Institut National de la Statistique et des Études Économiques (INSEE)
French National Institute of Statistics and Economic Studies

On the use of hedonic methods in the CPI:
an application to consumer durables and apparel

Jérôme Bascher and Thierry Lacroix

**Paper for the 4th International Conference of the Ottawa Group
Washington, April 1998**

INSEE
Direction des Statistiques Démographiques et Sociales
(Directorate for Demographic and Social Statistics)
Division Prix à la Consommation
(Consumer Prices Division)

Postal address:
Timbre F320
18 boulevard Adolphe Pinard
75675 Paris Cedex 14
France
E-mail: Jerome.Bascher@dg75-F320.insee.atlas.fr
Thierry.Lacroix@dg75-F320.insee.atlas.fr

Abstract

There are several methods for handling the quality effect in consumer price indexes (CPIs), and each method has practical and theoretical arguments to support it. The best approaches seem to be the hedonic methods, despite their high cost and the difficulty of implementing them for the massive processing operations involved in the monthly calculation of CPIs. However, the French experience suggests the approach is worth pursuing especially for products where the quality effect is both perceptible and measurable—and in particular when there are many replacements.

Introduction

Among the enduring problems that affect Laspeyres price indexes, one of the most difficult is basket constancy. The truth is that today's products—whether goods or services—undergo such intensive renewal in so many markets that it is unrealistic to claim that identical products are monitored over the entire lifetime of a calculation base. To avoid excessive bias, statisticians must minimize the number of replacements and estimate the quality effect. But they also need to preserve a representative sample of total consumption: this is an incentive to expand the number of replacements while choosing the right time to introduce new products, when possible. The French CPI seeks to take all these contradictory aspects into account.

In section 1, we describe the techniques used in the French CPI to measure these effects of new product introductions and replacements. Our discussion emphasizes the impact of these events at the time of their occurrence. In section 2, we argue that hedonic methods appear to be the best approach for dealing with quality effects, especially—in INSEE's experience—in consumer durables and apparel.

I. New products - Product replacements

To measure "pure" price changes, France—like other countries—has opted for a chained Laspeyres index, in which the numerator and denominator are based on a fixed weighting of prices by quantity.[1] Product introductions and replacements are potential sources of bias in price-index weightings. Regardless of how we define the problem—"volume-price decomposition" or "quality effect"—we will need to measure it before treating it, if we want to avoid a major bias in the CPIs (Boskin *et al.* 1996).

1.1. Measuring the phenomenon

The CPI basket must provide the best possible reflection of actual household consumption. Accordingly, some products have to be introduced over time (CD-ROMs, mobile telephones, etc.) while others are removed (vinyl records, for example). Because the weightings are fixed, it is impossible to add or remove these products in the course of the year. Thanks to the annually

---

[1]The weighting is fixed for a year since the French index is chained annually.

chained Laspeyres index, France updates its basket with a new set of weights every year, in December. At the item-grouping[2] level, the national accounts are generally detailed enough to provide the necessary estimates. At the item[3] level, the information sources are more irregular. New data from an aggregated scanner data panel will soon be used in France to cover mass-merchandised products. For the time being, the trade press is a valuable source. Suitable criteria are needed, however, for assessing whether a product has become significant in household consumption. Eurostat recommends the inclusion of any new product representing at least 0.1 per cent of total consumption. For obsolescent products, the main reason why CPI statisticians remove them is that they are hard to find on store shelves, since they are no longer representative of household consumption.

The annual chaining actually comprises four operations: a new set of items, an updated list of sales outlets, new weights, and—in consequence—a new sample of varieties.[4] This new sample is directly linked to the quality effect, as the following example shows. If, all other things being equal, the old sample of personal computers had no model with CD-ROM drive and the new sample consists only of PCs with CD-ROM drives, the measured universe will be biased but the bias will not appear in the price index. If this product replacement within the same family occurs at the linking stage, then the quality effect will be invariably canceled. In France, the cancellation effect is mild. About 10% of the varieties change at the chaining stage. Moreover, these changes are not one-way quality increases or decreases (Lequiller 1997). In any event, there is no explicit tracking of the effect of replacements on CPI movements.

Most of the changes in the sample take place not at the chaining stage but during the year. More than one-third of the sample is replaced in the course of the year: some categories like consumer durables or apparel have a replacement rate comprised between 70% and 90%. For these two categories, the high replacement rate is due to different reasons. In durables, the technology race makes products obsolete sometimes within just a few months. Apparel replacements are dictated by fashions and the switch between summer and winter collections. Furthermore, the products tracked are very specific and their description is based on many criteria that price collectors are required to observe on their visits. Indeed, in many cases, a change in even a single characteristic is deemed to generate a replacement. For example, if the product tracked was "a liter carton of brand X unskimmed milk" in one month and "a liter bottle of brand X unskimmed milk" in the next, then a replacement has to take place. This explains the high percentage of replacements observed. However, depending on the estimated kind of replacement, the statistical treatment differs.

---

[2]In French, *poste*, of which the French CPI contains 265.

[3]In French, *variété*, of which the French CPI contains about 1,000.

[4]In French, *série*.

1.2. A choice of treatments

1.2.1. Systematic methods and their offshoots

There are several different methods available, each to be used in well-defined cases. The items composing the index are divided into four broad and distinct groups: homogeneous items, heterogeneous items, tariffs, and fresh products. The last two groups are treated with specific methods not discussed here. The first two cover most of the quotations collected locally at sales outlets, or about 130,000 varieties accounting for 67% of the total CPI weighting.

- For homogeneous items (31% of the total CPI weighting)

All products composing a homogeneous item are regarded as similar. Consequently, the replacement of a removed variety by a new one allows an immediate price comparison between the two varieties. There is only one type of treatment, but it can be split in two: the product equivalent and the packaging-unit equivalent.
Example: hazelnut chocolate bar in paper wrapper.

Sub-case 1: Product equivalent

If of these prestige-brand bars is replaced by another bar from another prestige brand, the change is an equivalent. The new equivalent base price is calculated as follows:

New base price (P_B^N)= old base price (P_B^A)

Sub-case 2: Packaging-unit equivalent

This method is identical to the previous one, but applied to the packaging unit. The bar hitherto sold in a 200-gram package is now sold in a 250-gram package, with an identical proportion of hazelnuts. The "privileged variable" (see below, case 2) is calculated as follows:

New unit base price = old unit base price or, for the example chosen,
$P_B^N = 1.25 * P_B^A$

Interestingly, in this case, we are looking not for a product but for a quantity of product in a well-defined packaging unit.

- For heterogeneous items (36% of the total CPI weighting)

Heterogeneous items comprise products of different sorts, whose prices are therefore hard to compare. Several methods are available, depending on the requirement.

Case 1: "Direct Comparison" (Équivalent: EQ)

The removed variety is replaced by a very similar product with identical technical characteristics. The change is calculated on an equivalent basis using the same procedure as described for sub-case 1 above:

EQ: New base price (P_B^N)= old base price (P_B^A)

Despite the diversity of products, this method is employed for 30% of the replacements occuring in the heterogeneous items.

Case 2: "Privileged Variable" (Variable Privilégiée: VP)

This method is an equivalent applied to the packaging unit as described previously (cf. sub-case 2, homogeneous items). The calculation requires a quantitative variable closely proportional to the price, called a "privileged variable":

VP: New unit base price = old unit base price

Case 3: "Dissimilar" (Dissemblable: DI)

The removed variety is replaced by a slightly different product (for example, same type of coat but different fabric content and gilt buttons instead of zipper fastening). The change in the variety's price index is regarded as null since the products, being of different quality, are regarded as non-comparable. In consequence, the entire price change is attributed to the quality change.

DI: $P_B^N = \dfrac{P_m^N}{I_s(m-1)}$ where $I_s(m-1)$ is the index of the variety replaced in month m-1[5]

and P_B^N the observed price of the replacement variety in month m.

NB: This method relies on two assumptions: that there is no inflation[6] when the product is replaced, and that there is no price change between months m and m-1. The method eliminates the problem of measuring the price movement at the time of replacement and creates a downward bias when the indexes are positive. In keeping with the regulations adopted for the European Harmonized Index of Consumer Prices (HICP), INSEE is trying to eliminate this bias. We expect to have practically achieved this in 1998.

---

[5]m is always considered the replacement month, m-1 being the last month in which a price was collected in "normal" conditions.
[6]Inflation is defined here in the broad sense, i.e., an upward or downward movement in prices.

Case 4: "Adjusted Dissimilar" (Dissemblable Corrigé: DC) and "National Dissimilar" (Dissemblable National: DN)

The situation is identical to case 3. However, in the same region, for the same item, at least four "real" quotations have been collected (three in small regions). The price movement of the replacement variety is deemed identical to that of the (item x region) aggregate between months m-1 and m.

$$\text{DC: } P_B^N = \frac{P_m^N}{I_s(m-1)} * \frac{I_{va}(m-1)}{I_{va}(m)} \text{ where } I_{va}(m) \text{ is the index for the (item x region) in the}$$

month m and $I_s(m-1)$ is the index of the replaced variety in month m-1.

However, there are not always enough varieties for the estimation. In 1998, INSEE is therefore setting up a complementary procedure called the "national dissimilar." To estimate the "national dissimilar" price change between months m-1 and m, we take the change in the national item, under constraints that, by definition, are more easily complied with at an aggregate level. This procedure is used only in those residual cases for which a DC estimation was impossible.[7]

$$\text{DN: } P_B^N = \frac{P_m^N}{I_s(m-1)} * \frac{I_v(m-1)}{I_v(m)} \text{ where } I_v(m) \text{ is the index of the national variety in}$$

month m.

Case 5: "Discretionary" (Au Choix: AC)

As its name indicates, this method permits linking at the operator's discretion. Supplementary information is used to calculate price changes with the closest possible fit. The method is used only in exceptional cases—most often to directly introduce the variety's new base price when known, in order to obtain an effective display of the "true" price change over the year.

Example: bestselling books. The book's base price is estimated from a formula linking the price to the number of pages and the type of sales outlet. The price ratio is equal to the observed price divided by the base price calculated as above. In this specific example, the method actually resembles the econometric method described below[8].

---

[7]On average, a sub-item comprises 140 quotations, and between 120 and 130 actual observations are available for a monthly DN estimation, whereas there are only six or seven observations per region that can be used for the DC. Moreover, the sample size varies considerably from one item to another. Quotations are more abundant when the scatter of price changes is greater (Guglielmetti and Ardilly 1993). In particular, for "small items," there are often too few varieties to allow the use of the DC—hence the value of the DN.

[8] this way of use is not completely right since all the informations about the past evolution of the replaced product is lost. A change has just been realized in 1998 in order to treat quality adjustments as it is explained in §1.2.4 (hedonic models).

Case 6: Item-Level Linking (Raccord au Niveau Variété: VA)

Here, the replacement introduces the new product at the "average" level of the corresponding item.

VA: $P_B^N = \dfrac{P_m^N}{I_v(m-1)}$ where $I_v(m-1)$ is the index of the corresponding item in month m-1.

This method was used for durable goods, whose prices were, until 1997, typically collected every quarter. Its advantage was that it allowed statisticians to work with the small number of available monthly observations. However, there is an inherent bias risk, as the method does not keep a record of the variety's past history between the base month and the replacement month (we substitute the average change in the item price). As with the DI method, there is also a downward bias when the indexes are positive. The complete conversion of CPI varieties to a monthly basis and the introduction of the "National Dissimilar" (DN) will allow the VA method to be eliminated in 1998.

1.2.2. Respective importance in the CPI of the methods described above

In 1997, the number of forced replacements due to the removal of the initial product reached 46.1% of the varieties tracked in the sample (excluding tariffs and fresh products). The highest rates were recorded in « clothing and footwear » and « durable goods ». Taken together, these groupings, which make up 23.9% of the sample varieties, account for 53.8% of product replacements.

Figure 1 - Forced replacements in the CPI in 1997: number of cases and respective importance of the different methods

	All products	including:	
		Clothing, Footwear	*Durable Goods*
Number of forced replacements	58,801	*23,601*	*8,006*
Rate of replacements [1] **(%)**	**46.1**	***103.4***	***104.3***
Ratio for homogeneous items varieties [2]**:**			
- in the sample (%)	41.6	*12.4*	*0*
- in the replacements (%)	17.8	*7.9*	*0*
Methods used for replacements [part (%) according to the number of replacements]			
DI	12.4	*21.0*	*4.5*
DC, DN	37.4	*35.6*	*44.0*
VA	6.6	*0.0*	*48.3*
« Non direct comparison »	**56.4**	***56.6***	***96.8***
« direct comparison» (EQ)	**42.4**	***42.8***	***1.9***
« Explicit evaluation of quality effect » (VP, AC, EC)	**1.2**	***0.6***	***1.3***
All replacements	**100.0**	***100.0***	***100.0***

[1] Rate of replacement = number of forced replacements in the course of the year / number of sample varieties.
[2] Part (%) of varieties tracked in homogeneous items.

The methods used to treat replacements fall into three categories of uneven importance:

- « Non-Direct comparison » (DI, DC, DN, VA), used for 56.4% of the replacements (figure 1). This type of replacements prevails in « clothing and footwear », in« furnishings, household equipment, and routine maintenance of the house », in « recreation and culture », and in « hotels, cafés and restaurants ».A high proportion of these replacements still use a procedure that is based on a questionable or imperfect principle (DI or DC, respectively). The situation will improve in 1998, since the conversion of quarterly items to a monthly basis and the implementation of the "National Dissimilar" will practically eliminate DI and VA, to be replaced by DC and DN.

- "Direct Comparison" (EQ): 42.4% of all replacements are performed with this method, and more than 50% in food, "housing, water, electricity, gas, electricity and other fuels," and transportation, that is to say categories where a major part of the items are defined as « homogeneous ».

- "Explicit Evaluation of Quality Effect" (VP, AC, EC[9]): 1.2% of replacements are made with the use of a method that seeks in one way or another to determine an explicit measure of the difference in quality between the replaced product and the replacement product. This low

---

[9]EC: econometric or hedonic method, the subject of section 2 of this study.

percentage actually underestimates reality, because it does not take into account the "packaging-unit equivalent" replacements in homogeneous items.[10]

1.2.3. The limits of systematic methods: the example of vacuum cleaners

In the field, all the methods used require a perfect description of the tracked product. This explains why the collecting forms often include many headings on the technical characteristics of the product observed—especially for durables and apparel. The reason is that the necessary changes are identified by comparing characteristics, not prices. The massive use of dichotomous methods for measuring the quality effect leads to all-or-nothing choices that are bound to influence the values of the indexes calculated. Let us take the example of vacuum cleaners to illustrate the impact of the replacement method chosen on the item index.

Figure 2 - Change in vacuum-cleaner index between December 1995 and October 1996 determined with different methods of measuring quality effect in replacements
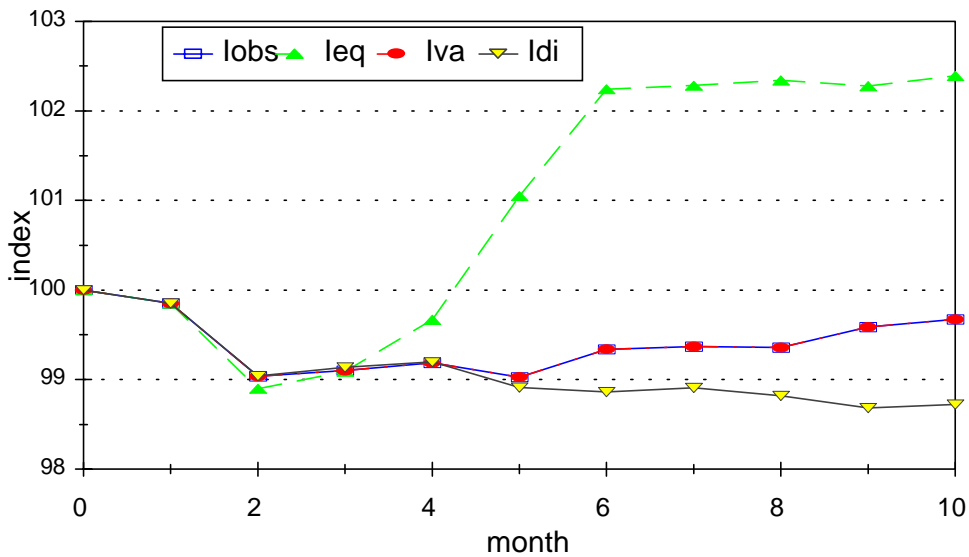
Iobs = observed index
Ieq = EQ for all replacements index
Iva = VA for all replacements index
Is = DI for all replacements index

In figure 2, we compare the movement in the vacuum-cleaner price index in the first ten months of 1996—in which the main method for treating replacements was VA—with the index change that would have been obtained by using the extreme methods, EQ ("Direct" comparison) and DI ("Dissimilar" comparison). The number of replacements in the vacuum-cleaner item is fairly high. There were 240 in the ten months examined, out of 248 quarterly quotations, one-third of

---

[10]Cars, for which specific methods are used (see §1.2.4) are also excluded from the analysis, since they are treated as tariffs, whose quotations are collected on a centralized basis. Rents are not taken into account either: they are calculated from a quarterly survey of 8,500 households using an explicit method of quality-effect measurement by sample stratification. Cars and rents account for a combined 9.1% of the total CPI weighting.

these being collected each month. The final divergences are therefore very wide: the CPI shows a slight decline, but the EQ method would have led to a rise of more than 2% and the DI method would have shown a drop of more than 1%. The main reason for these gaps is the difference between movements in prices of "permanent" products (which tend to fall) and of replaced-product prices (new models being typically more expensive than old ones).

We are forced to admit, therefore, that the actual change in the vacuum-cleaner price index is subject to significant uncertainty.

1.2.4. Hedonic methods

A more satisfactory treatment of product replacements calls for an explicit assessment of the quality effect that goes beyond the simplistic "all-or-nothing" technique. The most suitable approach is the use of econometric methods, which make a rational separation between quality effect and price effect, without prior assumptions. The price of a good is decomposed by product characteristic. Each characteristic is assigned a value in terms of price or price share. In France, the initial approach to this quality-adjustment method has been named the "Privileged Variable" (see §1.2.1, case 2 above). It actually consists in a very simple regression that may be regarded as hedonic. In fact, though, this type of decomposition is unsatisfactory, owing to the complexity of the products involved. It seems obvious, for example, that the price of a TV set is not wholly explained by its screen size.

Our goal here is to calculate a product price at a given date based on its characteristics. Econometrics provides the means to estimate the real or marginal costs of each characteristic. According to many experts, the sought-for expression may be written as follows:

$P_0^A, P_0^N:$ base price of old product (A) and new product (N);

$P_t, P_t^N, P_t^A:$ product price in month t, new and old, if an ambiguity exists;

$I_{t/0}^V, I_{t/0}^S:$ index of item, variety in month t, with month 0 = base;

$X_A, X_N:$ quality of the old and new products;

$f^0(X):$ estimated price of quality X at time 0;

m: replacement month;

i: product tag in the sample;

> *note:* $P_m^N, P_{m-1}^A, P_0^A, X_A, X_N$ are observed, $\hat{P}_{m-1}^N, \hat{P}_0^N$ are estimated.

$\hat{P}_m = P_m^N - \hat{P}_{m-1}^N:$ estimated "pure-price" difference between *m-1* and *m*

$\hat{d}_{m-1} = \hat{P}_{m-1}^N - P_{m-1}^A:$ value of estimated quality difference between new and old product in month *m-1;*

$\hat{d}_0 = \hat{P}_0^N - P_0^A:$ value of estimated base-price difference: read as a quality-difference price in 0.

Figure 3 - Replacement diagram
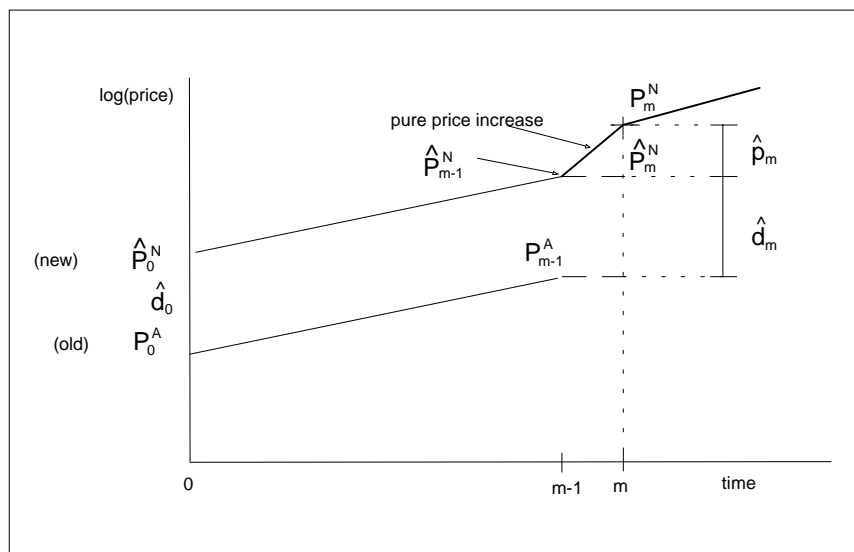
A price-log scale representation shows two parallel curves for the same relative movement in the same time interval (figure 3).

The basic hypothesis for any quality-effect measurement is that the quality price does not vary during the year, particularly between *m-1* and *m*. This is an ad-hoc assumption, owing to the uncertainty about the date at which the quality difference is measured. The possible error due to a real change in quality price during the year is probably negligible.

It therefore seems reasonable to choose the month of December—the usual base month for the price index—in order to carry out the following estimation[11]:

$$\hat{d}_0 = f^0(X_N) - f^0(X_A)$$

which, for the micro-indexes, gives:

$$I^S_{m/0} = \frac{P^N_m}{\hat{P}^N_0} = \frac{P^N_m}{P^A_0 + \hat{d}_o},$$

also written as:

$$I^S_{m/0} = \frac{P^N_m}{P^A_0 + f^0(X_N) - f^0(X_A)}$$

Annual revisions of the model are a desirable but costly goal. In fact, though, it may not even be sufficient, since models should be routinely reappraised with each use. The purpose of the exercise being the best possible measure of quality effect, it is preferable to have at one's disposal the most recent possible regression, which incorporates the last-minute innovations capable of modifying the market examined.

---

[11]Choosing prices collected in a single monthly period also avoids autocorrelation problems.

A method approaching that of pure econometrics consists in using the official or informal costs of the technical components of products. With automobiles, for example, we can determine the cost of parts and options. If ABS or airbags were optional on a car and are now standard, the quality effect can be estimated by using the price of the option when it is fitted as a standard part. In practice, such information is hard to establish with accuracy. At best, we can obtain the cost of the options before their series-assembly, which is much higher. Therefore, we often apply a discount (for example, 50%) to these "option costs" when trying to measure the quality effect they entail.

For these reasons, econometric methods requiring only product-description data seem more accessible, very complex products excepted. The French experience supports this argument, even if—with few exceptions—we are still in the test phase (see §2.1 and §1.2.2).

II. Problems and advantages of hedonic methods: the French experience

Consumer durables and apparel are two categories where the problem of quality-effect measurement is acute. Replacements are numerous and often involve wide price swings. Hedonic methods seem a natural choice here, especially since the products in both categories lend themselves well, in principle, to a detailed description of their characteristics.

2.1. Consumer durables

The management system for consumer-durables replacements is centralized—in contrast to non-durable heterogeneous and homogeneous items, for which the replacement process is managed by the price collectors themselves. When collectors in the field want to replace a durable good, they recommend two alternative replacement products, whose technical characteristics must resemble those of the replaced product as closely as possible. The "consumer durables" team at INSEE's head office decides which of the two recommended products to adopt as the replacement. The choice is based on the characteristics described on the collector's form, but also on manufacturers' data, such as catalogs, contacts with the manufacturers' sales departments, etc. The rationale for this approach is the following: There is a large technological component involved in choosing the product to be tracked. It seems logical, therefore, to entrust this part of the replacement process to a specialized team rather than to price collectors, who can hardly be expected to specialize in all fields covered by the index.

We began by defining a hedonic model for the **dish-washer**. The product is simple and stable enough over time to be described with accuracy, but also sufficiently diversified to ensure that the price decomposition by technical characteristic will be statistically significant—hence usable for operational purposes. The first step was to collect the largest possible set of reliable data, so as to construct regressions with a maximum number of variables. Most of the data are compiled from manufacturers' catalogs, held by the central team in charge of consumer durables. Model and price data are gathered "in the field." This explains why we were able to use the entire sample, with no wastage.

Once we define the first regressions, three broad categories of variables emerge: (1) technical characteristics of the appliance (temperatures, number of programs, noise level); (2) product brand; (3) sales outlets. Variable 1 is linked to those of the dish-washer's qualities that are directly measurable because they can be observed. Paying more for a more sophisticated product is not only natural but also theoretically justified by microeconomics. Variable 2 implies two different price determinants: first, a "marketing" effect that is not always closely correlated with the machine's intrinsic quality but is tied to the company's margin and production cost; second, a "reputation" effect that represents a quality recognized by the consumer but not measured in statistical terms, such as durability, mean time between failures, defectiveness, etc. Variable 3 separates the actual cost of the appliance—the "factory-gate" price—from merchandisers' margins and the service that merchandisers are supposed to offer (advice, warranty, range of products offered, delivery options, and after-sales support). We then need to establish how these factors interact to form the price.

The estimation covered 322 observations divided into 19 brands sold in four different kinds of outlet. This is the sample used for the December 1996 price index.

The explained variable is the price level. We also tested a log-linear model, which yields results of very similar quality. The main statistical indicators of the model's overall quality seem very satisfactory (the detailed results are reported in appendix 1).

- $R^2 = 0.87$
- adjusted $R^2 = 0.86$

Naturally, all the coefficients are significant at the 5% (and even 1%) limit, using the Student's test.

The model comprises two quantitative variables (number of programs and temperatures) and seven dummies. We do not obtain better results by dividing the quantitative variables into intervals. Looking at the explanatory power of each type of variable, we find that the most powerful is the brand name, at 40%, followed by sales outlets at 20%, and lastly the machine's intrinsic characteristics[12]. In short—as the discussion below will confirm—brand appears to be the crucial element in any approach to quality-change measurement.

The abundance of brands required their prior classification. This was carried out on the basis of initial tests and the expertise of INSEE statisticians specialized in durable goods. Brands were divided into four groups. Three of these, in fact, consisted of individual premier brands, accounting for a combined 20% or so of the price observations. The "ordinary" brands, which account for the remaining 80%, were all aggregated into a single group of their own.

Since September 1997, this econometric model is being used to calculate the French CPI. It is centrally managed by the product-replacement team. The extra cost incurred by comparison with standard techniques is acceptable, as the number of replacements averages about twenty a month. While hedonic models may come into more widespread operational use for durables, it is less likely that they can continue to be centrally managed.

---

[12]The explanatory power is measured by the difference between the R^2 values of the full model and the same model without the variable examined.

Moreover, we have noticed from the developments in the dish-washer market that the initial noise-level categories are already outdated for two different reasons: the rapid improvement in dish-washer performance, and the evolution in the technical noise standards themselves. Nine months after the model was created, we need to revise it to ensure that the measurements obtained will remain worthwhile and significant.

The description in terms of objective and often-identifiable technical characteristics is an incentive to use hedonic models for tracking durable goods; so is the centralized management of product replacements. In the apparel sector, the undertaking is more complex.

2.2. Apparel

Apparel product replacements are not centrally managed, but are performed by price collectors in the field. Collectors always try to choose the product that most closely resembles the previous one. For this, they examine the technical characteristics listed on the price-collection form. These data are captured, checked, and entered into the national computer system. While this method gives us a fairly substantial data base, there are several reasons why its quality is far below that of the consumer-durables base.

The first reason is the nature of the product category. In apparel, product characteristics are harder to define objectively without relying on esthetic judgment. A special problem here is the fashion effect. Does a white shirt possess the same utility as a red shirt, all other things being equal? There is no single answer to this question.

A second reason is the abundance of producers in the apparel industry, whereas there are few competitors in the consumer-durables market. There is no such thing as a complete catalog of apparel, or even a catalog of the main producers. Neither fashion magazines nor mail-order catalogs nor consumer-group surveys provide a reliable picture of the apparel market.

Third, a complete description of a clothing item requires many characteristics, and some notations are based on the collector's subjective opinion. For example, in descriptions of skirt length, we find "medium," knee," and "normal" used to characterize the same product. Data gathered in such conditions are obviously complex to analyze.

As a result, the potential samples in the national data base suffer quality losses—sometimes of major proportions—when we attempt to exploit the data for analytical purposes. For some items, the wastage is as high as 40%. Some information is missing through forgetfulness or lack of availability. By and large, however, there has been an improvement in the completion rate of price-collection forms and in the quality of the data entered. INSEE has enlisted the help of the apparel industry itself to improve the training of price collectors. Also, the price-collection forms have all been revised by the apparel-products team in the Institute's CPI Division (the apparel category comprises 142 items). The purpose of these initiatives was twofold: to incorporate industry expertise, and to facilitate the use of the data base by converting many characteristics into closed-ended questions.

The examination of the econometric results obtained yields a less clearcut picture than for durables. Our two examples here are the two items for which we estimated hedonic models: **women's suits** and **men's shirts**.

The main variables selected for *women's suits* are: brand, country of origin, type of sales outlet, fabrics, lining, belt (or lack of one), but also the "season"—summer collection or winter collection[13]. The data are for December 1996. The data base, which initially comprised almost 1,250 observations for nearly 20 variables, was reduced after validation to 1,110 observations and 12 variables. The model's explanatory power is modest ($R^2 = 0.33$), but the model is significant (for detailed results, see appendix 2a).

Once again, a large share—20%—of the model's explanatory power consists in the brand. However, product complexity removes most of the model's explanatory capability.

For *men's shirts*, the results are more satisfactory ($R^2 = 0.77$; detailed results in appendix 2b). Here as well, the variables chosen are brand, fabrics, country of origin, and type of sales outlet. Our sample was supposed to include 387 varieties. In the end, after validation the initial data base and making some empirical adjustments, we were left with only 212 usable basic data observations. These heavy losses are unlikely to be repeated in the December 1998 sample, where the quality of the basic quotation inputs will greatly improve. Likewise, we kept only four types of technical characteristics before any statistical test, as against a potential ten or so in the base at the outset. However, even in present conditions, our results are comparable to those obtained in the United States (Liegey 1994), Sweden (Norberg 1997), and Canada (Marckle 1997).

How can we account for such differences in the explanatory capability of the two models for apparel? The main reason, in our view, is the nature of the products. The "women's suit"' product may actually involve too many non-measurable phenomena for which no satisfactory estimator can be found. By contrast, for "men's shirts" (long-sleeved models, other than 100% cotton), the non-measurability is minimal. It should be noted that if we had been able to use all the variables (color, pattern, collar shape, number of buttons, etc.), we would certainly have found an even more explanatory model. It would also have been more firmly substantiated, rather than swollen artificially by the mere addition of variables. We can therefore track quality effects in this apparel

---

[13]A separation of the model into two sub-models, one for winter clothing, one for summer clothing, yielded results of fairly similar quality.

grouping provided that the items are homogeneous enough to reduce the number of variables required, but also heterogeneous enough to enable the model to contain price-explanatory variables.

Can such models actually be used in the CPI ? The answer is not yet clear. Replacements are performed on a decentralized basis, and the information is checked by INSEE's Regional Offices. The model cannot be managed by a centralized team at the national level, owing to the massive number of replacements—between 1,000 and 2,000 a month for the entire apparel category—and the lack of product standardization. The best solution seems to be the use of hedonic models by the CPI teams working in the Regional Offices. But this implies large-scale training and organization programs, as well as the construction of tools to manage the system and monitor its efficiency. This example shows the twofold difficulty of obtaining a model that is reliable in theory and usable in practice.

2.3. Hedonic models: an incomplete scope of application, but a variety of uses

Producing a good overall CPI requires a tracking of quality effects across the entire economy. It would not be enough to perform explicit quality-change adjustments in some industries while leaving others to the subjective judgment of price collectors. For durables, explicit procedures are feasible because the product is well defined with objective criteria. For apparel, we face an initial set of problems in measuring clothing quality. To assess quality in the service sector, we are sometimes unable to define precisely what the product is. As for measuring the flavor of a food product...

Let us take some examples to illustrate these problems, starting with musical entertainment. The price of the show, on any given day, depends on the artist's celebrity. How can one measure the quality change in a monthly index? Often, for theaters or stadiums, we can track the price of a season ticket and compare its change from one year to the next. But this method doesn't work for single performances. Many restaurants put "psychological" prices on their menus (ex. 99 francs). From one month to another, however, the portions in the plate can vary: the quantity of steak is reduced surreptitiously... This phenomenon, as well, is hard to evaluate. For a large share of consumer products, the use of hedonic models in the price index seems unfeasible. Better, then, to use "systematic" methods while improving the framework in which they are applied.

Despite the occasional difficulty in measuring quality effect in explicit terms, we can draw some lessons from econometric studies in an initial approach to the issue.

Two factors have a crucial influence on price levels: brand and sales outlet. This may sound paradoxical, since neither is—strictly speaking—a physical characteristic of the product. In fact, what we need to realize is that product quality, while not necessarily measurable, can be estimated on a proxy basis from the brand or the sales outlet. The brand serves as a guarantee, a label of reliability or performance for the consumer. Sometimes, it also denotes a fashionable or prestige item.

The sales outlet does not play exactly the same role as the brand. Some types of outlet are widely known to be more or less expensive for reasons of economy of scale or efficiency of distribution channels. However, the quality of product-related services also helps explain why a product is expensive. Warranty duration, delivery, alterations, and advice are all characteristics of the sales outlet. Bearing this in mind, we can readily accept that a share of the product price depends heavily on the type of sales outlet where the product is found. Thus, a classification of brands and sales outlets seems to be a necessary first step toward a systematic treatment of quality effect.

Experience shows that hedonic models are very expensive to set up. For large price samples, as in France, they are not easy to use on a daily basis. However, there are several possible approaches for analyzing the results of hedonic models. First, in cases where an explicit, reliable, and robust estimate can be obtained, we should not pass up the opportunity of using hedonic models for CPI production. Second, econometric studies should make it possible to improve the identification of products and specify the characteristics that need to be entered in the price-collection form: the collector can be given guidelines for this purpose. Third, we should rank the explanatory parameters of prices in order to allow the creation of quality-equivalence classes: replacements could then be carried out within these classes without further quality adjustments. In fact, this is a "checklist" approach (Marckle 1997).

Conclusion

When they can be used, econometric methods appear to provide the only viable explicit treatment of quality effect. The workload involved in using such equations to treat a large share of CPI products is colossal. To realize this, we need only count the number of countries using hedonic models for CPI production. Pooling studies on countries with similar markets should make it possible to save time and resources. A comparison of France's experience with that of its European and North American partners suggests some economies of scale might be achieved, even if the models are obviously not transferable "as is." In sum, preliminary studies and model maintenance are the two constant challenges facing price statisticians who seek a precise measurement of cost-of-living changes. The transition from theory to everyday practice, from equations to the complex and shifting reality in the field, is another equally daunting challenge.

References

Michael J. Boskin *et alii:* « Toward a more accurate measure or the cost of living » final report to the Senate Finance Committee, December 1996.

François Lequiller: « Does the French Consumer Price Index overstate inflation ? » Economie et Statistique n°303, April 1997.

Pascal Ardilly et Francis Guglielmetti: « La précision de l'indice des prix: mesure et optimisation », Economie et Statistique n°267, 1993.

Paul Liegey Jr: « Adjusting Apparel Indexes in the Consumer Price Index for quality differences », Monthly Labor Review, May 1994

Anders Norberg: « Quality adjustment in the Swedish Price Index for Clothing », a paper presented at the second International Conference on Prices Indices (Ottawa group), Stockholm, November 1995.

Terri Markle: « An application of the hedonic approach to clothing in the CPI », a paper presented at the Conference of European Statisticians, joint UNO/ILO meeting on consumer price indices, Geneva, November 1997.

Appendix 1

Hedonic model for dish-washers

The dependent variable is the price level. The model adopted is linear. It is based on 332 observations from December 1996. After validation, the model contains ten variables and the constant representing the reference situation: two quantitative variables (number of temperatures and number of programs) and eight dichotomous variables divided into three groups: sales outlet, noise level and brand reputation.

The model's overall indicators are:

- R^2=0.87
- adjusted R^2=0.86
- prob>F=0.0001
- Root MSE=350

VARIABLE TYPE	DESCRIPTION	NUMBER OF OBSERVATIONS (%)	COEFFICIENT	STUDENT'S T
.	Number of programs	100	93	5.325
Sales outlet	Hypermarket	17.4	reference	.
	Department store	4.6	302	2.909
	Small store	22.3	300	4.674
	Specialized large retailer	53.7	175	3.298
.	Number of temperatures	100	286	7.324
Noise level	Very noisy	3.8	reference	.
	Average noise	58.7	567	3.957
	Low noise	38.5	1117	7.282
Brand reputation	General class	81.7	reference	.
	Fair	9	345	4.333
	Good	2.2	596	4.206
	Very good	7.1	2643	33.303
.	Model constant	.	1986	12.743

Appendix 2a

Hedonic model for women's suits

The dependent variable is the price log. The model adopted is log-linear. It is based on 1,288 observations from December 1996. After validation, the number of observations was reduced to 1,110. The hedonic model contains the constant representing the reference situation and 12 dummies divided into six groups: sales outlet, majority fiber, finish, summer or winter collection, country of origin and brand reputation.

The model's overall indicators are:
- R^2=0.33
- adjusted R^2 =0.32
- prob>F=0.0001
- Root MSE=0.464

VARIABLE TYPE	DESCRIPTION	NUMBER OF OBS (%)	COEFFICIENT	STUDENT'S T
Country of origin	Non-EU	28.9	reference	.
	EU	6.6	0.275	4.546
	France	64.5	0.133	4.146
Sales-outlet	All outlet types	79.6	reference	.
	specialized large retailer	20.4	-0.223	-6.065
Fiber	All fibers	9.4	reference	.
	Wool>30%	73.6	0.205	6.731
	Cotton>65%	17	0.147	1.814
Brand reputation	General class	52.1	reference	.
	Low-end	0.9	-0.754	-5.211
	Unknown	27.8	-0.024	-0.741
	Good	9.3	0.337	6.559
	Very good	9.9	0.766	15.054
Collection	Winter	83.1	reference	.
	Summer	16.9	0.282	7.036
Finish	Belt	84.5	0.032	1.261
	Lining	69.6	0.103	3.248
.	Model constant	.	6.636	136.426

Appendix 2b

Hedonic model for men's shirts

The dependent variable is the price log. The model adopted is log-linear. It is based on 387 observations from December 1996. After validation, the number of observations was reduced to 212. The hedonic model contains the constant representing the reference situation and 9 dummies divided into four groups: sales outlet, majority fiber, country of origin and brand reputation. The model's overall indicators are:

- R^2=0.77
- adjusted R^2 =0.76
- prob>F=0.0001
- Root MSE=0.265

VARIABLE TYPE	DESCRIPTION	NUMBER OF OBSERVATIONS (%)	COEFFICIENT	STUDENT'S T
Country of origin	Outside France	59	reference	.
	France	41	0.120	2.764
Sales outlet	Hypermarket	24.1	reference	.
	Department store	10.8	0.349	3.693
	Small store	52.8	0.475	6.065
	Specialized large retailer	12.3	0.121	1.716
Majority fiber (>50%)	No majority fiber	9.4	reference	.
	Cotton	73.6	-0.230	-3.588
	Polyester	17	-0.461	-6.186
Brand reputation	General class	42.8	reference	.
	Fair	27.4	0.362	4.899
	Good	20.8	0.526	7.287
	Very good	9	0.792	8.509
.	Model constant	.	4.816	70.106