Static samples in a dynamic universe: the potential use of scanner data and hedonic regression in the compilation of consumer price indices.

(May 7, 2004 version)

David Fenwick and Adrian Ball, Consumer Prices and General Inflation Division, Office for National Statistics, 1 Drummond Gate, London SW1V 2QQ.
Email: david.fenwick@ons.gov.uk and Adrian.ball@ons.gov.uk

Abstract

The practical measurement of the change in the cost of a fixed basket of goods and services raises significant challenges. This is particularly so when set against the background of a dynamic market place where new outlets open and old ones close and more particularly where items frequently disappear from the shop shelves only to be replaced by new ones which may have radically different features. Yet this is typical of the situation that faces compilers of consumer price indices when dealing with fast moving hi-tech electrical goods.

How do compilers keep the basket fixed whilst at the same time ensuring it is up-to-date and representative? This paper provides some practical solutions based on recent results from an extensive study using scanner data for electrical goods. Scanner data is used to check the validity of the achieved sample against the corresponding sales data. Any resulting mismatch is addressed by using the same scanner data to provide a sampling frame for a more scientific and rigorous sample selection of items in retail outlets. Items are selected using probability sampling based on selecting a bundle of features or item characteristics that most influence price according to hedonic regression. This is different from the more traditional approach to probability sampling of selecting specific models. The same regressions can also be used to make explicit adjustments for changes in quality arising from differences in features between an old and a replacement item. This approach can significantly improve the representiveness of the sample and can more effectively address the potential problems of bias. This is illustrated with some examples. The solution that the Office for National Statistics has adopted represents an integrated process for sample selection and quality adjustment that makes common use of scanner data and adds to the coherence of the index construction.

Key words: dynamic sampling, measuring quality, scanner data, hedonic regression, integrated processes, coherent methodology.

Background

Compilers of fixed-basket consumer price indices are typically faced with the dual challenge of:

- Maintaining a representative and fixed basket of goods being priced.
- Imputing a new base price for an item that has been selected to replace an item that is no longer available in the shops to be priced. In the course of doing so prices are adjusted for changes in the characteristics of replacements so that like for like comparisons are being made.

The potential adverse effect on the statistical integrity of the index from a failure to satisfactorily meet these challenges depends on the nature of the product. In particular the problem increases with the observed turnover of the items concerned and with the level of product development associated with product replacement.

It can be observed that traditional methods employed by national statistical offices to overcome these problems are not optimal. This is partly caused by the practice adopted by most index compilers of using a “one solution fits all” approach that doesn’t take into account the varying degrees of product development associated with different goods and services. An example of this is sample replenishment of the whole index with the same periodicity via chain linking. This ignores, in large part, the fact that market evolution progresses at differing speeds depending, for example, on the stage in the product life cycle, and that for some goods and services market evolution is inherently more dynamic than for others.

It can also be observed that in practice the scope for coherence in compilation through the use of integrated datasets and methods is currently not fully realised. For example, procedures for the “quality-adjustment” of prices and processes used for sample replenishment are usually treated as operationally independent, whereas in reality they are closely connected with a number of inter-dependencies and with a reliance on common data sources. At the very least, procedures adopted for product replacement will have a real and practical impact on the need to quality adjust prices and information that informs decisions on product replacement is also the same information that should inform judgements on the measurement of quality adjustment.

Scanner data[1] can allow us to harmonise procedures, providing the opportunity for:

- Identifying specific areas of the index where achieving a representative sample can be particularly problematic;
- Delivering targeted tailor-made solutions to sampling;
- Addressing the associated issue of quality adjustment in a coherent and integrated way.

In summary, common methodologies using common data sets can be applied to the related issues of sample replenishment and quality adjustment.

The sampling and measurement issues: background

Consumer price indices are typically calculated in two steps:

- Price indices for elementary aggregates are calculated using a variety of different formulae, such as arithmetic and geometric means, to combine individual price quotes. The elementary aggregates are a set of goods or services that are chosen to be as homogeneous as possible. There is no explicit weighting of the price quotes.
- Higher-level indices are calculated as a weighted average of the price indices for the elementary aggregates where the weights relate to expenditure shares.

[1] Based on Electronic Point of Sale (EPOS) data recorded by bar-code readers.

The construction of elementary aggregates will be constrained by the availability of information but is often distinguished by geographical region and by outlet type (as part of the sample stratification). This multi-stage sampling is introduced to reduce the variation of prices within the elementary aggregate and the dispersion of the price movements. It involves the sampling of shopping locations, retail outlets within locations and items within retail outlets.

In reality, whilst many statistical offices including the UK Office for National Statistics (ONS), attempt to use some form of random sampling for the selection of shopping locations and retail outlets, this is not so for the selection of items within the retail outlets, usually because of the prohibitive costs involved. Typically, for each shop selected price collectors will instead ask the shopkeeper to identify for each pre-selected item in the Consumer Price Index shopping basket, the variety or product that is most representative of what customers buy (simplistically the most "typical")[2]. As a result the validity of the sample of prices that is collected is particularly dependent on:

- The validity of the outlets selected within shopping areas. If the sampled outlets are not statistically representative then neither will be the varieties or products that their customers typically buy.
- The validity of sample selection within outlets. In particular, the guidance given to price collectors for selecting varieties and products in shops, the way the guidance is followed in practice and the knowledge of the shopkeeper in providing a properly informed response to the questions posed by the price collectors. Additionally it is important to note that there tends to be inherently less control of local sampling within retail outlets and as a consequence more reliance on price collectors in the field.

The potential difficulties with procedures used by statistical offices for selecting purposive samples of varieties and products for the pre-determined basket of items is further exacerbated by the challenge of attempting to maintain a static sample in a dynamic universe. In particular:

- The extent to which the sample remains representative is highly dependent on the rules used for item replenishment when a particular variety or product disappears from the shelf of a particular outlet. Typically, price collectors are instructed to replace a missing item with one that is most similar to the one that has disappeared. This has the advantage of minimising issues relating to quality adjustment but, by fixing the sample on past sales, reduces the relevance of the sample over time;
- The problem increases with the rate of turnover in varieties and products and with the rate of product development.

Intuitively sampling and sample replenishment issues can be expected to be particularly challenging for high turnover hi-tech goods or for those areas of the

[2] Collectors are sent to shops with a generic item description, for example a 24" to 32" widescreen television or freestanding, 12 place setting, dishwasher. It is then up to the collector to choose the particular model selected for price collection. If there is more than one model available to the collector, the model most representative of customers' purchases in terms of sales is selected. Where possible, this is done in conjunction with store managers, who should know what sells well. When a replacement needs to be chosen they select the most similar model to the one that is no longer sold.

consumer market, such as electrical goods, that have traditionally experienced a high rate of product development. It is against this background that ONS compared achieved samples from its traditional approach to variety and product selection as described above with corresponding figures from scanner data on total sales[3]. The results can be summarised as follows:

- There seems to be over-selection of low-selling models of well known brands. Feedback from field supervisors and auditors suggests that this is caused, in part, by an element of brand loyalty by collectors in the selection of products. This is further accentuated by limited product updating at the time of the annual review of the basket. The latter focuses on identifying new items for inclusion in the basket and the dropping of old items rather than the selection of particular products or varieties in the retail outlets to represent those items.
- For each model, average prices calculated from prices collected in the shops are higher than the corresponding unit values from scanner data. This in part reflects the fact that scanner data will include some types of discounts, such as managers' special offers, that are typically not covered in a consumer price index.
- There are some consequent differences in price trends although not necessarily of a systematic nature, and hence not necessarily causing a bias. The results suggest more variation in price changes in the CPI sample than the scanner data.

It was this evidence that lead the ONS to develop, for a limited range of items, a system of local probability sampling and to develop in parallel explicit quality adjustment based on hedonics applied to scanner data.

Local Probability Sampling: the scheme adopted by ONS

In order to address the sampling problem ONS tested several alternative methods which were subsequently rejected – such as asking collectors to list all models available in a shop, and then randomly selecting from that list (a technique used by the Bureau of Labor Statistics in the US), or using scanner data to produce a random sample by giving each collector a prioritised list of model numbers to select from. The former was rejected because the method demanded, in a UK context, an unrealistic level of statistical expertise from the price collectors (which was only available at a high cost). The latter was rejected because, when piloted, it yielded a low coverage rate. Both were, for different reasons deemed not to be cost effective.

The solution implemented by ONS is a variant of the second method, based on groups of price determining characteristics rather than individual models. This has been developed as a practical and efficient way of ensuring the items priced are representative of consumer spending.

---

[3] Fenwick, Ball, Morgan & Silver "Price Collection and Quality Assurance of Item Sampling in the Retail Prices Index: How Can Scanner Data Help?

How does the scheme work?

The principle behind the particular approach adopted is to use the selling patterns of a combination of attributes (e.g. for televisions these are screen size, sound quality, picture frequency etc.) to create a basic frame for sample selection where the probability of selection reflects the corresponding sales volumes in monetary terms;

- Hedonic regressions are performed to identify the main price-determining attributes.
- Scanner data is used both to produce the hedonic regressions that identify the price determining characteristics and to populate the corresponding sales matrix where each combination of attributes is represented.
- For example in the case of televisions the following matrix was derived:

Brand	Screen size	Teletext	Sound	Frequency	Expenditure
High	28”-29”	Fastext	Stereo	50	16%
High	28”-29”	Fastext	Dolby	100	4%
Medium	28”-29”	Fastext	Stereo	100	20%
Low	28”-29”	Fastext	Stereo	50	8%
Low	28”-29”	No	Stereo	50	2%
High	30”-32”	Fastext	Stereo	50	10%
High	30”-32”	Fastext	Dolby	100	5%
Medium	30”-32”	Fastext	Stereo	50	6%
Low	30”-32”	Fastext	Stereo	50	8%

It is worth noting that, in this example, brands have been grouped together into high, medium and low price categories using an analysis based on hedonic regressions. Compared with precise brand specification, such a generic grouping increases the chances of an appropriate model being found in a particular shop, thereby increasing the effective response rate.

- The sales matrix is then used as the reference for a Probability Proportional to Sales (PPS) sampling scheme for the local selection of combinations of price-determining attributes. Each combination of attributes is given a chance of being included proportional to its total expenditure as shown by scanner data.
- Finally a list of six prioritised attribute groupings is generated for each price collector. Each collector is instructed to find an item matching the first attribute group on the list in their outlet. If this is not possible they move on to the second and so on. If none of the six given combinations is found, the collector reverts to the current method of looking for the best sold in the outlet. An example of a list is given below for widescreen televisions[4].

[4] The Brands represent the broad groupings according to market position as mentioned above.

Choice	Brand	Screen size	Teletext	Sound
1	Panasonic, Sony, Toshiba	32"	Fastext	Stereo
2	Hitachi, Goodmans, Orion	28"	Fastext	Stereo
3	Panasonic, Sony, Toshiba	32"	Fastext	Dolby
4	Samsung, JVC, Sharp	32"	Fastext	Stereo
5	Samsung, JVC, Sharp	24"	No	Stereo
6	Panasonic, Sony, Toshiba	28"	Fastext	Stereo

- The following month the collector attempts to price the same model. If this is not possible the collector searches for a comparable model from the same group, i.e. with the same price-determining attributes. If no such model can be found the collector searches for a model fitting the first attribute group, and the process repeats itself. A flow chart of this process can be seen in Annex 1.
- Finally the sales matrix is updated and the PPS sample re-drawn annually.

This process is further supported by the re-programming of collectors' handheld computers so that collectors can enter the brand and model number of the specific item selected and this can be checked against an attributes database containing the attributes of a number of models- also derived from scanner data. The handheld computer alerts the collector if the model selected by the price collector is contained in the database and the attributes do not match, and then the collector is required to confirm the attributes by re-entering the details - thereby providing further quality assurance.[5]

In summary within the context of probability sampling the use of hedonics and scanner data reduces collection costs with little adverse impact on the sample compared with traditional probability sampling based on model numbers. In addition the PPS sample is completely and systematically reviewed each year and PPS principles are essentially retained for sample replenishment of goods that are no longer available to price. Scanner data loaded on price collectors' personal computers also provides additional quality assurance.

The outcome

Initially the ONS investigated this method of PPS sampling for five items, where an initial analysis of scanner data had revealed significant differences between the samples obtained by price collectors in the field using purposive sampling and sales patterns from scanner data. The five items were: vacuum cleaners; dishwashers; washing machines; 14" televisions and widescreen televisions. Successive pilot collections were carried out in November 2002, March 2003 and July 2003 to refine the method.

[5] The method has also been adapted for items priced by ONS Head Office. The program works in much the same way, though in this case items and prices are recorded in a specially designed spreadsheet.

The main findings were that:

- The proportion of items where PPS probability sampling was successfully applied steadily improved over time, as the method was fine tuned, except for 14” televisions. Excluding the latter an overall success rate of about 80% was achieved.
- A χ^2 test (chi-squared test) of the achieved PPS sample against the sample population from scanner data showed no significant differences. This was in marked contrast to the previous sampling regime.

PPS sampling proved unsuccessful for 14" televisions in large part because of the nature of the market and in particular the relatively large turnover of models but with relatively little real product development in terms of new features[6]. This meant that price collectors were not always able to find a 14” television matching the specified attributes as outlets were reducing the number of models they stocked. This in turn reduced the effectiveness of the new sampling procedures, by reducing the number of times that the method could be successfully used and increasing the necessity to revert to the fallback option of following previous sample selection procedures. For this reason the new methodology was not implemented.

The associated issue of quality adjustment

Access to detailed and comprehensive scanner data also facilitates the explicit quality adjustment of prices when replacements are introduced whose characteristics do not fully match the replaced items. Without quality adjustment for changes in item specifications a consumer prices index will reflect price changes that extend beyond those associated with a fixed basket. One of the particular attractions of using scanner data in the context of local sampling is that it adds coherence to index construction by using a common source of information for both sampling and quality adjustment of prices. Varieties and products are initially sampled using scanner data as the sampling frame, sample replenishment uses the same scanner data and the latter is also deployed in the quality adjustment that is necessary when sample replenishment takes place. More generally scanner data informs choices about sampling methodology and quality adjustment methods, for example whether for specific groups of goods and services traditional sampling techniques and implicit quality adjustment methods are adequate.

ONS has been progressively moving to explicit quality adjustment methods and in particular hedonic regression techniques where the latter adds integrity to the index.

There has been past criticism that the application of hedonics for specific quality adjustment has not always been centred on the most productive expenditure items. This was the conclusion reached by the Schultze Panel, for instance, about the work undertaken by the US Bureau of Labor Statistics. The ONS have been addressing this issue by investigating indicators that might be used to identify in advance those goods and services where hedonics might best be applied.

[6] PPS sampling was subsequently dropped for 14” televisions but was additionally used for digital cameras which entered into the consumer price index for the first time at the beginning of 2004.

The approach being actively pursued is based on the two-dimensional matrix below and is a departure from the Implicit Quality Indices, which can be time consuming to compute[7]. Clearly if an index item is subject to both a high rate of turnover and of technological change then it will be a prime candidate for explicit quality adjustment. Similarly, the index compiler would not want to give priority to explicit quality adjustment where an item has low turnover and technological change. Both indicators are easily calculated, again from scanner data or from the raw data used to compile the CPI. The idea is being further developed, aimed at providing a more systemised and analytical approach to quality adjustment. It also fits neatly with the more strategic implementation of more rigorous PPS sampling.

	Low rate of technology change	High rate of technology change
Low turnover rate	NO	?
High turnover rate	?	YES

Experience indicates that a systematic and analytical approach is needed both to the initial construction of the hedonic model and to subsequent updates.

ONS application of hedonic methods

The ONS applies a predicted price approach to hedonic quality adjustment. The ratio of the prices predicted by the hedonic equation for the replaced and replacement items is used to adjust post-hoc the base price of the replaced item. Experience suggests that it tends to be more stable and the outcome less susceptible to the impact of multi-colinearity on the individual coefficients than other methods of application[8].

[7] Implicit Quality Indices (IQIs) were first developed by Jorgen Dalen and Don Sellwood in the context of the European HICP. They are a measure of the effect of the operational adjustments that have been made to the raw price data in order to obtain the published "quality adjusted" price index. That is the aggregate effect of adjustment, including explicit and implicit methods, to remove non-price effects and arrive at the "true" price change. Relatively large IQIs for specific items may indicate areas where particular attention needs to be given to the quality adjustment techniques

[8] There are three methods of applying hedonics to construct a quality-adjusted consumer price index:

- **The individual coefficient approach**. This uses the coefficients from single-reference period hedonic functions to adjust post-hoc the observed price of the replacement item to impute a new base price. It relies heavily on the reliability and stability of the individual coefficients estimated in the hedonic function
- **Time dummy variables**. This uses single regressions (with time dummies) covering all periods, which are re-run each time the index is compiled, with fixed "characteristics" coefficients and the difference between the time dummies is taken to represent the price change excluding quality differences.
- **Predicted prices**. This uses the ratio of the prices predicted by the hedonic equation for the replaced and replacement items to adjust post-hoc the price of the replacement item. It tends to be more stable and the outcome less susceptible to the impact of multi-colinearity on the individual coefficients.

The presence of colinearity between independent variables means that hedonic regressions may not be globally optimal and there may not be a uniquely best functional form. Rather the best that can be achieved is a local optimisation that gives the best regression fit possible with results that are compatible with what is expected of the market. For this reason a total reliance on regression diagnostic statistics (such as partial F-values and adjusted R-squares) should be deferent to a more systematic, multistage system relying on human intervention at key stages.

This iterative approach has been adopted with success by ONS following closely the Statistics Canada practice[9] and is repeated until a satisfactory result is achieved. It consists of the following steps;

1. Choose base values for dummy variables – these are left out of the regressions.
2. Run regression with all variables, and produce correlation matrix.
3. Examine correlation matrix, and associated statistics, for information on colinearity between independent variables. Look at colinear pairs and decide whether the variables can be combined, or one variable dropped.
4. Examine residuals for evidence as to whether to make continuous variables discrete dummies.
5. Re-run the regression with the amended variables.
6. Remove variables with low t-values (at this stage $t<1$).
7. Re-run the regression with the remaining variables.
8. Progressively add and remove variables, until a combination is reached that produces the best fit, with coefficients in line with market expectations (in particular positive coefficients for included variables).
9. Examine residuals and Cooks d-statistics for evidence of outliers.
10. Either remove the outlier observations or add extra attributes to remove their influence. Then re-run steps 1 to 9.
11. Look for evidence of missing variables. If this is present revisit data source for added information.
12. Group together dummy variables within brand, sound card and video card that are not significantly different.
13. Run final regression.

It should be noted that market knowledge as well as statistical expertise is required to successfully perform the hedonics. This is because judgement is needed on when to stop the iterations and the latter needs to be based both on an evaluation against market expectations and on statistical diagnostic tools

Regular updating of models to a pre-determined timescale is best avoided:

- It is not efficient as models may be updated unnecessarily.
- It is not fool proof as models may need to be updated between pre-planned updates.

[9] Personal conversation with Fred Barzyk, Statistics Canada, June 2001.

Similarly updating of models unsystematically should also be avoided because of the risk of missing a necessary update resulting from, for example, a significant market change.

For these reasons the ONS have adopted a mixed approach where updates of models are triggered by thresholds based on an analysis of predicted against actual prices but where these are underpinned by a pre-determined minimum number of updates. The difference between predicted and actual prices is used to trigger an update in the hedonic function but the latter are subject to a minimum number of updates. This approach has a number of advantages:

- It is scientific and makes maximum use of the data and the available statistical tools;
- It minimises the risk of not updating the hedonic regression when there has been a change in the market that was not manifest at the time through knowledge of the market;
- It avoids unnecessary updating.

The method applies the following routines:

1. The 99% confidence interval for the average difference between predicted and actual price is estimated each time the model is updated.
2. If the average difference in the current month is outside the 99% confidence interval then the impact on the corresponding sub-index index and on the all items index are estimated.
3. The model is then updated if this changes the published all items index. In the UK the latter is presented to one decimal point.
4. The model is updated at least every four months.

The rationale behind this approach emerged from an analysis of the comparative trends in actual and predicted prices for personal computers as shown in Chart 1 below. The findings were twofold:

- *The need for periodic model updates* because of an increasing divergence over time between actual price and the predicted price derived from the hedonic model. This is associated with the fact that as the hedonic model ages it will generally over price older attributes compared with the current market valuation (by ignoring new attributes entirely). The consequence is an over-adjustment for changes in item attributes and a quality adjusted index that has downward bias. This systematic difference is a logical reflection of the market and is dealt with by regular updates of the model. For instance, in the chart the model has been updated in May and October and the relationship between actual and predicted price is reset.
- *The need for additional checks to be in place* due to the unpredictability and impact of step changes in the retail market. For instance it is clear from the chart that the relationship between actual and predicted price changed dramatically in August. Subsequent investigation revealed that in this period both Intel and AMD significantly reduced the cost of their processors to PC producers. This in turn changed the relationship between the total cost of PCs and the cost of a processor, making the regression model out-of-date.

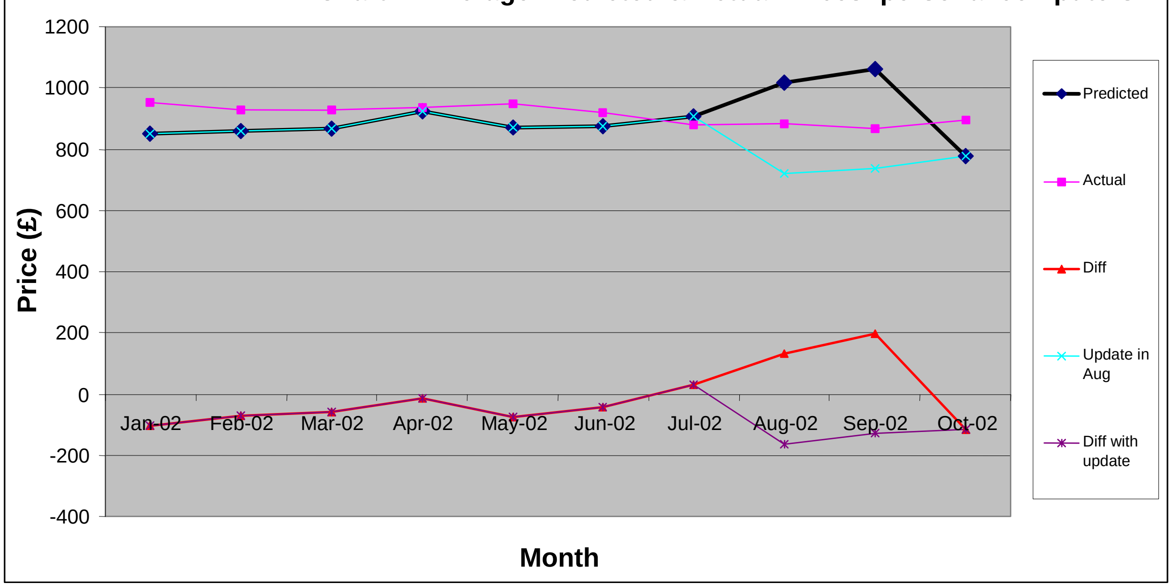


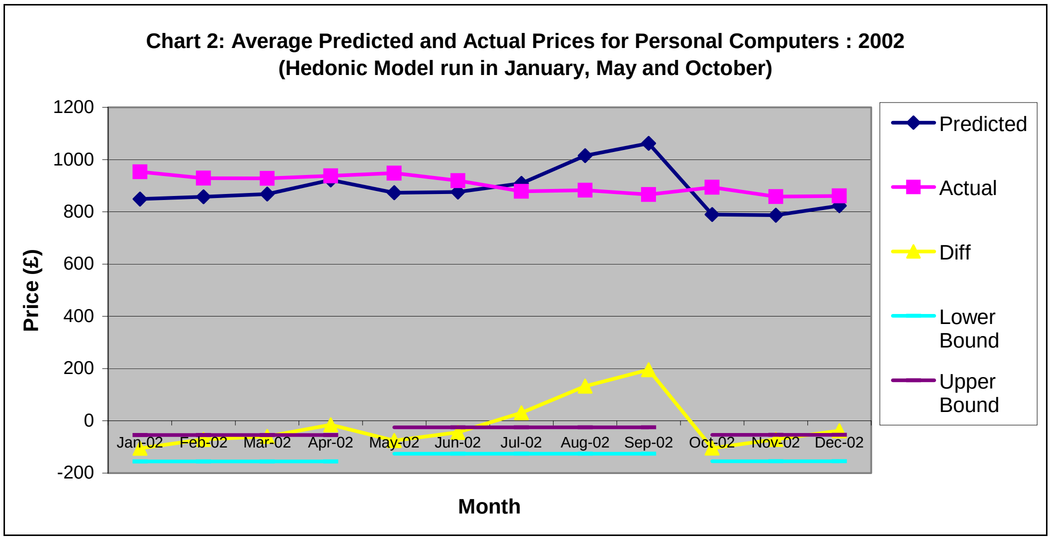
Chart 1: Average Predicted & Actual Prices: personal computers

It is an analysis of confidence intervals[10], as shown in Chart 2, which underlies the specific solution adopted to address the need for additional checks to be in place to ensure a fail safe mechanism for ensuring the hedonic models are up-to-date. It can be seen even more clearly that a rigorous and systematic but fail safe process for model updating is essential. This is because of the systematic nature of the inherent bias in the hedonic function over-time and the propensity for market shocks to invalidate the hedonic function. The use of the confidence interval[11] for the average differences between actual and predicted prices achieves this.

---

[10] Standard error of model prediction = $\frac{1}{n}\times\sqrt{\sum_i SE_i^2}$ = STDI. But as model produces predicted value of log price: Standard Error Price = $\Pr edicted\times\sqrt{\left(STDI^2-1\right)}\approx \Pr edicted\times STDI$.

[11] A 99% confidence interval was used. As the confidence interval was calculated from the differences between actual and predicted price and then applied to quality adjustment based on a ratio this broadly equates to the standard 95% confidence interval for the average difference between actual and predicted price.



Summary and conclusion

This paper argues for a more integrated and scientific approach to sample replenishment and quality adjustment through the use of scanner data and hedonic regression. It argues that such an approach, when applied to high-turnover hi tech items, will add to the statistical integrity of a consumer price index. In particular the approach will address potential issues relating to sampling and price measurement bias. The core elements of the approach are:

- The introduction of local probability proportional to sales (PPS) sampling based on bundles of item characteristics identified through the application of hedonic regression to scanner data.
- The use of the same PPS sampling for planned and unplanned sample replenishment.
- The use of the same hedonic methodology applied to the same scanner data for explicit quality adjustment.