Operationalising a Hedonic Index in an Official Price Index Program

- Personal Computers in the Swedish Import Price Index

Jörgen Dalén

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
1992:15

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.
Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1992:15. Operationalising a hedonic index in an official price index program – personal computers in the Swedish import price index / Jörgen Dalén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1992-X101OP9215

Operationalising a Hedonic Index in an Official Price Index Program

- *Personal Computers in the Swedish Import Price Index*

Jörgen Dalén

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
1992:15

Från trycket	December 1992
Producent	Statistiska centralbyrån, utvecklingsavdelningen
Ansvarig utgivare	Lars Lyberg
Förfrågningar	Jörgen Dalén, 08/783 44 94

© 1992, Statistiska centralbyrån
ISSN 0283-8680

SCB-Tryck, Örebro 1993 01 Miljövänligt papper

OPERATIONALISING A HEDONIC INDEX IN AN OFFICIAL PRICE INDEX PROGRAM: PERSONAL COMPUTERS IN THE SWEDISH IMPORT PRICE INDEX.

JÖRGEN DALÉN
STATISTICS SWEDEN

Abstract: Questions concerning the operationalisation and practical implementation of hedonic index computations in price index programs of national statistical agencies are discussed with emphasis on the choice of index formula. The example used in the discussion is a hedonic index for personal computers introduced into the Swedish import price index in 1991. Some tentative conclusions from the work are:
1) An adjustment method for dealing with quality changes is proposed as the standard technique.
2) A simple regression model with few variables may be sufficient which is important considering limited resources.
3) Weighting with market share weights should be considered in running the regressions.

Acknowledgement: The details on the Swedish PPI program in Section 4 has been provided by Hans Agrell, who is responsible for this program. I have also benefited from discussions with Daniel Thorburn at the University of Stockholm.

CONTENTS **page**

1. Introduction

In this paper we will discuss some problems in implementing hedonic index computations in a previously existing price index system of a statistical agency. Our example will be personal computers in the Swedish Import Price Index (IMPI).

For some years experiments have been going on with hedonic methods for computer equipment in the Swedish PPI program (covering Home Market, Import and Export Price Indexes). Since Sweden does not produce computers itself the index directly concerned is the IMPI.

In Dalén (1989) some initial experiments were reported in which market surveys from computer magazines were used in the estimation process. Since then experiments have continued and from January 1991 the estimates for personal computers in the IMPI are based on a hedonic technique to be presented in this report.

In Section 2 we give a brief summary of the theoretical background. Section 3 deals with the operationalisation problems with respect to index formulas, functional forms etc. In Section 4 the Swedish PPI program is presented with particular emphasis on the handling of quality changes and the computer indices. In Section 5 we present the hedonic technique for personal computers that we have implemented from 1991. Section 6 presents a number of numerically calculated indices and discusses differences between them.

2.Hedonic functions and hedonic price indices

The theoretical literature on hedonic methods is by now immense. The major theoretical contribution on hedonic functions is Rosen (1974). Triplett (1987,1989) provide additional explanations on the theory and the following is a brief summary.

A *hedonic function* is a relation between prices of varieties or models of a heterogeneous product and the quantities of characteristics contained in them:

$$\Pi = \Pi(\mathbf{x}) + \varepsilon \qquad (1)$$

where Π is the expected price of a variety containing $\mathbf{x}$, a vector of characteristics. These characteristics have a value to the user and are costs for the producer. The theory of hedonic functions depends on the

> "*hedonic hypothesis* that goods are valued for their utility-bearing attributes or *characteristics*. Hedonic prices are defined as the implicit prices of attributes and are revealed

to economic agents from characteristics associated with them." (Rosen,1974, p 34)

In our case the hedonic hypothesis implies that a computer transaction is a tied sale of a bundle of characteristics.

The form of the function $\Pi(.)$ is, except in particular instances, not determined by economic theory but is an empirical matter. The hedonic function is neither determined merely from the demand nor merely from the supply side.

Traditional price index theory deals with the weighting of price changes for baskets of *homogeneous* goods. Hedonic price index theory could perhaps be viewed as extending traditional index theory to the case of *heterogeneous* goods, which could be considered as tied bundles of characteristics with implicit prices $\pi_j = \partial\Pi/\partial x_j$ attached to them.

For computers a variety is a particular machine (or possibly a computer system), which is identified uniquely by its characteristics vector. Characteristics are specifications such as speed, memory size, storage capability etc., which are, in the case of a producer price index, outputs for producers and, usually, inputs into a production process for buyers. The U.S. work on computers is descibed in e.g. Cole et al (1986), Triplett (1989), Sinclair and Catron (1990) and Gordon (1990). In Europe work has chiefly been done in France - see Moreau (1992). EUROSTAT (1991) gives a quick and easy introduction to the hedonic approach.

In a consumer price index a variety is usually a brand and characteristics are properties of a product that are directly valued by consumers. For example in Armknecht and Weyback (1989), a regression equation for women's suits is given, where characteristics are type of brand, material of which the suit is made (wool, silk, polyester, cotton and so on) and style variables.

The Swedish construction price index for single- and multi-family dwellings is also based on a hedonic method - see SOU(1971). In this case a "variety" is a project (loan unit) consisting of a number of houses and apartments. Characteristics are living area, other area variables, equipment standards, region, etc.

3. Operationalising the hedonic theory

Economic price index theory in general is not operational; it does not provide enough specifications to guide actual measurement and computation. This is true for traditional price index construction but even more so for hedonic index construction. In this section we shall discuss more closely the various steps that must be taken before we have an operational index definition in the hedonic case. There are two major technical, statistical choices to be made:

- the choice of index type and
- the choice of functional form for the regression.

Our perspective will be that of official price indices. Triplett(1989) discusses these operationalisations from the perspective of specially designed long-term comparisons.

3.1 Traditional operationalisations

There are two basic forms of operationalisations in statistical index practice. The first one is the choice of index formula: Laspeyres, Paasche, Edgeworth, Fisher etc. The second one is the periodicity of chaining: annual links or less frequent chaining. These choices also affect the operationalization of the hedonic indexes.

Below we will take, as an example, a Laspeyres' two-period index as our starting point for other operationalisations.

3.2 The type of hedonic index

First we must give a more exact specification of the hedonic function. A general way of doing this specification which covers the proposals that have been put forward by various researchers is by one of the following two models:

The time dummy model: $g(\Pi) = h(\mathbf{x}, D) = \alpha + \Sigma_j \beta_j x_j + \gamma D$ (2)

The time specific model: $g(\Pi) = h_t(\mathbf{x}) = \alpha_t + \Sigma_j \beta_{tj} x_{tj}$ (3)

Here $\mathbf{x}$ is the vector of characteristics used in the model and D a time dummy which is equal to 1 if the time period is 1 and 0 if the time period is 0 (we are only concerned with two-period comparisons). g is an invertible transformation: logarithmic in the semi-log or double-log models, a power transformation in the general Box-Cox model or the identity transformation in the linear model. h is a linear function. Π is interpreted as the "true" price given the model but also as the direct function - $g^{-1}h$ - of the characteristics.

We now distinguish five different types of hedonic indexes in an applied context.

The dummy variable index (DV): The index is based on the dummy variable coefficient γ in (2).

The direct model index (DM): We hold the characteristics constant and compare the Π:s from the respective functions at time 0 and 1.

The adjustment index (AD): We make a quality adjustment of either of the two observed prices based on the hedonic function. An adjustment index could be viewed as a formalisation of a common procedure in official price index methodology, called direct quality adjustment.

The imputation index (IM): We keep the actual price in one end of the comparison and impute the model-based price with the same **x** in the other end of the comparison.

The characteristics index (CH): The index is based on the characteristics prices $\pi_j = \partial\Pi/\partial x_j$, the partial derivative of the hedonic function with respect to quality characteristic j. In general the π_j are not constant over the characteristics space but are functions of **x**.

Below, explicit formulas for these types are given for a Laspeyres index and an arbitrary functions g and h. Σ_k is a summation over all models sold at a certain time (0 in this case) and Q_{k0} is the quantity sold of that model. P_{kt} is the actual unadjusted price of model k at time t (having quality characteristics $\mathbf{x}_{kt}$), while as $\pi_{tj}(\mathbf{x})$ is the implicit price of characteristic j evaluated at **x** in the characteristics space. Σ_j is a sum over all quality characteristics. The DV index is of course based on (2) above while the other four are base on (3).

DV:
$$\frac{\sum_k Q_{k0}(g^{-1}[g(P_{k0})+\gamma])}{\sum_k Q_{k0}P_{k0}} \qquad (4)$$

DM:
$$\frac{\sum_k Q_{k0}\Pi_1(\mathbf{x}_{k0})}{\sum_k Q_{k0}\Pi_0(\mathbf{x}_{k0})} \qquad (5)$$

AD:
$$\frac{\sum_k Q_{k0}(g^{-1}[g(P_{k1})+h_1(\mathbf{x}_{k0})-h_1(\mathbf{x}_{k1})])}{\sum_k Q_{k0}P_{k0}} \qquad (6)$$

IM:
$$\frac{\sum_k Q_{k0}\Pi_1(\mathbf{x}_{k0})}{\sum_k Q_{k0}P_{k0}} \qquad (7)$$

CH:
$$\frac{\sum_k Q_{k0}\sum_j \pi_{1j}(\mathbf{x}_{k0})x_{jk0}}{\sum_k Q_{k0}\sum_j \pi_{0j}(\mathbf{x}_{k0})x_{jk0}} \qquad (8)$$

The specification of CH in (8) deserves a few comments. It corresponds to the formulation in Dulberger (1989, equation 8). With this formulation it is different in principle from the others in that it does not aim at measuring price change for the product as a whole but only its characteristics. If we wanted it to measure the whole product, the intercept term should also be included. In the linear case (see 1 below) CH would then be equivalent to DM.

The AD index has a special appeal to index practitioners, since it resembles the usual practice of direct quality adjustment. But it also has a very desirable theoretical property: **An AD index (but not the other four) converges to a standard index formula as quality differences for substitutions approach zero.** It thus leaves less room for errors due to model misspecification and estimation.

3.3 Functional form

Rosen (1974) gives an economic-theoretic argument for the hedonic function to assume a non-linear form. The characteristics prices are not constant since neither the cost of producing nor the demand for purchasing more or less of a characteristic is independent of the cost of or demand for other characteristics. Therefore some kind of multiplicative form for the hedonic function should be chosen. Those that have been mostly discussed and empirically tested are the following:

dl: The double-logarithmic form:

$$\log(\Pi) = \alpha + \Sigma_j \beta_j \log(x_j) \qquad (9)$$

sl: The semi-logarithmic form:

$$\log(\Pi) = \alpha + \Sigma_j \beta_j x_j \qquad (10)$$

bc: The general Box-Cox form

$$\Pi^\lambda = \alpha + \Sigma_j \beta_j \log(x_j) \text{ or } \qquad (11a)$$

$$\Pi^\lambda = \alpha + \Sigma_j \beta_j x_j \qquad (11b)$$

where λ is a constant usually between 0 and 1.

l: For reference we also write down the linear form:

$$\Pi = \alpha + \Sigma_j \beta_j x_j \qquad (12)$$

For the characteristics prices we obtain the following expressions in these cases: $\pi_j = \beta_j \Pi/x_j$ (9*)

$$\pi_j = \beta_j \Pi \qquad (10^*)$$

$$\pi_j = \beta_j \Pi^{1-\lambda}/\lambda x_j \qquad (11a^*)$$

$$\pi_j = \beta_j \Pi^{1-\lambda}/\lambda \qquad (11b^*)$$

$$\pi_j = \beta_j \qquad (12^*)$$

Crossing these five functional forms with the types of indexes in Section 3.2 leads to 25 different indexes altogether. If other basic index formulas than the Laspeyres are considered the possibilities multiply further. There is no need to write them all down explicitly. A few of them are of particular interest, however.

DV combined with a logarithmic (sl or dl) functional form leads to an index equal to exp(γ) although the γ:s are of course different in the two models. This index is what Triplett(1989) refers to by "The Dummy Variable Method" although in his case he includes multiple time periods in the regression equation. But DV could be combined with other functional forms too, giving different but sensible numerical results.

3.5 Weighted or unweighted regression

In an existing index system observations are usually weighted by their market shares and so the question arises as to if regression should also be done based on these weights. Intuitively one wants to give larger weights to more important observations. From the point of view of regression theory weighting with w is optimal (BLUE) if the error variance of the residuals is σ^2/w. In our practical experince there is some reason to believe that there are larger errors associated with models with smaller market shares, which gives some theoretical justification to the use of weighted regression.

3.6 Demarcation of the product group

The demarcation of the product group for which to apply the hedonic regression is another specification problem. For computers there are mainframes, minicomputers, PCs, displays, printers etc. Some computers are sold, packaged into systems, sometimes with software included. This is, however, more common in the final consumer stage than it is for imports.

3.7 Selection of variables

The selection of variables is essentially an area of subject-matter expertise combined with certain statistical principles of regression theory. We will discuss our choice in the computer area in more detail below. A particular problem is whether to include what Gordon (1990) calls make effects, which would imply using dummies for (possibly groups of) companies/outlets.

3.8 Sampling and estimation

We view the procedure of calculating a price index as a process with the following steps: 1) Ideal definition, 2) Operational definition, 3) Population parameter and 4) Estimate based on a sample. See Dalén (1992) for more details on this philosophy. The specifications done above essentially take us only up to step 2. An important problem is that we do not have access to quantities or market share weights for items at the micro (elementary aggregate) level. For example, in the Swedish PPI system we generally have market share weights down to the company level but not for different products within the company. A purposive sample of one or a few products is taken after consulting the company itself. The resulting hedonic index formula after these steps are taken will be somewhat different from those in (4) to (8) above. More below.

4. The computer industry in the Swedish PPI program

4.1 General

The Swedish PPI program consists of five different index series. Three of them are primary:

- The Home Market Price Index (HMPI)
- The Export Price Index (EXPI)
- The Import Price Index (IMPI)

Two index series are derived from the primary ones:

- The Producer Price Index = HMPI + EXPI
- The Domestic Supply Price Index (DSPI) = HMPI + IMPI

All indices are chain indices with annual links from December to the current month. Basically, each annual link is computed as a weighted average of price changes from the item at the enterprise level over intermediate commodities and industry levels to the all-item index

$$\mathbf{I} = \Sigma_{\mathbf{k}}\, \mathbf{w_k I_k}\,.$$

Summation is in practice over several levels and the weights w_k are obtained from production values in industrial statistics or from trade values in foreign trade statistics, in both cases multiplied with the price change up to the reference period for the annual link.

4.2 Treatment of quality changes

Each year there are about 600 substitutions out of a total of about 4500 price observations.That is, about 13% of the observations are, within an index link, affected by smaller or larger problems of comparability. There are at present three common ways to treat substitutions in the PPI program:

1) Overlapping. When a substitution occurs, there is a period of overlap in which prices are collected for both varieties. Price change up to the overlap period is then measured using the old variety and price change from the overlap period forwards is measured with the new variety. If both varieties are present on the market for an extended period of time, then the price difference between them could be interpreted as the consumers' valuation of the quality difference. Overlap pricing accounts for some 35% of all substitutions.

2) Link-to-show-no-change. The two varieties are considered non-

comparable and the new variety is linked in so as to show no price change in the current month. This technique is used for about 50 % of all cases of substitution. There is a rather great potential for error with this method since large price changes could be "linked off" this way.

3) Direct evaluation of the quality difference. This evaluation is based on information given from the producing/importing company. A special case of this is direct comparison, where the quality change is judged to be zero. Direct evaluation is used for about 15% of all substitutions.

The computational form of direct evaluation is in our current index practice

$$I = P_1/(P_0+d), \quad (13a)$$

where d is the value of the quality adjustment. This form could be interpreted as adjusting the base price to the quality level of period 1 and could thus be interpreted as a Paasche index. The corresponding Laspeyres computation would be

$$I = (P_1-d)/P_0 \quad . \quad (13b)$$

The hedonic technique is naturally seen as a formalisation of (13a or b). By keeping the actual price in one end of the comparison it mostly resembles the AD or IM hedonic index types in Section 3.2.

4.3 Computers in the PPI system

Computers have been measured in the Swedish PPI system since the 1970s. Index series that are presented according to SNI (the Swedish standard of industrial classification, very similar to ISIC) are labelled SNI 38251. In this SNI group computers of all sizes are included as well as computer parts. All products are classified according to a statistical number in the Harmonized System and is under its chapter 84.71.

Since Sweden does not produce computers itself, computers are only a part of the IMPI. Price data are collected monthly from importers of computer equipment. The price is in principle supposed to reflect the monthly average of invoiced prices. Value added tax and other taxes are excluded.

Within the IMPI the weight shares of computers are the following:

Of all goods (SNI 1+2+3)	3.7%
Of all manufactured goods (SNI 38)	7.8%
Of all machinery	23.8%

The computer weight itself was distributed in the following way:

PC's	26.1%
Minicomputers and larger	9.7%
Peripheral equipment	30.3%
Computer parts	33.9%

It should be noted that large computers are sometimes imported in parts.

Up to 1990 the method mostly used in cases of substitutions in the computer area was the link-to-show-no-change.

In table 1 below price indices for the groups above in the IMPI are given:

Table 1: Import Price Index (1980=100)

SNI/ year	1+2+3	38	382	38251
1981	111,4	107,6	108,7	109,8
1982	127,3	122,0	123,9	131,5
1983	143,6	140,2	140,6	145,9
1984	151,6	144,8	143,6	150,4
1985	156,3	149,6	149,5	159,3
1986	140,2	157,4	155,1	147,1
1987	143,3	164,4	161,9	138,3
1988	147,9	168,6	166,9	135,0
1989	156,9	172,8	170,0	128,3
1990	162,3	179,2	176,4	122,9
1991	163,4	180,8	176,3	111,0

The rise of computer prices in the first half of the 80's is mainly due to the rising exchange rate of the dollar vs the Swedish Krona.

Price changes within the computer groups were:

	Dec 90-Dec 91	Dec 91-July 92
PC's	-29.6%	-16.7%
Minicomputers and larger	-25.2%	- 9.6%
Peripheral equipment	- 6.2%	- 2.6%
Computer parts	+ 1.3%	- 9.5%

Only for PC's the hedonic method was used. For the other three groups traditional methods for handling quality changes were used with emphasis on the link-to-show-no-change.

5. The hedonic technique for PC's used from 1991

Beginning in 1991, the sample of enterprises importing PCs was increased to 16 and a new form was introduced in which they were asked to provide information on their three most sold varieties with regard to quantity. PCs were defined as a processing unit plus hard disk but without display and keyboard. (In the U.S. - Sinclair and Catron (1990) - they grouped computers by type of processor ("clones") and included larger computers.)

The variables for which information was gathered were

PRIS - price in SKr, in some cases transformed from another currency,
KLFREK - clock frequency in Mhz,
DISKMB - size of hard disk in MB,
MINNE - memory size in MB RAM,
BARBAR - portability (1 if laptop, 0 otherwise) and
ATKTID - access time in milliseconds

Since we failed to obtain information on ATKTID from all companies and since its coefficient in the preliminary regression calculations was not significant, it was not included in the final model.

Compared with Moreau (1992) we excluded type of processor, since we felt that the quality of the processor was fairly well covered by KLFREK and MINNE and since their inclusion in earlier experiments (reported in Dalén, 1989) had created problems with insignificant coefficients and erroneous signs. Instead of using both minimal and maximal RAM we only used maximal RAM in MINNE. We estimated all companies together in one single regression model instead of having one equation for each company. Since displays were excluded there were no variables pertaining to displays. We did not use dummy variables for the different companies/brands, the reason being the relatively small sample we are working with (in 1991 there were 16 brands among 48 observations). This issue will be considered more in the future.

For determining a certain year's regression coefficients price data for December in the previous year are used. In the beginning of 1991 these data comprised 3 different PC models from 16 different enterprises, all together 48 PC varieties. For these models there was, besides the variables in the regression, also the WEIGHT (the same type of weight as elsewhere in the PPI system) that was used in combining the different enterprises in the item group to an aggregate index. WEIGHT was used in the regression so that the weighted residual sum of squares was minimized. The reason for choosing weighted regression was a) that it was considered reasonable to give companies with larger market shares an influence proportionate to their size, b) there were some anomalies in the data for a few of the smaller brands (in some cases they had higher prices for varieties with lower quality for all the measured characteristics) and c) the estimated coefficients got larger t-values and there was less problems with erroneous sign when using weighted regression.

The model formulation chosen was the double-logarithmic:

$\log(PRIS_t) = a_t + \Sigma_k b_{tk}\log(x_{tk}) + b_{tB}x_{tB}$ (14),

where the sum over k is over MINNE, KLFREK and DISKMB+1 and x_{tB} stands for portability.

In our computer case, the double-logarithmic, multiplicative model makes intuitive sense in that the value of an additional amount of a certain characteristic depends on the price level instead of being constant as in a linear model. It gave a somewhat better fit than the semi-log model in 1991 with the same variables. We obtained the following estimates for 1991 (R^2=0.82) and 1992 (R^2=0.83), t-values in brackets:

Table 2	1991		1992	
a	5.429	(14.3)	5.684	(18.7)
b_{MINNE}	0.2337	(2.78)	0.2759	(4.47)
b_{KLFREK}	1.041	(7.47)	0.8537	(7.89)
b_{DISKMB}	0.1863	(4.07)	0.1320	(2.59)
b_{BARBAR}	0.5247	(3.59)	0.4790	(5.01)

These coefficients were in principle used for adjustment whenever a substitution occurred. (In some cases there were discontinuations of an old specification without substitution and then the old index figure was retained for the rest of the year.) The price index is thus an AD index according to the terminology of Section 3.2.

The following index formula was applied for the whole PC item group.

$$\mathbf{I} = \Sigma_k w_k I_k \text{ with } I_k = \frac{\sum_{j \in k} p_{1j}}{\sum_{j \in k} p_{0j} c_j} \qquad (15)$$

where the quality adjustment factor
$c_j = \exp\{\Sigma_k b_{0k}\log(x_{1jk}/x_{0jk}) + b_{0B}(x_{1jB}-x_{0jB})$,
p_{1j} (p_{0j}) is the price of model j at time 1 (0),
x_{1jk} (x_{0jk}) is the value of characteristic k in PC model j at time 1 (0) and b_{0k} is the estimated regression coefficient for characteristic k.

The form of I_k is motivated by the fact that the companies were asked to provide information on their three most sold varieties in quantity terms. Under the precondition that quantities are equal the q:s in the price index formula can be reduced, leading to (15). In practice no substitutions with regard to portability occurred.

This formula is actually a hybrid, in that I_k is a Paasche AD-dl index which is weighted to higher levels with the usual weights relating to an older weight base period. Such hybrid index forms are often forced upon index practioners by practical necessity.

6. Numerical outcomes from alternative indices

In table 3 we present results from a kind of sensitivity analysis of our hedonic index. Based on the same data set of monthly prices and quality characteristics for PC's from December 1990 to December 1991, indices based on some combinations of the above methods are calculated.

The **official hedonic** index of 70.4 is the actual, published outcome of our production system. It is in principle based on an IM, Paasche, weighted regression, double-log model. However, there were a few ad hoc decisions in situations of substitutions, which leads to a different result than the straight-forward ex-post computation giving the figure 69.2.

The **direct comparison** index is based on direct paired comparison between the models in December 1990 and December 1991, i.e. an assumption of 100% comparability without quality adjustment.

The **link-to-show-no-change** index is computed so that the index shows no change for an item in those months where one model is substituted for another. That is, new models are "linked in" with an index of 100.

The **DM** and **AD** indices are based on (5) and (6) with weighting according to (15). In the Paasche cases the adjustment is done in the denominator instead of the numerator.

DV1-DV3 are different variants of the Dummy Variable index. In practical computation it has the form:

$$\frac{g^{-1}(g(\bar{P}_0+\gamma))}{\bar{P}_0}, \quad (16)$$

which reduces to $\exp(\gamma)$ for the dl and sl functional forms and to $\frac{\bar{P}_0+\gamma}{\bar{P}_0}$ for the linear form. $\bar{P}_0$ is the average price of all computers.

DV1 indices are based on a single regression for all months together with 13 dummy variables for the months. DV2 indices, in addition, include enterprise dummies. DV3 are chained indices, a product of 12 two-month DV indices without enterprise dummies.

In comparing all these results a striking feature is that all of them, including the traditional non-hedonic indices, show considerable price decreases - index numbers are from 63.0 to 80.2. The hedonic indices are, as expected, lower than the traditional indices. The weighted indices are generally lower than the unweighted ones, but this is a data effect and is difficult to interpret.

Including enterprise dummies gives a sizeable increase in R^2 as could be seen by comparing DV2 to DV1 in table 4. More importantly, the standard errors of the estimated coefficients decrease markedly. In the future we will therefore seriously consider the inclusion of these in the regression model, although we would not want to use them for quality corrections.

The choice between different functional forms or the choice of weighted vs unweighted regression could not be done on the basis of R^2 figures such as those in table 4. This is mainly because these figures are not comparable when the dependent variable is transformed. But in Diagram 1 we have calculated R^2 for a number of different values of λ in (11a) and (11b) for the 1990 price data with the dependent variable transformed according to Weisberg (1984, equation 6.7) to make R^2 comparable. The optimal value of λ is about 0.15 in (11a) giving R^2=0.8274 as compared to 0.8246 with λ=0 (double-log), which is also the overall optimum with respect to R^2. Our feeling is that it is usually not worthwhile to do this optimization procedure in a production situation but rather to stick to a simple functional form.

7. Conclusions

The hedonic technique for dealing with quality changes in price index series holds great promises. A major reason for this is that its competitors - the traditional techniques of overlapping, link-to-show-no-change etc. are so poor.

Its major difficulty is, however, its costs. Large-scale implementation of hedonic techniques require separate regressions run for many, small item groups, everyone of which has to be dealt with by a combination of subject-matter and statistical expertise. For example Sinclair and Catron (1990) run different regressions for computers with different wordsizes and processors while as Moreau (1992) uses different regression models for different companies. It is also highly desirable to recalculate the coefficients often. Given the limited resources available to statistical agencies, it will take a very long time until hedonic methods are applied on a large scale if complicated methods are used. In the meantime old, inferior methods of quality adjustment will prevail, seriously biasing official index numbers. It therefore seems worthwhile to look for reasonable simplifications in the technical solutions with few variables and simple regression functions and index formulas.

We propose the use of an adjustment type of index, recalculated at regular intervals, preferably annually. We also suggest the use of market share weights in the regressions.

TABLE 3: PRICE INDEX FOR PC'S FROM DECEMBER 1990 TO DECEMBER 1991 ACCORDING TO DIFFERENT METHODS

Official hedonic	70.4
Direct comparison	75.2
Link-to-show-no-change	80.2

	Weighted regression			Unweighted regression		
	Double-log	Semi-log	Linear	Double-log	Semi-log	Linear
Lasp-AD	67.7	67.7	65.9	68.2	70.0	67.7
Paasche-AD	69.2	68.7	63.7	69.3	70.3	63.0
Lasp-DM	68.0	68.0	65.7	77.0	77.4	73.3
Paasche-DM	66.7	67.3	66.2	75.5	77.9	75.2
DV1	65.4	65.9	63.8	69.5	75.4	71.1
DV2	66.4	66.3	64.8	70.4	73.1	70.4
DV3	66.1	73.2				

TABLE 4: R^2 FOR HEDONIC MODELS UNDERLYING PRICE INDICES IN TABLE 3

Official hedonic	0.82
Direct comparison	..
Link-to-show-no-change	..

	Weighted regression			Unweighted regression		
	Double-log	Semi-log	Linear	Double-log	Semi-log	Linear
Lasp-AD	0.83	0.78	0.80	0.74	0.67	0.75
Paasche-AD	0.76	0.77	0.67	0.67	0.66	0.67
DV1	0.76	0.77	0.73	0.72	0.69	0.73
DV2	0.90	0.89	0.85	0.85	0.84	0.81
DV3	0.93*	.	.	0.89*	.	.

* = average over 12 months

8. References

Armknecht, P.A.,Weyback, D. (1989): Adjustments for Quality Change in the U.S. Consumer Price Index. Journal of Official Statistics, 5, 107-123.

Cole, R., Chen, Y.C., Barquin-Stolleman, J.A., Dulberger, E., Helvacian, N., and Hodge, J.H. (1986): Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment. Survey of Current Business, 66(January): 41-50

Dalén, J. (1989): Using Hedonic Regression for Computer Equipment in the Producer Price Index, R&D Report 1989:25, Statistics Sweden.

Dalén, J. (1992): Computing Elementary Aggregates in the Swedish Consumer Price Index, Journal of Official Statistics, 8 129-148.

EUROSTAT (1991): Hedonic Price Indices. Eurostat/D2/ISTI/91/18.

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

Moreau, A., (1992): A Price Index for Microcomputers in France. INSEE, Paris, paper to the seminar "International Comparison of Survey Methodologies", Athens, 30th March-1st April 1992.

Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, 92:34-55.

Sinclair, J., and Catron, B., (1990): An Experimental Price Index for the Computer Industry, Monthly Labor Review, October 1990, 16-24.

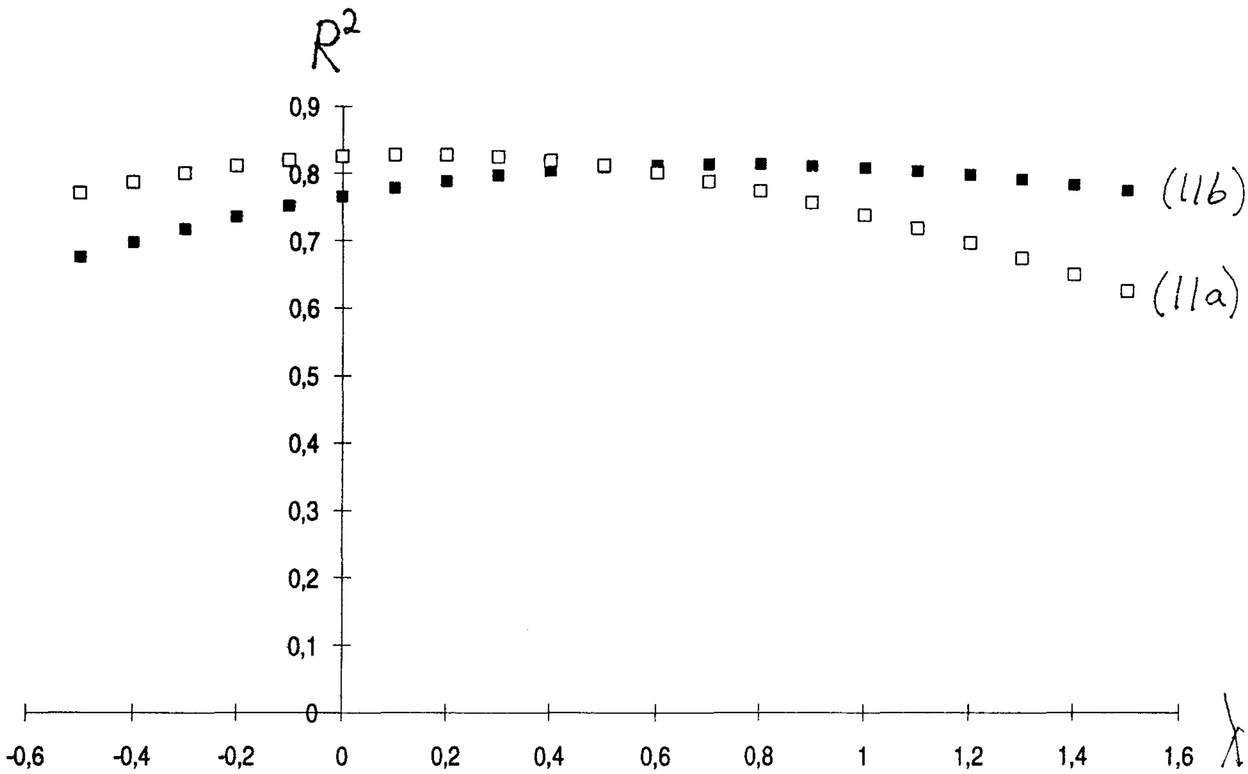
SOU (1971): Byggnadsindex för bostäder, Statens offentliga utredningar 1971:79, Stockholm 1971 (In Swedish).

Triplett, J.E., (1987): Hedonic Functions and Hedonic Indexes. In John Eatwell, Murray Milgate, and Peter Newman, eds., *The New Palgrave: A Dictionary of Economics*. New York:Stockton Press, Volume 2:630-634.

Triplett, J.E., (1989): Price and Technological Change in a Capital Good. In Jorgenson and Landau (editors): Technology and Capital Formation, the MIT Press, Cambridge Massachusetts and London England.

Weisberg, S. (1985): Applied Linear Regression, 2nd ed, Wiley.

DIAGRAM 1:



R^2
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,6
-0,4
-0,2
0
0,2
0,4
0,6
0,8
1
1,2
1,4
1,6
λ
(11b)
(11a)

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie, som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna. I serien ingår även **Abstracts** (sammanfattning av metodrapporter från SCB).

R & D Reports Statistics Sweden are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers. In addition, abstracts are published three times a year (light brown/beige covers).

Reports published during 1992:

1992:1 (grön)	Industrins konkurrenskraft och produktivitet i fokus - en utvärdering av statistiken **(Margareta Ringquist)**
1992:2 (grön)	Automated Coding of Survey Responses: An International Review **(Lars Lyberg and Pat Dean)**
1992:3 (grön)	TABELLER ,... TABELLER ,... TABELLER ,... - Variation och Förnyelse **(Per Nilsson)**
1992:4 (grön)	Basurval vid SCB? Studier av reskostnadseffekter vid övergång till basurval **(Elisabet Berglund)**
1992:5 (beige)	Abstracts I - sammanfattning av metodrapporter från SCB
1992:6 (grön)	Utvärdering av framskrivningsförfarande för UVAV-statistik **(Kerstin Forssén & Bengt Rosén)**
1992:7 (grön)	Cross-Classified Sampling for the Consumer Price Index **(Esbjörn Ohlsson)**
1992:8 (grön)	Bortfallsbarometern nr 7 **(Mats Bergdahl, Pär Brundell, Anders Lindberg, Håkan Lindén, Peter Lundquist, Monica Rennermalm)**
1992:9 (beige)	Abstracts II - sammanfattning av metodrapporter från SCB
1992:10 (gul)	Organizing the Metainformation Systems of a Statistical Office **(Bo Sundgren)**
1992:11 (grön)	CLAN - ett SAS-program för skattningar av medelfel **(Claes Andersson, Lennart Nordberg)**

1992:12 (grön) KVALITETSRAPPORTEN - Utveckling av kvaliteten för SCBs statistikproduktion **(Jan Eklöf, Per Nilsson)**

1992:13 (grön) The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey **(Jan Hörngren)**

1992:14 (beige) Abstracts III - sammanfattning av metodrapporter från SCB

Kvarvarande **beige** och **gröna** exemplar av ovanstående promemorior kan rekvireras från Inga-Lill Pettersson, U/LEDN, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 49 56.

Kvarvarande **gula** exemplar kan rekvireras från Ingvar Andersson, U/LEDN, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 47.

G-090693-93 01016 - 1

CATARINA ELFFORS

A/BF-S