. However, it may be imperative to let the coefficients of some central. An example in the digital camera context could be that the coefficient for picture accuracy (measured in megapixels) changes in time while the effect of brand stays constant in medium term. In the framework above this may be done simply by defining the estimation function accordingly. Although this was tried, the results are not reported in this article since no remarkable differences were observed. All methods give reasonable results if the assumptions are sufficiently fulfilled.

Hedonic methods that are used as structural models to "correct" the observed price for changes in some quality characters do not easily fall into any of these categories. However, it may also be written in terms of the above estimation function $\boldsymbol{f}$ so that for observation $\boldsymbol{i}$ it returns the observed price plus quality correction

$$f^t(x) = p_i^t + b'\left(x_i^t - x_i^t\right).$$

A variation of this method and used with missing observations is patched model (See Pakes, 2002), that avoid the dangers of structural correction. Patched model uses imputation for those observations that are missing, or where the quality has changed, for the period in question and matching prices for observations that can be found in both periods[11].

We don't want to promote the use of hedonics as in (12) and assign interpretations for the regression coefficients. We started from the complex joint distribution of all measurable and immeasurable characters and come down to, at best a simplified, set of characters and a reduced form of expectation. By this time, we cannot really claim that the individual coefficients themselves represent, if we may borrow from economic jargon, "shadow prices" without any knowledge of the true distribution. The price forecast, on the other hand, is just as good as the reduced form regression model.

IV Description of the Data Set

The quarterly price data were recorded from various sources, mainly from issues of Journal of Popular Photography on microfilm. For two last quarters of 2002, the price data were collected from the internet at *www.pricescan.com*. The model specific quality characteristics were compiled mainly from the website *dpreviw.com*.

Total of over 1300 prices of 288 different digital camera models were collected. Some of these prices were averaged over advertising retailers at the time of entry. Although the share of these multiple observations was not recorded, it accounts for approximately 10 to 15 per cent of the total sample. Regardless of the number of observations used for each recorded model price quote, they are treated as single observations.

I) The sample design

It was not feasible to use random sampling in the study because of scarce data sources. For the early years, all possible models with a price quote in the Journal were recorded. When the price collection was changed to internet, generally the lowest price was recorded and almost all available makes and models were included. There were no effort to follow same models, and market entries and exits occurred when new models were first advertised or they were no longer available. In other words, the sample was not designed to mimic any typical statistical agency approach.

After the data collection, we decided to exclude SLR[12] type digital cameras and also restrict the time sample to the years 1999 through 2002. The SLR-type digital cameras are used by professionals and they are much more expensive and have partly different properties and characteristics. The data from years 1996 and 1997 proved to be too limited.

---

[11] The term Patched model was used by Pakes (2001).
[12] SLR - single lens reflex.

The quality characteristics may have some variation between retailers. However, just one reference attribute is used throughout the lifespan of a model over all retailers. These differences could not be observed in detail and are most likely limited to different memory cards included in cameras. For this reason, memory was excluded from some of the regression models.

The number of priced models (and hence price observations) follow the number of adds in the Journal and the number of different models each manufacturer makes. One might argue that the data is self-weighting in a sense that the more models a manufacturer has, the larger its share is in the data. No explicit data was available for the weights and all indices are calculated as equally weighted geometric means. In digital camera case, we assume that the number of different models advertised is a proxy for market share of that make. This view may be challenged, but no other data was available to support or contradict this hypothesis.

II) Data reservations

There are some issues in the data that might be a hindrance to the index:

1) *"Call for price"*. A number of times the Journal adds do not show the price directly, and they could only be obtained by calling the retailer. This practice is still in use and may be set by either the manufacturer or the retailer.
2) *Weight data*. There were no data on model nor manufacturer turnover or other data on relative importance of manufacturers.
3) *Unbalanced samples*. The price sample is unbalanced towards the last year and the last two quarters when the price collection changed fundamentally.
4) *Limited data on the early years.*

These concerns were treated by the following practices, which are commonly used. However, their effects should always be estimated.

1) *"Call for price"*. These prices were not collected at all. Our assumption is that the "Call for price" is more widely used with new introductory models than older models. Although the basis for reasoning is not relevant in our descriptive indices, one may argue that models entering the market have higher mark-ups while models exiting are being sold out with smaller or even negative markups. Since we are not referring to marginal cost –pricing, as we are only interested in the actual prices paid, we assume that there are no differences in the price determining processes (or marginal distributions) between the advertised and non-advertised prices.
2) *Weight data*. As already noted, since we had no data on quantities sold by model or by manufacturer, we assumed that the number of models advertised provides a proxy for the manufacturer weights. Thus, each model has an equal weight. Out of all observations Sony counted for most (18%) and then Olympus (15%), Kodak (13%), Fuji (12%), Canon (11%) and Nikon (8%) of total observations. The rest 13 manufacturers account for the rest 24 % of observations. This self-weighting seems reasonable – especially since there were not great differences in pricing between

manufacturers.[13] However, it would be very interesting to see how model specific weights would have affected the series.

3) *Unbalanced samples.* There is a clear change in the number of observations per quarter when the data source was changed to the internet after the second quarter of 2002. There may be a systematic reduction in prices due to this change, since the data selected from Pricescan.com usually refer to "Best price" which is the lowest advertised price within a selection of online retailers. These prices usually do not include shipping and there may be a tendency for the low price retailers to charge more for the shipping than others. However, this is clearly not the case every time and the same may also be true for the Journal adds as well. The degree of possible bias from this has not been quantified. There were total of 1052 price quota for the years 1998 – 2002 and the distribution of observations is presented in table 3.1.

Table 3.1 The distribution of price observations

Quarter / Year	1998	1999	2000	2001	2002
Q1	19	41	79	75	50
Q2	20	34	44	29	94
Q3	27	54	63	83	155
Q4	37	31	55	41	124
Total	103	160	241	228	423

The change in the data collection is likely to have an effect also on the index. The prices collected from the internet are usually the lowest prices for each model and not averaged over advertised prices, as is the case in the price data from the Journal adds. By collecting overlapping prices, one could estimate the magnitude of a shift change but this exercise was not carried out. Another feature of internet purchases is the shipping and handling fees. Although not included in the prices, it could be argued that in some cases a part of the actual price is actually charged as shipping and handling, which is rather evident with regard to some special offers of other consumer goods. However, a quick sample did not confirm the negative relationship between price of the good and the handling fees for same models.

4) *Limited data on the early years.* There are additional 21 price observations for 1996 and 38 for 1997 that have not been used in the analysis, because these data include missing information on the characteristics. With a distinctive model or other method the series could be extended a few years back, but for this study this exercise was not carried out. Also, if the time period were further reduced to, say the last three years, differences between quality adjusted and unadjusted price indices would not appear so large[14].

These questions, as important as they may be, are not addressed further in this study. Our main purpose was to compare different methods of controlling quality and the index series

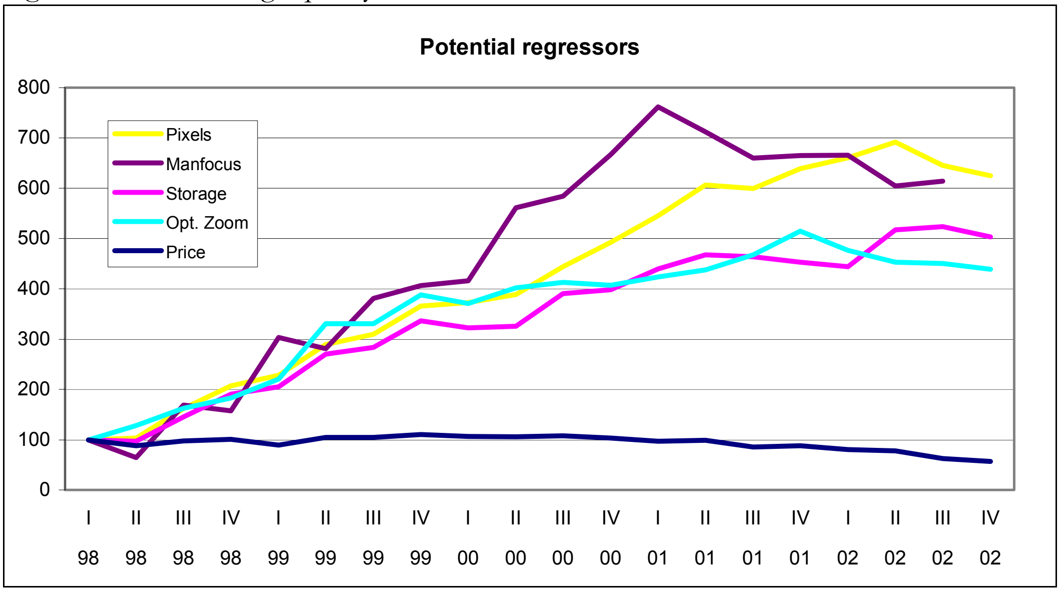
[13] As tested with classification models.

[14] This will become apparent later.

they produce. Some of the questions could have been solved with some complementary data or additional data collection.

III) Quality characters
The regression models to be laid out in section 5 need input variables, namely regressors. Potentially good explanatory variables for hedonic regression would be such that have variability in time and relatively large spread within each time period. Additionally, they should be correlated with the price and not heavily with each other. In figure 4.2, there are some potential variables to be used in the regression models in form of a simple average relative change thus showing only the variability in time.

Figure 4.2 Some average quality variables as indices



Changes in the average quality characteristics are large in time compared to the evolution of average prices. Of these variables, the manual focus is a binary variable and the index presentation should be interpreted as the evolution of the share of digital cameras having the feature in question. When used in average form e.g. in imputation, all binary variables should be interpreted similarly. All available variables in the data set are presented in table 2 in Appendix 1. The most useful are the ones indicating a sharpness of the picture, a memory capacity, an optical zoom ratio, and manual focus, an external flash and movie options.

V Results

As a group digital cameras (compact or ultra compact models) is a rather homogenous set of product varieties (compared to some other transactions in the economy). For many high technology goods, the average price does not seem to change very much but at the same time average (technological) quality characteristics change considerably. This is true for digital cameras too.

I) Grand unit-value

The price for a typical new digital camera model often starts with a stable introductory price (may be set by the manufacturer) and then the dispersion of offer prices becomes larger in time. Often the highest asking price stays the same (or decreases moderately) while the lowest price declines sharply. In the data, there is just one price for a model at one time, but additional sales information would be interesting to obtain.[15]

Without controlling for quality changes the average price fluctuates around $500 until early 2001 and then drops to under $350 during the last year. The grand unit-value index refers to the geometric means. Changes in quarterly gross unit prices are presented in figure 4.1.

Figure 4.1 Unit value index series (1998 = 100)

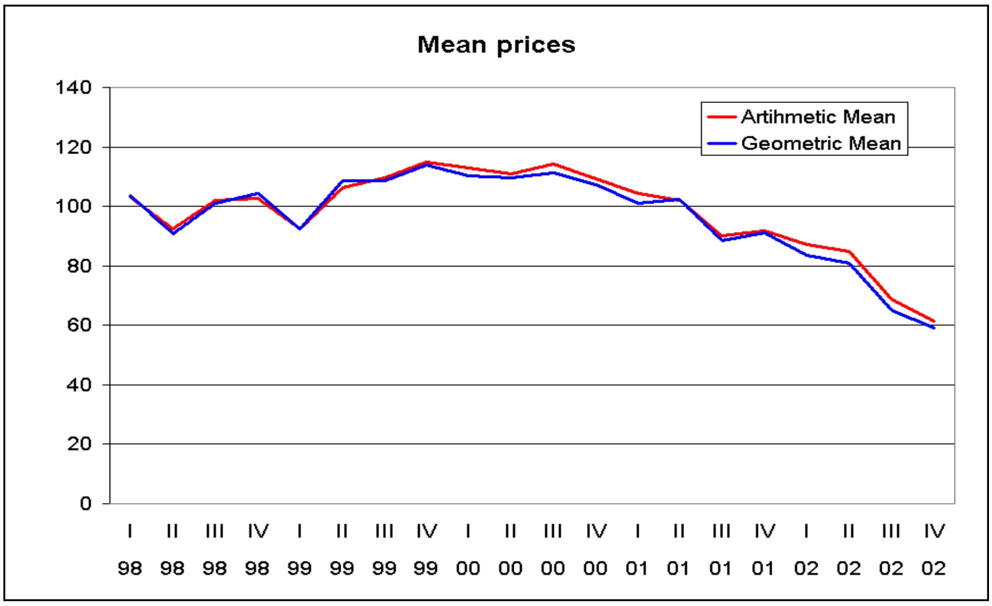


Figure 4.1 confirms the intuition we had on digital cameras. The prices of 'a good camera' have been around $500 for some years and just during the last few years the prices have begun to drop, though less than the picture would indicate[16]. Of course, these prices do not count for the changes in performance and other characteristics of the equipment. Equally importantly, these average prices do not follow the prices of same camera models. The drop during the last few quarters may be partly explained with the change in data collection – both because the recorded price refers to the lowest price and also because there may be more 'low end' models in the data set.

In terms of the model in section 2, calculating unadjusted geometric means is equal to "estimating" a hedonic regression model

[15] See e.g. www.pricescan.com.This property brings some additional problems for the price index that were not accounted for in section 3. The data collected does not allow to take into account this behavior.

[16] We believe the same goes with some other high-tech equipment, e.g. PC's. For some years now you've got a 'good computer' – not the very state of the art, but the next best thing – for some $1500.

$$(13') \; f^t(x) = \hat{\beta}'^t x^t = \hat{\alpha}^t .$$

This can be estimated from pooled data consisting all quarters and the resulting model fit measured in adjusted R^2 is little over 20% and most of individual t-test statistics for time indicator coefficients suggest that they do not deviate from *0*[17]. We call this a trivial model. It is actually used by some statistical agencies in imputing missing observations, though not usually explicitly in this form.

II) Class unit-value

The usual practice statistical agencies use in tackling the changes in quality comes as a by-product of the sampling or selection process. Just follow the prices of same goods in time! In case of digital cameras, the 'same' would mean the same store and same model of the same make. No measurable quality change to control for – by definition. This could be interpreted as using a tight classification where each model forms one class. The index is based on the class average prices. For the sake of illustration, we classify (ex post) the cameras to similar (homogenous) groups and calculate the group mean price changes. This classification method is actually often used by statistical agencies for missing observations or replacements and may be a good method in connection with some products.

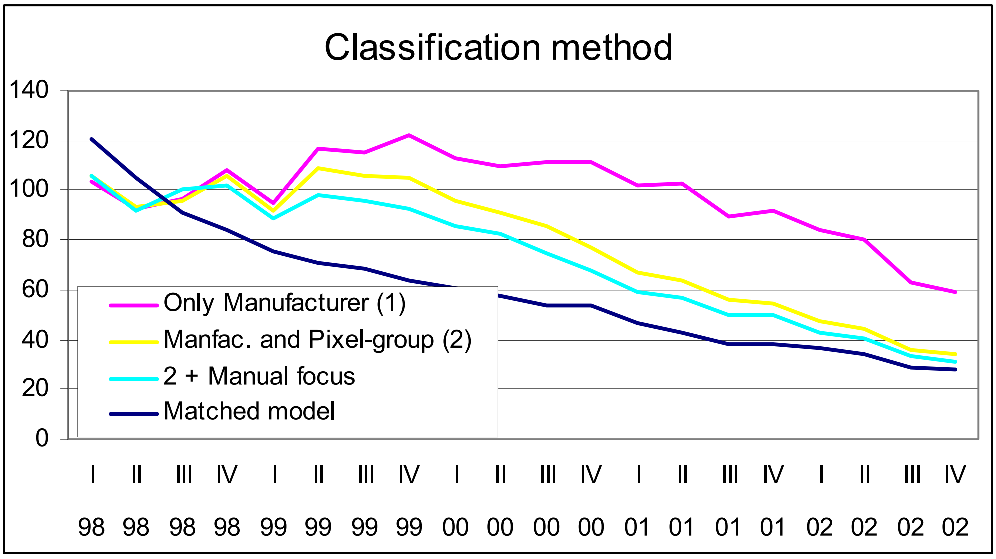
The classification method is based on classifying the cameras according to some rule – most likely by their characteristics – and calculating the class means. The matched model index above is an extreme case of this method. Each model is classified as its own group and 'empty' classes appear every time when the model is not found in the next period.

A more realistic case would be to use, for example, the manufacturer as a classification rule. With the classification, one hopes to remove as much as possible of the within class price variation at each time. This method is typically used in repeated measurement experiments in natural sciences and called analysis of variance. It is usually not called that way in the price index context and usually none of the available test statistics from the analysis are either estimated or presented.

The first classification model in figure 5.2 is based on the manufacturer. The index may be calculated from the changes in make-specific average prices[18]. The second model adds pixel group – a variable, which classifies the camera models into five categories according to the available accuracy of the picture (less that 1 megapixel, 2 Mp, 3Mp and over 4 Mp cameras). The third model further classifies the data to models with or without a manual focus option (autofocus is the norm). As with unadjusted averages, we do not actually calculate the class means but instead use regression models (without cross effects).[19]

[17] Actually t-tests for H0-assumption of all individual α:s being zero would be rejected at 5% confidence level for only last two quarters if the first quarter is used as a reference.
[18] The actual calculation is based on a time dummy regression model with indicator variables for each manufacturer.
[19] So we are actually not cross-classifying the models, but only using the 'main effects'.

Figure 5.2 The classification method



Compared to the matched model index, these indices have more volatility around the trend in the first two years. Adding the classifying factors clearly smoothens the series, but practical usability suffers because the number of class means to be calculated grows exponentially and empty classes start to appear.

The classification method may be presented in the notation of section II as follows:

$$(14)\ f^t(x) = \hat{\beta}^{\prime t} x^{t} = \alpha + \beta_1 D_{MANUF} + \beta_2 D_{MF} + \cdots + \sum_{t=2}^{T} \gamma^t D_t,$$

where Ds are binary variables for classifying variables and time indicator variables. When estimated from pooled data the R-squares for the three models above are 33, 62 and 69% respectively. As already mentioned, the index series may be constructed directly from the time indicator coefficients. However, the same resulting series is also reached by estimating the price for each observations and using the index number formulas presented in table 1.1.

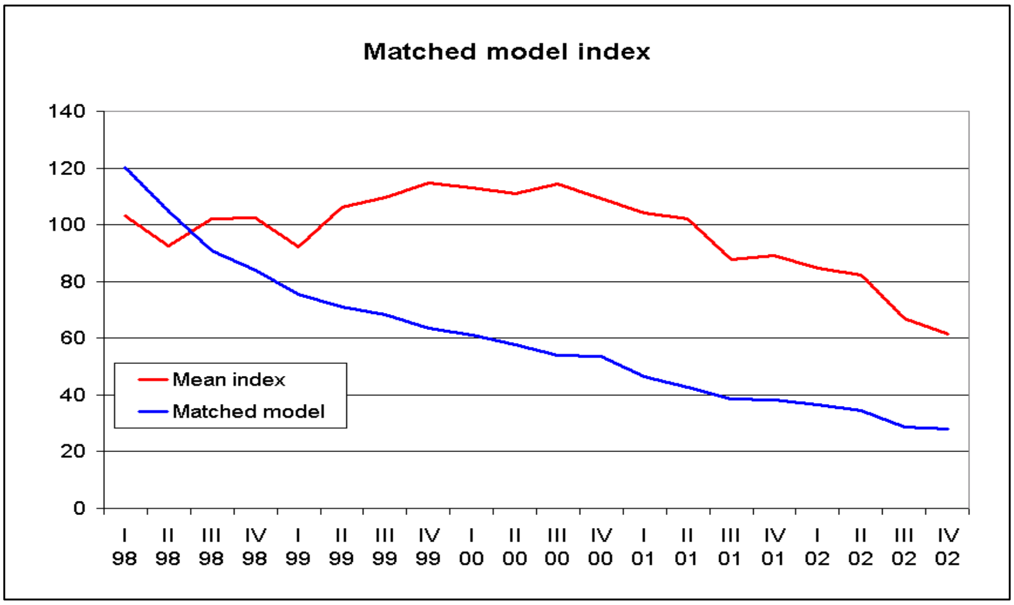
III) Matched model

Before using the classification method, we first calculate a digital camera price index based on a matched model method. As mentioned above in section 3.1, the sampling frame was not meant to be used for calculation of a statistical agency –type matched model index. In a sense, this sampling could be described as "quarterly re-sampling"[20]. Since we did not initially plan to calculate the matched model index at all, it is provided only as an example and methodological criticism should not focus on inadequacy of this matched model method. Also, no patched models were compiled from the data.

[20] It would not be true to claim that the samples were truly independent from one another since we used the same magazine having mostly the same advertisers over time. However, the notion of independence should not be too far from true and we would expect it to have only minor effect for the price index.

There were observations from at least two quarters for almost all of the 288 models in the data set. However, the turnover of camera models was rather fast. The number of models for which price were found over more than 4 quarters' period was 167 and over 6 quarters' just 52. In traditional statistical agency practice, this would have meant a large number of replacement models to be found and an alternative method to account for those models at times of no price observation. Our matched model index does not include estimates for the missing models, either for the ones entering or exiting. The observed price change over more than one quarter is divided by the number of quarters and addressed only to the first quarter. With these reservations, the (log mean) index series are presented in figure 4.1 together with an unadjusted average price index. As can be seen, especially in the first six to eight quarters, the two methods differ considerably, and the matched model index is much smoother in decline. We will get back to some interpretation in section 6.

Figure 5.1 The matched model index and simple average (1998 = 100)



As already noted in section II, all models use natural logarithmic of price for the dependent variable and either logarithmic or identity transformation functions for the explanatory variables. In the first sub-section we introduce the index formula decompositions. I the next section we estimate models restricting the coefficients of all quality characteristics to be constant over time, while in 5.3 these restrictions are relaxed.

IV) Time dummy pooled hedonic regression
In this section all period *t* models are estimated as

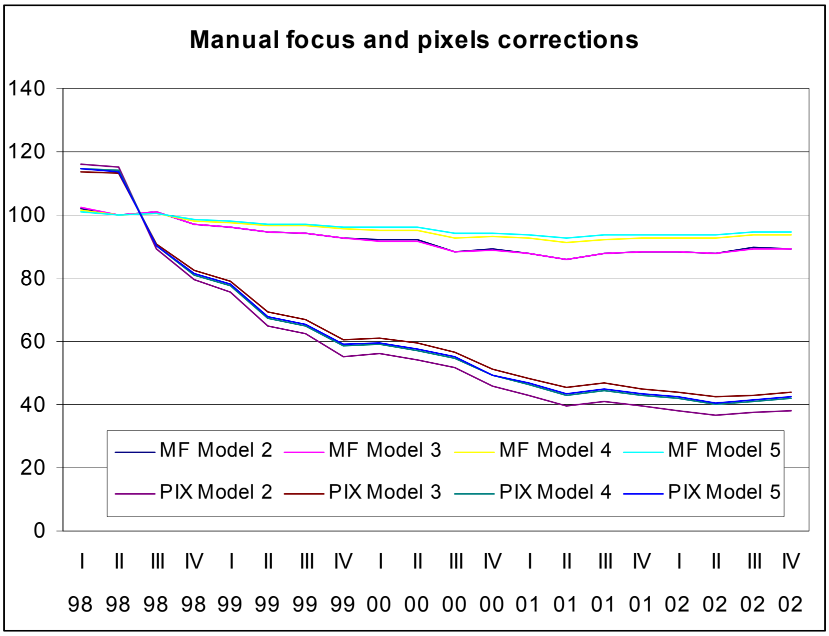
$$(18) \quad \ln\left(\hat{p}_i^t\right) = \hat{\beta}^t x_i^t = \hat{\alpha} + \hat{\beta}_1 x_{i1}^t + \cdots + \hat{\beta}_K x_{iK}^t + \sum_{t=1}^{T} \hat{\delta}^t D_t ,$$

where an indicator variable D_t gets value 1 at period *t* and 0 otherwise. We will use four slightly different models to illustrate how the model selection affects the quality correction and the index. The models are:

Model 1: manual focus + ln(megapixels) [R-square 66%]
Model 2: manual focus + ln(megapixels) + ln(megabytes) [75%]
Model 3: manual focus + ln(megapixels) + optical zoom [77%]
Model 4: manual focus + ln(megapixels) + optical zoom + ln(megabytes) + external flash [80%]

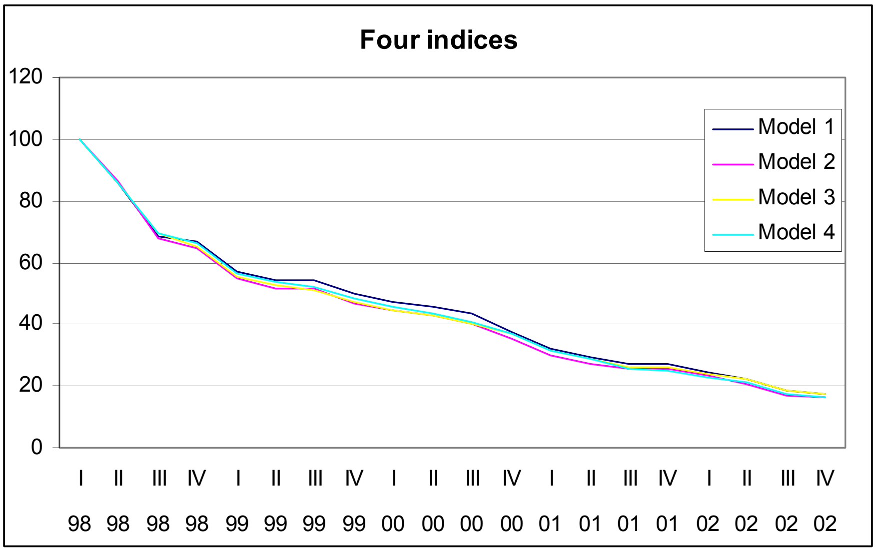
Adding more quality characteristics increase the overall model fit somewhat, and the differences with quality correction factors from each model become even smaller. A summary table for the coefficients and quality correction terms of the below models are presented in appendix 4. In figure 6.1 quality correction terms for two characteristics, manual focus and pixels, are presented.

Figure 6.1 Quality correction factors for four models



As one can see there is some variation between the four models when the OLS adjusts the hyperplane in price-quality coordination. The feature of forecast model is that by adding explanatory variables into the model the individual coefficients, and quality correction factors, adjust so that best overall fit is achieved. This means that the individual quality corrections factors contribute a part of their value to the new variable depending on the amount of multicollinearity it has with the variables already in the model. However, when taken together with all quality characteristics in the model, the total quality correction factor may have very little 'dispersion' between models, as in figure 6.2.

Figure 6.2 Quality adjusted log-Laspeyres indices for the four models



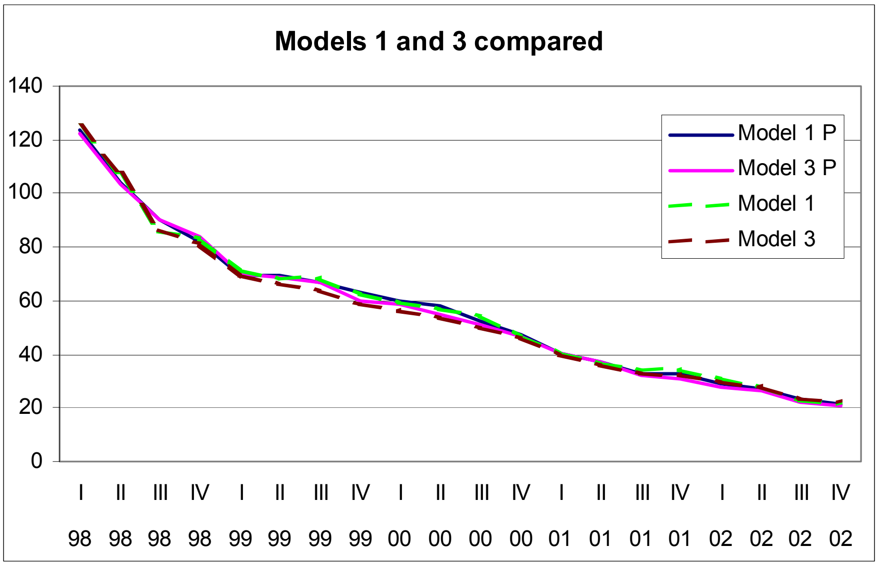
The overall picture of the quality adjusted price index for 1999 – 2002 does not change when adding new quality variables into the hedonic regression. Also, for practical reasons it may be feasible to collect even high frequency data on just few quality characteristics together with an existing price collection, e.g. the CPI.

V) Time dummy 2-period hedonic regression

In case of rapid quality change, the assumption that quality – price relation stays the same except from the constant term over a long period of time should be questioned. We will use two different methods to allow more flexibility in the models. The first one is to apply the time indicator model to data that pools data together only two consecutive periods, and estimate 19 independent time indicator models (for all pairs). The second is to estimate separate models for each 20 quarters and impute the matching prices as suggested by the index formulas. With the latter we calculate different log-Laspeyres and "log-Palgrave" indices using the model from period *t* and *t-1*, respectively and present the hedonic Törnqvist price index. We call these models pairwise pooled and full hedonic models.

As the results will show, again there are no large changes in the quality adjusted indices and thus we will use just two models, Model 1 and Model 3 from the previous section.
Now, the model R2:s vary between 60 and 85%. See a summary table of estimation results in Appendix 4. The resulting index series for pairwise pooled indices are presented in figure 6.3 together with ones from the previous section. The P refers to pairwise model, and as one can see, the two quality correction magnitudes are very similar with the completely pooled data models.

Figure 6.3 Pooled estimation models



In this case individual observations have much more effect for the coefficients, especially in the early years. Consequently, the quality correction factors for individual characteristics do vary little more, but the total quality correction factors are not affected as much, as expected. One could modify this method by adding the number of consecutive periods to the estimation, which would have a further smoothing effect but still gradually take into account possible changes in quality – price relation.

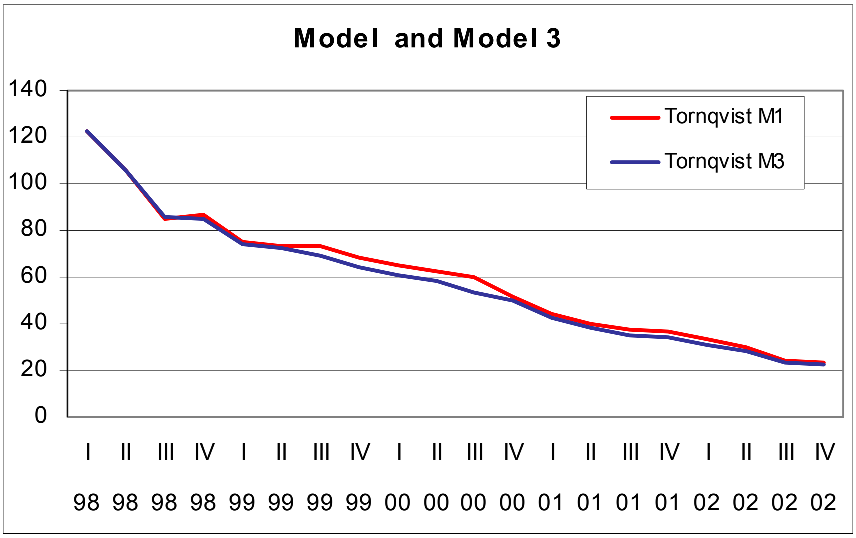
VI) Full hedonic imputation

The second method could be called a true full hedonic method. Estimated models stay the same but instead of calculating just one index we will use all the data and present the resulting quality adjusted price indices as a Tornqvist index[21]. This would theoretically be the most comparable with an index based on time indicator model[22]. See result in the summary table in Appendix 4. The two indices are presented in figure 6.4.

---

[21] Difference of chained Laspeyres and Palgrave are at most 7 index points and average to very close to 0.
[22] Since both use the data from the two periods to estimate the model(s). See appendix 1.

Figure 6.4 Full hedonic models



While it may be difficult to see real difference in the quality adjusted price indices, there are some differences in the quality correction factors. The difference between the two model pairwise quality correction factors is some 8% at maximum. With the pooled data the difference is at most 6%. With individually estimated Fisher quality correction factors between the two model differ again at most some 8%.

VII) Summary
Tähän yhteenvetotaulu

VI Conclusions

As shown above, regardless of different methods and hedonic models used to adjust for quality changes in digital camera price index, all reasonable models produce indices that come very close to each other. Since we apply a forecast model we prefer a simpler model if no substantial benefit is achieved from adding more variables into the hedonic model. If one has to choose between a simple model and more complicated one we think the simpler is better.

As a starting point we should expect and allow the models differ between periods. Only if there is no strong evidence against changing coefficients, we may use time restricted models. If data allows the use of pooled regressions, hedonic models seem to be rather robust in choice of estimation and of model specification. The advantages from using somehow pooled data are that it gives more stable regression coefficients and 'smoothens' the quality adjusted index series. Especially with high frequency indices it may also make the hedonics more feasible since it demands less data. This is again not a bad thing.

Since relatively simple models may work well enough, hence large scale characteristics collection may not be necessary, and may not be any greater restraint for statistical agencies than a matched model approach.

As long as the matched model index does not produce outside the sample bias it works very well. However, if the distribution of characteristics changes in time, as it does with high technology products, frequent sampling is needed and quality changes in mismatches must be dealt somehow. Typical statistical agency procedures may not be suitable for simultaneously dealing with quality change and sampling. We argue that hedonic approach is a good and often feasible way to produce indices so that sampling may be separated from the quality adjustment process.

The main findings indicate that, in an aggregate context, such as price index, relatively simple hedonic models may be sufficient for accurate quality controlling even in high technology products. Further, if compared with a matched model framework, the collection of characteristics data for hedonics may not need to exceed the precision already needed to "make the match". This suggests that it may be feasible to use hedonic indices even in high frequency index compilation.

As Pakes (2004), we see no real obstacles for using some form of hedonic approach instead of more fixed matched model approach. To validate this, we claim that additional cross sectional explanatory power from a set of added quality characteristics in hedonic models have only marginal longitudinal effects in the index series.

References

de Haan, J. (2003), Direct and Indirect Dummy Approaches to Hedonic Price Measurement, Revised version of paper presented at International Working Group on Price Indices, Paris

Diewert, W. E. (2001), Hedonic Regressions: A Consumer Theory Approach, Discussion Paper No.: 01-12, University of British Columbia, Vancouver

Diewert, W. E. (2002), Hedonic Producer Price Indexes and Quality Adjustment, Discussion Paper No.: 02-14, University of British Columbia, Vancouver

Heravi, S. and Silver, M. (2002), On the Stability of Hedonic Coefficients and their Implications for Quality-Adjusted Price Change Measurement, Cambridge, Mass: NBER Summer Institute 2002.

Hulten C. R. (2002): Price Hedonics: A Critical Review. NBER Summer Institute 2002, Cambridge, Mass.

Feenstra, R.C. (1995), Exact Hedonic Price Indexes, Review of Economics and Statistics, LXXVII, 634-54.

Gordon, R.J. (1990), The Measurement of Durable Goods Prices, Chicago: University of Chicago Press

IMF (2002): PPI Manual

Koev, E. and Suoperä, A. (2002), Omakotitalojen ja omakotitalotonttien hintaindeksit 19985=100, Statistics Finland report.

Koskimäki, T. and Vartia, Y. (2001),

Vartia, Y.O. and Vartia, P.L.I. (1984), Descriptive Index Number Theory and the Bank of Finland Currency Index, Scandinavian Journal of Economics 86 (3), 352-64.

Pakes, A. (2002): A reconsideration of Hedonic Price Indexes with an Application to PC's, Harward University.

Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentation in Pure Competition

Silver, M. and Heravi, S. (2001), Scanner Data and the Measurement of Inflation

Triplett, J.E. (2002), Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products, OECD Directorate for Science, Technology and Industry, Draft, OECD Paris

Appendix 1. Decomposition of hedonic geometric indices

It was proposed that hedonic log-Laspeyres index may be decomposed as:

$$(A1)\ \ P_{t-1}^{t}(La)=\exp\left[\sum_{i=1}^{N^{t-1}}\left(w_i^{t-1}\left(\hat{p}_{x_i^{t-1}}^{t}-p_i^{t-1}\right)\right)\right]$$
$$=\frac{G(P^t)}{G(P^{t-1})}\times\exp\left(\hat{\beta}^{\prime t}\left(\overline{x}^{t-1}-\overline{x}^{t}\right)\right)\times\exp\left(\operatorname{cov}\left(\frac{w}{\overline{w}},\dot{p}\right)\right).$$

To see this, we develop the basic hedonic log-Laspeyres formula by adding and subtracting estimated equally weighted pure price change

$$(A2)\ \ P_{t-1}^{t}=\exp\left[\sum_{i=1}^{n}\left(w_i^{t-1}\left(\hat{p}_i^{xt}-p_i^{t-1}\right)\right)\right]$$
$$=\exp\left[\sum_{i=1}^{n}\left(\left(w_i^{t-1}-\frac{1}{N^{t-1}}\right)\left[\hat{p}_{x_i^{t-1}}^{t}-p_i^{t-1}\right]\right)+\frac{1}{N^{t-1}}\sum_{i=1}^{n}\left(\hat{p}_{x_i^{t-1}}^{t}-p_i^{t-1}\right)\right]$$
$$=\exp\left[\sum_{i=1}^{n}\left(\left(w_i^{t-1}-\frac{1}{N^{t-1}}\right)\left[\hat{p}_{x_i^{t-1}}^{t}-p_i^{t-1}\right]\right)\right]\times\exp\left(\frac{1}{N^{t-1}}\sum_{i=1}^{n}\left(\hat{p}_{x_i^{t-1}}^{t}-p_i^{t-1}\right)\right).$$

Now manipulating the second part by again adding and subtracting period *t* average price it may be written as:

$$(A3)\ \ \exp\left[\frac{1}{N^{t-1}}\left(\sum_{i=1}^{N^{t-1}}\hat{p}_{x_i^{t-1}}^{t}-\sum_{i=1}^{N^{t-1}}p_i^{t-1}\right)\right]$$
$$=\exp\left(\hat{\beta}^{\prime t}\,\overline{x}^{t-1}-\overline{p^{t-1}}+\left(\overline{p^{t}}-\overline{p^{t}}\right)\right)$$
$$=\exp\left(\overline{p^{t}}-\overline{p^{t-1}}\right)\times\exp\left(\hat{\beta}^{\prime t}\,\overline{x}^{t-1}-\hat{\beta}^{\prime t}\,\overline{x}^{t}\right)$$
$$=\frac{G\left(P^t\right)}{G\left(P^{t-1}\right)}\times\exp\left(\hat{\beta}^{\prime t}\left(\overline{x}^{t-1}-\overline{x}^{t}\right)\right).$$

The first term is the relative of geometric means of the two periods' prices. It is an unadjusted or unit price index from period *t-1* to *t.* The second term is a multiplicative quality correction term that may be further factored into each characteristic. After estimating the regression coefficients, this equally weighted decomposition may easily be used for index calculation, since the quality correction term only depends on the average quality change.

The second term of (A1) may be written as

$$(A4)\quad \exp\left[\sum_{i=1}^{N^{t-1}}\left(w_i^{t-1}-\frac{1}{N^{t-1}}\right)\left[\hat{p}^t_{x_i^{t-1}}-p_i^{t-1}\right]\right]$$

$$=\exp\left[\frac{1}{N^{t-1}}\sum_{i=1}^{N^{t-1}}\left(\frac{w_i^{t-1}}{\overline{w}^{t-1}}-1\right)\left[\hat{p}^t_{x_i^{t-1}}-p_i^{t-1}\right]\right]$$

$$=\exp\left[\operatorname{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}},\left[\hat{p}^t_{x_i^{t-1}}-p^{t-1}\right]\right)\right]$$

$$=\exp\left[\operatorname{cov}\left(\frac{w}{\overline{w}},\dot{p}\right)\right]$$

Also, to emphasize the hedonic model, we could further decompose the covariance term into *systematic and random* parts. Since the average weight depends on the number of observations, we may also write $\operatorname{cov}\left(w/\overline{w},\dot{p}\right)=\operatorname{cov}\left(N^{t-1}w,\dot{p}\right)$. To see the effect of selecting the model type this covariance term may further be written as

$$(A5)\quad \exp\left[\operatorname{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}},\left[\hat{p}^t_{x_i^{t-1}}-p^{t-1}\right]\right)\right]$$

$$=\exp\left[\operatorname{cov}\left(\frac{w}{\overline{w}},\left[\hat{p}^t_{x_i^{t-1}}-\hat{p}_i^{t-1}-e_i^{t-1}\right]\right)\right]$$

$$=\exp\left[\operatorname{cov}\left(\frac{w}{\overline{w}},\hat{p}^t_{x_i^{t-1}}-\hat{p}_i^{t-1}\right)+\operatorname{cov}\left(\frac{w}{\overline{w}},-e_i^{t-1}\right)\right]$$

$$=\exp\left[\operatorname{cov}\left(\frac{w}{\overline{w}},\left(\hat{\beta}'^{t}-\hat{\beta}'^{t-1}\right)x^{t-1}\right)+\operatorname{cov}\left(\frac{w}{\overline{w}},-e_i^{t-1}\right)\right].$$

Now, it's easy to see that for any time indicator model, for which regression coefficients stay constant, the covariance is between the weights and forecast error. Errors of course depend on how many periods are used in the estimation of time indicator model.

To derive a symmetric hedonic index that makes use of both period weights and regression models we start with repricing period *t* observations using period *t-1*model. It is easy to show that this Palgrave type formula is almost the same:

$$(A6)\ P^t_{t-1}(Pa)=\exp\left[\sum_{i=1}^{N^t}\left(w_i^t\left(p_i^t-\hat{p}^{t-1}_{x_i^t}\right)\right)\right]=\cdots$$

$$=\frac{G\left(P^t\right)}{G\left(P^{t-1}\right)}\times\exp\left(\hat{\beta}^{t-1}\left(\overline{x}^{t-1}-\overline{x}^t\right)\right)\times\exp\left[\sum_{i=1}^{N^t}\frac{1}{N^t}\left(w_i^t-\frac{1}{N^t}\right)\left[p_i^t-\hat{p}^{t-1}_{x_i^t}\right]\right]$$

$$=\frac{G\left(P^t\right)}{G\left(P^{t-1}\right)}\times\exp\left(\hat{\beta}^{t-1}\left(\overline{x}^{t-1}-\overline{x}^t\right)\right)\times\exp\left(\operatorname{cov}\left(\frac{w}{\overline{w}},\dot{p}\right)\right).$$

Finally, the hedonic Törnqvist index may be obtained as a geometric average of the two. The traditional Törnqvist index formula uses arithmetic mean of weights as

$$(A7)\ P_{t-1}^{t}(T\ddot{o}) = \exp\left(\sum_{i=1}^{N}\left(w_i^t + w_i^{t-1}\right)/2\left(p_i^t - p_i^{t-1}\right)\right).$$

But since we may have a different number of observations in the two periods we define the hedonic Törnqvist index as geometric mean of hedonic log-Laspeyres (A1) and the current period weighted "Palgrave-type" index (A6). Using notation $\overline{\hat{\beta}}$ for average of the two period estimated regression coefficients it is rather straightforward to show that:

$$(A8)\ P_{t-1}^{t}(T\ddot{o}) = \sqrt{P_{t-1}^{t}(La)\times P_{t-1}^{t}(Pa)} = \frac{G\left(P^t\right)}{G\left(P^{t-1}\right)}$$

$$= \frac{G\left(P^t\right)}{G\left(P^{t-1}\right)}\times\exp\left(\overline{\hat{\beta}}'\left(\overline{x}^{t-1} - \overline{x}^{t}\right)\right)\times\exp\frac{1}{2}\left(\mathrm{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}}, \hat{p}_{x^{t-1}}^{t} - p^{t-1}\right) + \mathrm{cov}\left(\frac{w^t}{\overline{w}^t}, p^t - \hat{p}_{x^t}^{t-1}\right)\right),$$

where the last term could be simplified - if the number of observations would stay the same – as

$$(A9)\ \frac{1}{2}\left(\mathrm{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}}, \hat{p}^t - \hat{p}^{t-1}\right) + \mathrm{cov}\left(\frac{w^t}{\overline{w}^t}, \hat{p}^t - \hat{p}^{t-1}\right)\right) + \frac{1}{2}\left(\mathrm{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}}, -e^{t-1}\right) + \mathrm{cov}\left(\frac{w^t}{\overline{w}^t}, e^t\right)\right)$$

$$= \mathrm{cov}\left(\frac{1}{2}\frac{w^{t-1}}{\overline{w}^{t-1}} + \frac{1}{2}\frac{w^{t-1}}{\overline{w}^{t-1}}, \hat{p}^t - \hat{p}^{t-1}\right) + \frac{1}{2}\left(\mathrm{cov}\left(\frac{w^t}{\overline{w}^t}, e^t\right) - \mathrm{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}}, e^{t-1}\right)\right)$$

$$= \mathrm{cov}\left(w_T, \dot{p}\right) + \frac{1}{2}\left(\mathrm{cov}\left(\frac{w^t}{\overline{w}^t}, e^t\right) - \mathrm{cov}\left(\frac{w^{t-1}}{\overline{w}^{t-1}}, e^{t-1}\right)\right).$$

In (A9) the weights are now Törnqvist weights and also the regression coefficients used in the quality adjustment term follow the Törnqvist, in a sense that they are arithmetic means. Actually, any index formula based on log-change may be decomposed in the above way, just the covariance terms associated with the weighting scheme change.

Appendix 2. Average prices and the classification method

Table A1.1 Average prices

Year	Quarter	Quarterly mean prices $	Mean index
98	I	528.9	103.0
98	II	474.5	92.4
98	III	524.1	102.0
98	IV	527.0	102.6
99	I	474.4	92.4
99	II	546.0	106.3
99	III	562.4	109.5
99	IV	590.1	114.9
00	I	580.5	113.0
00	II	569.4	110.9
00	III	587.0	114.3
00	IV	561.4	109.3
01	I	535.7	104.3
01	II	524.2	102.1
01	III	463.6	87.7
01	IV	471.6	89.2
02	I	447.3	84.6
02	II	436.0	82.4
02	III	353.4	66.8
02	IV	315.8	61.5

Table A1.2 Data set variabels

Variable	Description	type of measure
lnp	log of price	dollars
lnpix	log of sharpness	megapixels
lnsto	log of memory included	megabytes
movie	movie feature	0 - 1 variable
remote	remote control	0 - 1 variable
flash_ex	external flash	0 - 1 variable
manfocus	manual focus	0 - 1 variable
zoomo	optical zoom	scale of optical magnification
zoomd	digital zoom	scale of digital magnification
USB	usb connection	0 - 1 variable
serial	serial connection	0 - 1 variable
bat_re	battery recharger	0 - 1 variable
type	type of camera	compact, ultacomp, SLR-type
multires	choices of various resolutions	number, or '0 - 1 variable
ISO	number of different iso	number, or '0 - 1 variable
manufac	manufacturer	18 manufacturers

Table A1.3. Classification method

Year	Quarter	Only Manufacturer (1)	Manufacturer and Pixel-group (2)	2 + Manual focus
1998	I	103.4	105.4	106.0
1998	II	92.3	93.3	91.9
1998	III	96.4	95.4	100.4
1998	IV	108.0	105.8	101.7
1999	I	94.6	92.0	88.9
1999	II	116.5	108.6	97.9
1999	III	114.8	106.0	95.3
1999	IV	122.2	104.9	92.6
2000	I	112.8	96.0	85.4
2000	II	109.8	91.3	82.8
2000	III	111.3	85.3	74.9
2000	IV	110.9	77.3	67.9
2001	I	102.0	66.5	59.0
2001	II	102.5	63.9	56.4
2001	III	89.6	55.8	49.9
2001	IV	91.9	54.4	49.5
2002	I	83.9	47.7	43.1
2002	II	80.1	44.1	40.4
2002	III	63.3	36.0	33.5
2002	IV	58.7	33.9	31.5

Appendix 3. Some regression results

Time indicator Model 4 estimation results. The model is estimated as

$$\ln\left(\hat{p}_i^t\right) = \hat{\beta}^t x_i^t =$$

$$\hat{\alpha} + \hat{\beta}_1 LNPIX + \hat{\beta}_2 LNSTO + \hat{\beta}_3 MANFOCUS + \hat{\beta}_4 ZOOMO + \hat{\beta}_5 FLASH\_EX + \sum_{t=10}^{28} \hat{\delta}^t Q_t$$

The SAS System 09:33 Monday, May 12, 2001

The REG Procedure
Model: Model 4
Dependent Variable: lnp

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	24	170.35142	7.09798	160.64	<.0001
Error	938	41.44658	0.04419		
Corrected Total	962	211.79800			

Root MSE	0.21020	R-Square	0.8043
Dependent Mean	6.05545	Adj R-Sq	0.7993
Coeff Var	3.47133		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.56490	0.07584	86.57	<.0001
lnpix	ln(PIXEL)	1	0.50031	0.02129	23.50	<.0001
Manfocus	Manfocus	1	0.13551	0.01722	7.87	<.0001
ZoomO	Opt Zoom	1	0.06494	0.00562	11.56	<.0001
lnsto	ln(STORAGE)	1	0.12207	0.01539	7.93	<.0001
Flash_ex	Ext flash	1	0.11777	0.01872	6.29	<.0001
Q10		1	-0.16623	0.09449	-1.76	0.0789
Q11		1	-0.34812	0.08709	-4.00	<.0001
Q12		1	-0.49367	0.08241	-5.99	<.0001
Q13		1	-0.69502	0.08191	-8.48	<.0001
Q14		1	-0.77254	0.08240	-9.38	<.0001
Q15		1	-0.80633	0.07965	-10.12	<.0001
Q16		1	-0.90879	0.08450	-10.76	<.0001
Q17		1	-0.93108	0.07829	-11.89	<.0001
Q18		1	-0.97396	0.08171	-11.92	<.0001
Q19		1	-1.07752	0.08038	-13.41	<.0001
Q20		1	-1.15870	0.08123	-14.26	<.0001
Q21		1	-1.29944	0.08044	-16.15	<.0001
Q22		1	-1.37906	0.08708	-15.84	<.0001
Q23		1	-1.50989	0.08095	-18.65	<.0001
Q24		1	-1.52072	0.08508	-17.87	<.0001
Q25		1	-1.59569	0.08398	-19.00	<.0001
Q26		1	-1.65902	0.08160	-20.33	<.0001
Q27		1	-1.83195	0.07983	-22.95	<.0001
Q28		1	-1.90307	0.08005	-23.77	<.0001

Appendix 4. **Model 3 and 4 coefficients**

Models 3 and 4 estimated from total and pairwise pooled data

Table A4.1 Parameter estimates for pairwise dummy model and full dummy model 3

Year	Qrt	Intercept	lnpix	Manfocus	ZoomO	Time D
98	I					
98	II	6.81	0.65	0.07	0.06	-0.167
98	III	6.52	0.52	0.13	0.07	-0.135
98	IV	6.33	0.42	-0.06	0.08	-0.079
99	I	6.25	0.43	0.04	0.07	-0.153
99	II	6.09	0.49	0.05	0.08	-0.025
99	III	6.05	0.54	0.14	0.07	-0.047
99	IV	5.99	0.53	0.21	0.07	-0.078
00	I	5.89	0.58	0.16	0.08	-0.037
00	II	5.84	0.60	0.15	0.08	-0.045
00	III	5.83	0.54	0.17	0.07	-0.085
00	IV	5.76	0.55	0.14	0.07	-0.089
01	I	5.66	0.60	0.14	0.07	-0.141
01	II	5.51	0.61	0.16	0.06	-0.089
01	III	5.41	0.59	0.08	0.09	-0.110
01	IV	5.28	0.57	0.10	0.10	-0.015
02	I	5.30	0.51	0.21	0.08	-0.083
02	II	5.15	0.53	0.26	0.09	-0.054
02	III	5.13	0.53	0.23	0.08	-0.186
02	IV	5.03	0.46	0.23	0.08	-0.063
Full model		**6.69**	**0.53**	**0.18**	**0.08**	

Table A4.2 Parameter estimates for pairwise dummy model and full dummy model 4

Year	Qrt	Intercept	lnpix	lnsto	Manfocus	ZoomO	Flash_ex	Time D
98	I							
98	II	6.68	0.65	0.18	0.00	0.04	0.00	-0.153
98	III	6.46	0.51	0.06	0.10	0.07	-0.12	-0.153
98	IV	6.28	0.41	0.04	-0.06	0.08	-0.02	-0.088
99	I	6.17	0.42	0.06	-0.02	0.07	0.03	-0.184
99	II	5.97	0.47	0.08	0.07	0.05	0.12	-0.039
99	III	5.78	0.34	0.11	0.16	0.07	0.22	-0.019
99	IV	5.75	0.35	0.08	0.18	0.10	0.20	-0.095
00	I	5.55	0.47	0.13	0.12	0.11	0.13	-0.015
00	II	5.60	0.53	0.08	0.09	0.12	0.09	-0.054
00	III	5.52	0.46	0.11	0.08	0.13	0.06	-0.089
00	IV	5.37	0.44	0.16	0.06	0.11	0.10	-0.074
01	I	5.40	0.55	0.14	0.11	0.04	0.11	-0.143
01	II	5.26	0.57	0.14	0.14	0.02	0.13	-0.087
01	III	5.22	0.64	0.09	0.05	0.06	0.11	-0.134
01	IV	5.11	0.68	0.07	0.04	0.06	0.13	-0.027
02	I	5.15	0.78	0.01	0.14	0.04	0.09	-0.094
02	II	4.88	0.68	0.10	0.15	0.06	0.08	-0.071
02	III	4.79	0.56	0.15	0.13	0.06	0.12	-0.170
02	IV	4.80	0.49	0.11	0.18	0.05	0.10	-0.073
Full model		**6.56**	**0.50**	**0.12**	**0.14**	**0.12**	**0.06**	

Table A4.3 Parameter estimates for full hedonic model 3

Year	Qrt	Intercept	Inpix	Manfocus	ZoomO
98	I	6.75	0.58		0.07
98	II	6.72	0.71	0.06	0.05
98	III	6.35	0.47	0.13	0.07
98	IV	6.25	0.38	-0.11	0.08
99	I	6.09	0.48	0.13	0.07
99	II	6.08	0.51	0.00	0.08
99	III	5.99	0.54	0.22	0.07
99	IV	5.91	0.51	0.20	0.07
00	I	5.84	0.61	0.14	0.08
00	II	5.79	0.57	0.17	0.08
00	III	5.77	0.53	0.16	0.07
00	IV	5.66	0.59	0.11	0.08
01	I	5.51	0.60	0.17	0.06
01	II	5.42	0.65	0.16	0.06
01	III	5.28	0.60	0.06	0.10
01	IV	5.30	0.52	0.18	0.09
02	I	5.22	0.50	0.24	0.08
02	II	5.06	0.55	0.28	0.09
02	III	4.98	0.53	0.20	0.08
02	IV	5.01	0.39	0.26	0.07

Table A4.4 Parameter estimates for full hedonic model 4

Year	Qrt	Intercept	Inpix	Manfocus	ZoomO	Insto	Flash_ex
98	I	6.63	0.59		0.05	0.16	
98	II	6.58	0.71	-0.01	0.03	0.19	
98	III	6.36	0.48	0.12	0.08	0.00	-0.05
98	IV	6.12	0.36	-0.11	0.09	0.08	-0.02
99	I	6.01	0.48	0.04	0.07	0.05	0.04
99	II	5.80	0.47	0.12	0.03	0.16	0.14
99	III	5.79	0.33	0.18	0.10	0.06	0.23
99	IV	5.58	0.34	0.18	0.10	0.12	0.18
00	I	5.50	0.52	0.09	0.12	0.13	0.11
00	II	5.72	0.58	0.09	0.13	-0.02	0.05
00	III	5.28	0.38	0.06	0.13	0.19	0.08
00	IV	5.40	0.53	0.07	0.07	0.14	0.10
01	I	5.26	0.57	0.15	0.02	0.14	0.13
01	II	5.19	0.57	0.13	0.03	0.14	0.13
01	III	5.08	0.66	0.01	0.07	0.08	0.10
01	IV	5.15	0.73	0.07	0.05	0.02	0.18
02	I	5.04	0.80	0.17	0.04	0.00	0.00
02	II	4.72	0.62	0.14	0.07	0.15	0.12
02	III	4.67	0.54	0.13	0.05	0.15	0.11
02	IV	4.88	0.43	0.24	0.05	0.06	0.07

Table A4.5 Average quality characteristics

Year	Qrt	lnp	lnpix	Manfocus	ZoomO	lnsto	Flash_ex
98	I	6.22	-0.96		0.59	0.85	
98	II	6.09	-0.93	0.09	0.75	0.82	
98	III	6.20	-0.49	0.06	0.96	1.22	0.06
98	IV	6.23	-0.23	0.15	1.08	1.49	0.04
99	I	6.11	-0.14	0.14	1.30	1.56	0.14
99	II	6.27	0.10	0.28	1.95	1.84	0.34
99	III	6.27	0.17	0.26	1.95	1.89	0.33
99	IV	6.32	0.33	0.35	2.28	2.06	0.42
00	I	6.29	0.35	0.37	2.18	2.02	0.34
00	II	6.28	0.39	0.38	2.36	2.03	0.35
00	III	6.30	0.53	0.51	2.43	2.21	0.41
00	IV	6.26	0.63	0.53	2.39	2.23	0.33
01	I	6.20	0.73	0.61	2.49	2.33	0.33
01	II	6.21	0.84	0.69	2.58	2.39	0.46
01	III	6.07	0.83	0.65	2.75	2.38	0.34
01	IV	6.10	0.89	0.60	3.03	2.36	0.34
02	I	6.01	0.92	0.60	2.80	2.34	0.26
02	II	5.98	0.97	0.60	2.66	2.49	0.22
02	III	5.76	0.90	0.55	2.65	2.50	0.18
02	IV	5.66	0.87	0.56	2.59	2.46	0.20

Table A4.6: Quality correction factors

Year	Qrt	Model 2P	Model 2Full	Model 4P	Model 4Full	Model 2	Model 4
98	I						
98	II	-0.04	-0.05	-0.02	-0.05	-0.03	-0.02
98	III	-0.24	-0.25	-0.26	-0.30	-0.27	-0.31
98	IV	-0.11	-0.16	-0.12	-0.18	-0.12	-0.13
99	I	-0.06	-0.06	-0.06	-0.09	-0.06	-0.06
99	II	-0.18	-0.20	-0.20	-0.26	-0.18	-0.20
99	III	-0.03	-0.03	-0.02	-0.04	-0.03	-0.03
99	IV	-0.13	-0.13	-0.14	-0.16	-0.13	-0.14
00	I	-0.01	-0.01	0.02	0.01	-0.01	0.02
00	II	-0.04	-0.04	-0.05	-0.05	-0.04	-0.05
00	III	-0.10	-0.10	-0.10	-0.12	-0.10	-0.10
00	IV	-0.06	-0.06	-0.04	-0.05	-0.06	-0.04
01	I	-0.08	-0.07	-0.08	-0.08	-0.08	-0.08
01	II	-0.08	-0.08	-0.10	-0.09	-0.09	-0.10
01	III	-0.01	0.00	0.01	0.00	0.00	0.02
01	IV	-0.06	-0.05	-0.06	-0.06	-0.06	-0.06
02	I	0.00	0.00	-0.01	0.02	0.00	-0.01
02	II	-0.01	-0.01	-0.04	-0.02	-0.01	-0.03
02	III	0.05	0.05	0.05	0.04	0.05	0.05
02	IV	0.02	0.02	0.02	0.03	0.02	0.02

These are derived as $\sum \hat{\beta}_1 '\left(\overline{x}_1^{t-1} - \overline{x}_1^{t}\right) + \ldots + \hat{\beta}_K '\left(\overline{x}_K^{t-1} - \overline{x}_K^{t}\right)$

Table A4.7 Quality correction factors

Year	Qrt	Model 2P	Model 2Full	Model 4P	Model 4Full	Model 2	Model 4
98	I	117.7	118.5	117.6	121.5	119.1	120.2
98	II	113.4	115.0	114.9	117.6	115.2	117.5
98	III	89.1	89.9	88.7	87.8	87.7	86.3
98	IV	79.8	76.7	78.8	73.1	78.0	76.0
99	I	75.5	71.8	74.0	67.0	73.7	71.4
99	II	63.3	58.9	60.6	51.6	61.7	58.4
99	III	61.3	57.0	59.4	49.8	59.7	56.9
99	IV	53.7	50.0	51.5	42.3	52.4	49.4
00	I	53.4	49.7	52.4	42.7	52.1	50.3
00	II	51.2	47.9	49.9	40.8	49.9	47.9
00	III	46.4	43.4	45.0	36.3	45.2	43.2
00	IV	43.8	41.0	43.4	34.6	42.6	41.5
01	I	40.6	38.1	39.9	31.8	39.5	38.2
01	II	37.3	35.2	36.1	29.0	36.2	34.6
01	III	37.1	35.2	36.6	29.0	36.2	35.2
01	IV	35.0	33.6	34.5	27.5	34.2	33.2
02	I	35.0	33.6	34.2	28.0	34.2	32.9
02	II	34.6	33.1	33.0	27.3	33.8	31.8
02	III	36.4	34.7	34.6	28.6	35.6	33.5
02	IV	37.0	35.4	35.3	29.3	36.2	34.1