Using Hedonic Regression for Computer Equipment in the Producer Price Index

Jörgen Dalén

SCB

R&D Report
Statistics Sweden
Research-Methods-Development
1989:25

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 1989:25. Using hedonic regression for computer equipment in the Producer Price Index / Jörgen Dalén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1989-X101OP8925

Using Hedonic Regression for Computer Equipment in the Producer Price Index

Jörgen Dalén

SCB

R&D Report
Statistics Sweden
Research-Methods-Development
1989:25

Från trycket December 1989
Producent Statistiska centralbyrån, Utvecklingsavdelningen
Ansvarig utgivare Åke Lönnqvist
Förfrågningar Jörgen Dalén, tel. 08-783 44 94

ISSN 0283-8680
Printed in Sweden
Garnisonstryckeriet, Stockholm 1989

USING HEDONIC REGRESSION FOR COMPUTER EQUIPMENT IN THE PRODUCER PRICE INDEX

Jörgen Dalén
Statistics Sweden, 115 81 Stockholm, Sweden
August 1989

Key words: price index, hedonic regression, computers.

Abstract

A technique for using hedonic regression for computer equipment in current price index calculations is presented and discussed. Data from consumer information in computer journals is used. Results for one product group - personal computers - is presented in some detail.

1. Introduction

The hedonic regression technique for estimating price indexes is used as a means of eliminating quality differences in cases where the old selected item has to be replaced by a new item. With this technique multiple regression is used with price as the dependent variable and (some function of) quality characteristics influencing the price as independent variables. The estimated regression coefficients are then used for imputing a price in either the base or the reference period for a product with the same quality characteristics as in the other period.

There is, by now, an extensive body of literature on the theory of hedonic methods in price indexes in general and on these methods applied to computers in particular. Griliches (1971) book is classical and some important papers concerning computers are Cole et al (1986), Triplett (1986) and Triplett (1989).

However, there is surprisingly little done when it comes to implementing hedonic procedures into official price index calculations at national statistical agencies. In Statistics Sweden, we have done some work in order to develop practicable procedures for measuring price changes for computer equipment in the Producer Price Index. We have also taken a preliminary decision to implement these procedures for one category of equipment - personal computers - in the index calculations for 1990. In this paper we report on our work so far.

2. The Producer Price Index

The Swedish Producer Price Index is computed "from bottom to top" as a weighted average of indexes for different cells, where a cell on the lowest level typically consists of a certain item in a certain enterprise in a certain industry. There is a time lag of about three years between the year of weight calculations and the year of price measurement.

For an index cell an index number is computed with last year's December price as the base price. If there is a change of model during the year the models are usually declared non-comparable implying no price change between them. For computers, where a model change is usually accompanied by a quality improvement, this practice leads to a large upward bias in the index estimates.

3. Hedonic price functions

The hedonic price index theory regards the price as a function of a set of characteristics of a product:

$P = h_t(\mathbf{x})$, where $\mathbf{x}$ is a vector of quality characteristics of the product and t denotes a certain period or point of time.

The choice of functional form is generally regarded as an empirical matter and not one of economic theory - see i.e. the discussion in Triplett (1989). Several forms have been tried in the literature of hedonic price indexes. The most common seem to be the linear, the semilog and the double-log forms which look as follows:

LINEAR: $$P = \alpha + \Sigma\, \beta_j x_j + \varepsilon \qquad (1)$$

SEMILOG: $$\log(P) = \alpha + \Sigma\, \beta_j x_j + \varepsilon \qquad (2)$$

DOUBLE-LOG: $$\log(P) = \alpha + \Sigma\, \beta_j \log(x_j) + \varepsilon \qquad (3)$$

Summation is over the quality characteristics j. For the linear form the regression coefficients β_j may be interpreted as the implicit prices per unit of characteristic j in a certain time period. This implies that for the linear form a unit change in one of the quality characteristics always changes the price with the same amount. In the semilogaritmic form a unit change in one of the quality characteristics always leads to the same proportional price change. Finally, in the double-log form a proportional change in a quality characteristic always leads the same proportional change in the price.

4. Hedonic price indexes

In a hedonic index, the prices should be compared between two time periods for the same values of the quality characteristics. In a Laspeyres index these are held constant at their base period (0) values and in a Paasche index at their reference period (1) values.

The procedure that we consider amounts to "adjusting" one of the prices in the comparison for the change in quality. This adjustment is based on the estimated coefficients in the hedonic function.

We denote the adjusted prices in the following way:

$P_0' = h_0(x_1)$, the adjusted base price or

$P_1' = h_1(x_0)$, the adjusted reference price,

where the indexation 0 or 1 refers to the respective time period. We observe that P_1' is the hedonic function at time period 1 evaluated for the quality characteristics of the computer variant of time period 0.

The corresponding price indexes are:

$$I^1_{01} = \frac{h_1(x_0)}{h_0(x_0)} = \frac{P_1'}{P_0} \quad \text{and}$$

$$I^2_{01} = \frac{h_1(x_1)}{h_0(x_1)} = \frac{P_1}{P_0'} \ .$$

The first index requires estimation of a new hedonic function for each new month in the reference period, which is not practical. The second index, on the other hand, only needs one hedonic function for each year, calculated for the base month, december. For practical reasons we therefore choose the second form of the index.

We also intend to calculate the hedonic indexes on the elementary item level, which means that P_1 and P_0 denote actual observed prices for a certain item in a certain enterprise.

For the three functional forms considered we obtain the following expressions for the adjusted base price and index, respectively:

LINEAR:

$$P_0' = \alpha_0 + \Sigma\ \beta_{0j}x_{1j} + \varepsilon = P_0 + \Sigma\ \beta_{0j}(x_{1j}-x_{0j}) \quad \text{and}$$

$$I^2_{01} = \frac{P_1}{P_0 + \Sigma\beta_{0j}(x_{1j}-x_{0j})}$$

SEMILOG:

$$\log(P_0') = \alpha + \Sigma\ \beta_{0j}x_{1j} + \varepsilon = \log(P_0) + \Sigma\ \beta_{0j}(x_{1j}-x_{0j}) \quad \text{and}$$

$$I_{01}^{2} = \frac{P_1}{P_0 \exp\{\Sigma \beta_{0j}(x_{1j}-x_{0j})\}}$$

<u>DOUBLE-LOG:</u>

$$\log(P_0') = \alpha + \Sigma\beta_{0j}\log x_{1j} + \varepsilon = \log(P_0) + \Sigma\beta_{0j}\log(x_{1j}/x_{0j})$$

$$I_{01}^{2} = \frac{P_1}{P_0 \exp\{\Sigma \beta_{0j}\log(x_{1j}/x_{0j})\}}$$

We observe that when there is no quality change ($x_{1j}=x_{0j}$ for all j) $I_{01} = P_1/P_0$ for all the functional forms and so the hedonic indexes reduce to the simple price index.

5. Estimating the hedonic functions

As a first step towards analyzing the potentials of the hedonic technique and for comparing the different functional forms we used market surveys from Swedish computer journals during 1988. These consisted of data collected by the editorial staffs of these journals for specific computer models. For each model a certain supplier gave information on price and some specified quality characteristics. We studied four types of computer equipment: personal computers, matrix printers, laser printers and displays. Here we present our results concerning personal computers for the preferred regression model in the following table:

Table 1: Estimated models for personal computers

Variable	Functional form					
	Linear		Semilog		Double-log	
	b	t-value	b	t-value	b	t-value
Intercept	1073	0.16	9.1	35.3	6.0	5.9
Memory size	5.6	1.50	.00019	1.37	.305	2.07
Clock frequency	1560	3.30	.062	3.47	.478	2.20
Size of fixed disk	202	5.75	.0038	2.87	.435	4.13
Access time	-101	-1.46	-.0032	-1.21	-.176	-1.60
R^2	0.78		0.69		0.76	
"Relative likelihood"	-538		-533		-527	

The selected regression model was chosen on grounds of R^2-value, significance of coefficient estimates and the reasonability of these estimates from a subject-matter point of view.

The "relative likelihood" in the table refers to the likelihood procedure proposed by Box and Cox (1964) for comparing models based on transformations of the dependent variable. The version of this procedure used here is from Weisberg (1980). The model with the largest "relative likelihood" is the double-log model and is therefore the best in this sense.

6. Examples using the estimated coefficients

Here we present three examples of index calculations at the elementary level using the three functional forms. The three examples are all based on actual model changes for personal computers in the Swedish PPI in 1989. In table 2 below we present the examples. It is important to note that the present practice usually declares the models non-comparable, which implies a price index of 100.

Table 2: Examples of index calculations for model changes

	Example 1		Example 2		Example 3	
	Old model	New model	Old model	New model	Old model	New model
Price, Skr	28099	26597	22703	17620	5353	8281
Memory size	192	256	256	640	No change	
Clock frequency	No change		4.77	8	4.77	8
Size of fixed disk	No change		No change		No change	
Access time	No change		65	40	65	40
Linear						
Adjusted base price	28460		32441		12927	
Price index	93.5		54.3		64.1	
Confidence interval	91.9-95.0		47.9-62.8		48.6-93.8	
Semilog						
Adjusted base price	28450		32343		7078	
Price index	93.5		54.5		117.0	
Confidence interval	91.8-95.2		46.2-64.3		100.2-136.6	
Double-log						
Adjusted base price	30678		41889		7467	
Price index	86.7		42.1		110.9	
Confidence interval	79.8-94.2		31.3-56.6		88.5-139.0	

We note the large differences between the three index estimates for example 3. In this example the prices are unusually low which leads to large differences between the index estimates. The linear index is intuitively worse in this situation, since it gives the quality characteristics the same large value for this cheap machine, whose low price probably reflects a lower service degree and reliability in general. In this situation it is natural also to value the technical quality characteristics lower.

7. The quality of a hedonic index

The quality of a hedonic index estimate is a complicated issue with many aspects. Here we only discuss the quality of the elementary index calculated as above. The only stochastic elements in the price index formulas are the regression coefficients. (It is reasonable to treat the observed prices P_0 and P_1 as fixed constants, if we consider an index based on identical models as having zero error. Alternatively, we could say that we condition on the outcomes P_0 and P_1.) The variance of the linear part of the index estimates may thus be estimated using the variance-covariance matrix of the regression. Confidence interval endpoints can be calculated, which are then transformed to refer to the index estimate.

More formally, for the semilog index we take the following steps:

$$I_{01} = \frac{P_1}{P_0 \exp(B)}, \text{ where } B=\sum_j b_{0j}(x_{1j}-x_{0j})$$

$$\text{Var}(B) = \sum_{ij}\sum (x_{1i}-x_{0i})(x_{1j}-x_{0j})\text{Cov}(b_{0i},b_{0j})$$

$$\underline{B} = B - 1.96\{\text{Var}(B)\}^{\frac{1}{2}} \quad \text{and} \quad \bar{B} = B + 1.96\{\text{Var}(B)\}^{\frac{1}{2}}$$

$$\underline{I}_{01} = \frac{P_1}{P_0 \exp(\bar{B})} \quad \text{and}$$

$$\bar{I}_{01} = \frac{P_1}{P_0 \exp(\underline{B})}$$

The last two expressions we interpret as the lower and upper bounds of an approximatively 95% confidence interval for the semilog index estimate. The procedure for the linear and the double-log indexes is completely analogous. The results appear on the confidence interval rows in table 2. We notice that the confidence interval in most cases do not contain the value 100 implied by the present procedure.

8. Summary

Our experiences in working with hedonic functions and price indexes for computers at Statistics Sweden so far lead us to the following preliminary conclusions.

1) The hedonic technique is promising in this area. This is due partly to the fact that computers have a few major technical variables which are relatively good indicators of the quality of the product. But the major reason is in our opinion that the alternative is so poor. The present practice of declaring new models to be non-comparable with the old ones is believed to have a huge upward bias on the index estimates, since it completely disregards the rapid quality improvements for computer products.

2) The regressions should be done annually, using the base price as the dependent variable.

3) There must be a large enough sample of each item (for example personal computers). In Sweden we will try to use at least 20 observations for each item. We are not yet sure if this is enough. Resource restrictions may call for a corresponding reduction of the number of representative items in an item group.

4) The independent variables should be selected according to several criteria. One is their accessibility: the dealers must be able to provide the data. Another is their stability over time: the regression model must be relevant for a certain time period, in the case of the Swedish PPI one year. Thirdly the estimated regression coefficients should be logically reasonable and statistically significant.

5) Our results so far lead us to prefer the double-log form before the other functional forms. This preference is not strong but the linear form has some undesirable properties.

9 References

Box, G.E.P. and D.R. Cox (1964). "An analysis of transformations (with discussion).", J. R. Statist. Soc. Ser. B, 26, 211-246.

Cole, Rosanne, Chen, Y. C., Barquin-Stolleman, Joan A., Dulberger, Ellen, Helvacian, Nurhan and Hodge, James H. (1986): Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment. Survey of Current Business 66, 41-50.

Griliches, Zvi, ed. (1971): Price Indexes and Quality Change: Studies in New Methods of Measurement. Cambridge, Massachusetts: Harvard University Press.

Triplett, Jack E. (1986): The Economic Interpretation of Hedonic Methods. Suvey of Current Business 66, 36-40.

Triplett, Jack E. (1989): Price and Technological Change in a Capital Good: a Survey of Research on Computers. In Technology and Capital Formation, edited by Dale W. Jorgenson and Ralph Landau, MIT Press.

Weisberg, S. (1980): Applied Linear Regression, Wiley.

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 earlier during 1989 are:

1989:1 (grön)	Går det att mäta produktivitetsutvecklingen för SCB? **(Rune Sandström)**
1989:2 (grön)	Slutrapporter från U-avdelningens översyn av HINK och KPI **(flera författare)**
1989:3 (grön)	A Cohort Model for Analyzing and Projecting Fertility by Birth Order **(Sten Martinelle)**
1989:4 (beige)	**Abstracts I** - Sammanfattningar av metodrapporter från SCB
1989:5 (gul)	On the use of Semantic Models for specifying Information Needs **(Erik Malmborg)**
1989:6 (grön)	On Testing for Symmetry in Business Cycles **(Anders Westlund and Sven Öhlén)**
1989:7 (grön)	Design and quality of the Swedish Family Expenditure Survey **(Håkan L Lindström, Hans Lindkvist and Hans Näsholm)**
1989:8 (grön)	Om utnyttjande av urvalsdesignen vid regressionsanalys av surveydata **(Lennart Nordberg)**
1989:9 (grön)	Variations in the Age-Pattern of Fertility in Sweden Around 1986 **(Michael Hartmann)**
1989:10 (grön)	An Application of Generalized Precision Functions in the 1985 Swedish Family Expenditure Survey **(Håkan L Lindström and Peter Lundquist)**
1989:11 (gul)	Statistics production in the 90's - decentralization without chaos **(Bo Sundgren)**
1989:12 (grön)	Översyn av urvalen i de objektiva skördeuppskattningarna **(Erling Andersson)**
1989:13 (gul)	ADBs roll i statistiken **(Bo Sundgren)**

Var god vänd!

1989:14 (grön) Krisgruppsarbetet och räknarexperimentet i HUT 88 **(Håkan L. Lindström)**

1989:15 (grön) On evaluation of surveys with samples from the revised Zimbabwe master sample frame **(Bengt Rosén)**

1989:16 (grön) Bortfallsbarometern nr 4 **(Sonia Ekman, Tomas Garås, Hans Pettersson, Monica Rennermalm)**

1989:17 (beige) **Abstracts** II - sammanfattning av metodrapporter från SCB

1989:18 (gul) Conceptual modelling as an instrument for formal specification of statistical information systems **(Bo Sundgren)**

1989:19 (grön) Rapport från ett besök vid U.S. Bureau of the Census **(Gösta Forsman)**

1989:20 (gul) Regler och rekommendationer vid konstruktion av användargränssnitt **(Hans Irebäck)**

1989:21 (beige) **Abstracts** III - sammanfattning av metodrapporter från SCB

1989:22 (grön) The Design of Pilot Studies: A Short Review **(Lars Lyberg, Pat Dean)**

1989:23 (grön) Om framskrivning av statistik avseende varutransporter med lastbil ("UVAV"-statistiken") **(Folke Carlsson, Kerstin Forssén, Bengt Rosén)**

1989:24 (grön) Behovsanalys och statistiska beslutsregler **(Håkan Zetterström)**

Kvarvarande **beige** och **gröna** exemplar av ovanstående promemorior kan rekvireras från Elisabet Klingberg, U/STM, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 78.

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