

Internet retail channels in Price Indices: the challenges involved in including non-traditional retailers in the UK Retail Prices Index

David Fenwick, Anna Brueton and Adrian Ball

UK Office for National Statistics

1. Introduction

According to market research information online shopping in the UK even by 1999 accounted for 10 per cent of computer software retail sales, 5 per cent of books and 3 per cent of music. ONS's own Omnibus Survey shows that by 2002 46% of households had access to the Internet either at home or work.

There are three reasons why we would want to separate out purchases over the internet from other purchases:

- For the statistical integrity of the index. Prices and price trends might be different;
- For quality assurance, for example to help check weights;
- Public expectations and perceptions.

Against this background, the ONS is expanding the scope of price collection to include goods sold over the Internet in the calculation of the Retail Prices Index (RPI), and its European equivalent the Harmonised Index of Consumer Prices (HICP). Currently prices for books, CD roms and supermarket goods are included explicitly from Internet sources, and the Internet is used as a convenient source of prices data for outlets where goods can be bought at the same price on the Internet and from traditional stores. As expenditure through this channel becomes more significant, we will wish to collect prices for more goods and introduce appropriate weights in to the index, to ensure that such purchases are included with the correct relative importance.

Preliminary investigations have been undertaken in many of these areas, and results are included in the main part of this paper.

2. The inclusion of internet prices in the RPI: a critique of the current strategy

The approach currently adopted is essentially a pragmatic one that exploits the systematic processes already in place for sample replenishment. These are described in the paragraph that follows. But first there is a matter of definition and coverage.

What is the definition of internet expenditure? Does it require the money transfer for a purchase to be online or just the commitment to purchase or order to be placed on line? Clearly this is not a trivial question:

- There will be those that order on line and pay by post, fax or in person, e.g. on delivery and receipt of an invoice;
- Consistency of approach with other remote forms of shopping, such as by catalogue, also needs to be considered. It can be argued that internet shopping is distinguished from catalogue shopping by the very nature of it being totally “proactive” and “interactive”. Thus, unlike catalogue shopping, the purchaser can pull information off the web site and complete a transaction totally within the web site.
- Consistency between weights and prices domains. Online transaction prices might differ from others.

The definition used in ONS’ Expenditure and Food Survey is that the customer has to pay online. However, in the case of services there can be a practical problem of distinguishing between purchase and payment. For instance, if the initial purchase of a service- say insurance- is made over the internet, under what circumstances would we classify the renewal of that service as an internet purchase? Would this be if the initial discount given for purchasing the service over the internet continues even if the renewal is done over the telephone or only if the renewal is paid for online? Should we take each transaction in isolation? How in practice would we differentiate online renewal from the renewal of the same service which is initially purchased by other means? Clearly, we could easily end up with a mismatch in the index between weights and prices (the final bullet point above).

How does ONS cover internet shopping in its consumer price indices? It is essentially covered by our standard procedures for keeping the sample of shops and items up-to-date and representative. The updating of the sample of retail outlets whose prices are sampled for the RPI is done systematically on a rotating basis, using a random sample stratified by region and size of shopping population and by outlet type. A quarter of locations are re-sampled each year with complete replacement over a four-year cycle. ONS conduct a full enumeration of retail outlets in sampled areas to ensure that price collection is taken across the full range of retail outlets found in the chosen locations. In addition market research information and data from other ONS enquiries is used to identify new types of outlets and shifts in purchasing patterns including towards use of the internet. The latter is taken on board in the RPI using the following criteria to assess the potential impact on the index:

- ***Expenditure threshold.*** As a guideline (but not a hard and fast rule) we refer to a draft HICP sampling regulation. This requires that the sampling frame covers any distinguishable subset of outlets that accounts for more than 1 to 2 per cent of total household expenditure.
- ***Dissimilarity in price movements.*** Clearly, in order to maintain representiveness, it is particularly important to include new outlets types if the price movements of the goods they sell are different from those observed for outlets already included in the RPI sample. This can sometimes be difficult to monitor and the results not necessarily predictable, for instance in a competitive market where lower Internet prices can encourage a traditional retailer to discount. In such circumstances it is possible that the appearance of Internet shopping will have lead to a one off fall in prices in traditional shops that will already have been picked up by the index prior to any direct inclusion of Internet prices. In the UK this applies, for instance, to books where the advent of Amazon has led to more competitive discounting of “best sellers” in traditional shops.

- **Perceptions.** The integrity of the index can be undermined if there is a perception that the index doesn't cover all appropriate retail activity even though exclusion has no significant numerical impact.
- **Conceptual and definitional considerations.** For example it may be considered inappropriate to include Internet shopping where this involves membership of an organisation. This depends on the rules for compilation of the particular index.
- **Feasibility.** For example it is considered impractical to cover auctioned and bartered goods even if we wanted to (this includes Internet bidding).

Our general feeling is that our current review arrangements pick up new forms of shopping such as online via the internet reasonably well and that the guidelines for inclusion are robust in the current retail environment despite the fact that they are guided by pragmatism rather than a rigorous conceptual argument. However, it is arguable whether our current methods will provide a suitably robust approach in the future particularly if internet shopping increases in significance. For example, from a more strategic viewpoint we need to consider:

- Whether internet shopping should be considered as separate "outlets", in which case, it will need to be fully integrated into the systematic 4-yearly enumeration process described above. Additionally, such an approach will also have an impact on the structure of elementary aggregates and the RPI sample. In particular, consideration will need to be given to whether internet shopping should be treated as a separate stratum for sample selection and elementary aggregation.
- Whether items purchased from the internet should be treated as different items from those purchased from more traditional outlets on the grounds that a different type of service is being provided. In this regard, it is worth asking whether internet purchases are essentially any different, at least in principle, from mail order by catalogue. Certainly in the UK the latter has been a significant part of retailing over a number of years and has recently experienced a resurgence from active participation in this sector by fashionable retailers such as the UK clothing company Next. A number of considerations arise:
 - The increasing strength of this sector may be attributed, at least in part, to the convenience for which shoppers are willing to pay a charge for postage. In circumstances where exactly the same good can also be bought for exactly the same price from a physical outlet then the additional cost of postage and packaging may be considered as the minimum consumer valuation and the price paid for the convenience. It is therefore indicative of the "quality adjustment" that should be applied to the total gross price. It could thus be argued in some circumstances that delivery charges are netted out of the price index as a quality adjustment.
 - If it is accepted that there is very little difference between mail order shopping and purchasing via the internet, conceptually at least, then as a matter of principle a consistent approach should be followed. Alternatively and more pragmatically, at the very least the experience gained from measuring the former should help to provide practical solutions to the measurement issues associated with the latter. It should be noted that in the current RPI design, mail order shopping is treated as an outlet dimension to sampling and is included as a sub-stratum of the stratum for "multiples".

The latter also includes standard retail chains (where prices are collected either from local outlets or via Head Office) and mail order.

- The distinctions between internet and mail order shopping and also with traditional shopping are fast becoming blurred in that, for example, mail order shopping often provides the facility to order and pay by post, over the telephone or over the internet. In addition, for a significant number of goods and service customers have a choice over how they shop with, for example, many supermarket chains offering purchasing over the internet “for added convenience” as well as purchasing at a physical outlet. Thus a number of retailers offer a variety of modes of purchase for identical goods, again making the distinction less clear. This also has implications for measurement.

Whichever view is taken of the treatment of internet shopping in relation to other forms of shopping, the practical measurement issues share common ground with other types of shopping and index compilation. Thus, to summarise, whether internet shopping is any different from any other form of retailing is arguable and perhaps the main distinction from other forms is the practicality of measurement, i.e. of collecting price observations proportional to sales. But what is, perhaps, lacking is a strategic approach focussing on the longer-term. This is something that the ONS is currently addressing in its research programme and it will be looked at in the context of a wider consideration of other relevant forms of shopping.

This paper, amongst other things, seeks to stimulate debate on the need and formulation of an overall strategy, and specifically whether there is a need, at this stage, to include expenditure via the Internet as a fully disaggregated separate “outlet” or “item”, where the expenditure share is the primary issue. In addition, there are a number of more technical areas that would benefit from further analysis and discussion.

3. Issues for further analysis and discussion

The production of price indices for goods bought over the Internet is an issue of growing importance in the calculation of the Retail Prices Index. In 2000 ONS introduced the explicit pricing of books purchased from the Internet. From 2003 it will be pricing some food, alcohol and household consumables from supermarkets ordered over the Internet. However there is growing evidence from ONS’s Expenditure and Food Survey (see later paragraphs for further detail) that this should be extended to other areas, for example music CDs, and with the projected growth in Internet expenditure this is likely to extend more widely.

However, there are many issues to be considered before we can include Internet expenditure in the index for many items, and in a more systematic way. As mentioned earlier, there are two basic strategic choices: either make internet collection conform with more traditional price collection or extend definitions and compilation procedures to encompass the distinct practices associated with Internet retailing. Several issues need to be addressed, both conceptual and practical, in considering the relative benefits of these two alternative approaches. The most important of these are:

- Initial identification of products where internet sales to households have become significant, and sources of sales information for weighting;

- The treatment of delivery charges. Should they be included in prices, or charged separately? How should they be calculated given that there is often a fixed minimum charge?
- Differentiation between purchases for domestic consumption and those for business;
- The treatment of sales from overseas websites, and whether they are within the scope of the RPI/HICP (i.e. does the transaction take place within the usual geographic boundaries associated with the index);
- Whether there is a quality adjustment issue arising from a change in the mode of shopping (e.g. greater convenience);
- The production of a sampling frame of Internet outlets, from which a random sample of outlets can be taken;
- The volatility of the Internet market, and how to deal with situations where internet outlets become unavailable because they close.

4. Main sources of information for the monitoring the levels of Internet purchasing

4.1 High level monitoring of shopping patterns

ONS's Expenditure and Food Survey is a sample survey of around 7,500 households which collects information on household spending, and is used to provide weighting information for the RPI. From April 2001, information has been collected on purchases over the Internet. The first year's results show that four per cent of households ordered goods or services over the internet (compared with 40 per cent of households who had internet access), and this accounted for 0.9 per cent of expenditure. Details are given in table 1.

A more detailed analysis of the data summarised in table 1 shows that spending is frequently concentrated in only a small number of areas within the broad heading. For example, the highest spending was on transport, but with most spent on airfares, rather than on other kinds of transport goods and services. In general the information so far produced gives a very good idea of which sorts of products we should be looking to price as internet purchases. However, with only a small number of households spending on internet purchases, and just one year's data, we need to take care when using the data for weighting purposes. Despite this, the data is indicative of levels and future trends and of those areas that warrant further consideration.

Table 2 shows where ONS currently price Internet purchases, and where the EFS data suggest that there is potential for further collection. Items in bold in the second column have been introduced this year. The majority of these are part of a collection of non-perishable goods priced at four supermarket chains. This collection was prompted by the analysis of the EFS data, which showed that such purchases were not only found in food and non-alcoholic drinks, but also within alcoholic beverages, household goods, recreation and culture and miscellaneous goods. The scope is limited in the first year, comprising 30 items priced in 2 varieties at 4 supermarket chains offering Internet shopping. If the collection is successful we will expand it in future years. Note that the current procedures do not treat these items as a special category, rather they are integrated within the current structure of the RPI.

Table1

	Recording households in sample	Number of households ordering goods/ services over internet	Households ordering goods/services over internet as a % of all households	Average weekly internet expenditure for households ordering goods/services over the internet (£)	Internet expenditure as percentage of all expenditure on the commodity or service
Food and non-alcoholic drinks	7,444	64	0.9	36.80	0.8
Alcoholic drinks	3,886	29	0.4	10.50	0.7
Clothing and footwear	5,282	34	0.5	27.00	0.3
Household goods and services	7,010	70	0.9	10.60	0.3
Health	3,817	8	0.1	11.30	0.3
Transport	6,578	26	0.3	93.30	0.6
Communication	7,186	8	0.1	40.30	0.4
Recreation and culture	7,430	186	2.5	30.50	1.4
Miscellaneous goods and services	7,315	80	1.1	6.00	0.2
All expenditure groups	7,473	299	4	45.50	0.9

Table2

COICOP group	PRODUCTS BEING PRICED	POTENTIAL FOR INTERNET PRICING
Food and non-alcoholic drinks	non-perishables at supermarkets offering internet shopping	full range
Alcoholic drinks	supermarkets offering internet shopping	full range
Clothing and footwear	none	full range (though some purchases are of specialised goods such as motorcycle kit and replica football shirts)
Household goods and services	household consumables at supermarkets offering internet shopping	mainly household consumables plus some furnishings
Health	none	small spend on OTC medicines & similar
Transport	air fares	
Communication	none	none
Recreation and culture	books, computers & computer parts & accessories, computer software & games	CDs, DVDs, toys, holidays
Miscellaneous goods and services	toiletries	insurance, financial services

4.2 Data for weights

In order to correctly include Internet expenditure in the RPI it is clearly important to know, for each included item, what proportion of sales take place over the Internet. This will allow us to either correctly weight price indices for goods bought over the internet together with other indices (i.e. introduce a specific elementary aggregate), or collect the correct number of price quotes to include them in the index without explicit weighting. The data available from the EFS is helpful in this respect, but as it covers only a small number of households spending on internet purchases, and just one year's data, we need to take care in using the data for weighting purposes. This is not a major issue at this stage but could be in the future if the range of Internet shopping increases significantly and the level of accuracy of the EFS remains insufficient. In these circumstances we may need to consider introducing a supplementary sample. But this could be costly due to the need to filter out those that don't buy online and there would also be a respondent burden issue. Alternatively we could seek other more imaginative approaches to the issue including the use of administrative and accounting data from Internet and associated companies. The experience to date on the latter, based on a joint project with a major credit card company, has not been encouraging. Further details are given towards the end of this paper. The data issues to be confronted are in many ways similar to scanner data but without the convenience of an active group of suppliers who collate and quality assure the primary input.

4.3 Purchasing patterns

In addition to providing information on which products are purchased on the Internet (and the associated expenditure), household budget surveys are also capable in principle of providing more detailed information on purchasing patterns. For instance, they can show what sort of goods were bought at the same time, and how much was spent- information which can be useful in determining delivery charges (see later). In addition they have the advantage over most other sources of generating valuable information on household profiles and can thus provide the necessary data to identify the purchasing habits of households covered by a consumer price index. For example, the UK RPI excludes pensioner households that derive at least three-quarters of their total income from state benefits and pensions and high-income households, defined as those households whose total household income lies within the top 4% of all households. Such a differentiation would not normally be available from an administrative or accounting source.

4.4 Supplementary sources: market research data

Where possible we have tried to verify and supplement EFS data with market research data, though the latter usually includes business as well as household spending, tends to be collated from numerous incompatible sources and can be expensive to purchase. The limited information we have shows similar patterns of purchases. However, one area identified by market research data, but not in the EFS, are internet car purchases, where prices are very different. We will be investigating this further in 2003. The latter is a typical example of an infrequent but high value purchase on which a household budget survey will generally not be capable of providing reliable information. Market research information from retailers is likely to provide a more reliable source.

5. Purchasing services over the Internet: pricing issues

In the case of EFS data collected on services purchased over the internet we have found it difficult to distinguish clearly between true purchases and the payment of a service through

internet banking. In terms of pricing, the issue may appear to be more clear-cut in that what we are concerned to identify is products where different prices are offered for internet purchase. However, this is simplistic without giving recognition to the fact that internet shopping provides a different form of service and may therefore imply a quality difference.

5.1 Delivery Charges

The issue of delivery charges is partly practical (not all sites explicitly state delivery price), and partly conceptual. With the help of Eurostat draft guidelines, we are able to make a broad statement:

“Delivery price should be regarded as part of the price where the good cannot be consumed unless delivery takes place.”

Putting the quality change issue to one side this implies that for Internet purchases, where clearly you can not make use of the product unless it is delivered to the home, delivery charges should be included as part of the price of the good. But behind such a statement lies many differences of opinion and interpretation. This can be seen from extracts of the debate that was held and from the fact that there is no regulation on the issue:

“The issue of delivery charges and their assignment to COICOP/HICP concerns the issue of weights as well as the measurement of prices in the HICP. There are some COICOP groups where the delivery charges are explicitly assigned to the same COICOP groups as the delivered products themselves, e.g. delivery of major household appliances (COICOP 05.3.1/2).

COICOP/HICP follows the NA principle only in those cases where COICOP refers explicitly to delivery and installation charges (COICOP 05.3.1/2). Concluding from this explanation the price representatives to be followed in this COICOP group should include delivery and installation if consumers usually buy the products including these services (such as washing machines).

The HICP does not to use the full NA wording. The part “including any transport charges paid separately by the purchaser to take delivery” has left out of the definition of price on purpose. The decision was based on the argument that further consideration was needed as to whether delivery charges should be generally covered and to which COICOP group they should be assigned.

For those COICOP groups where COICOP does not explicitly give advice where the delivery charges should be assigned, illustrated examples have been produced to the assignment of delivery charges by applying the principle established for the NA. Following this principle the delivery services should be assigned to the same COICOP group as the delivered products. This approach was favoured by some but not all Member States because, e.g. it was consistent with the weights which were based on NA data. The UK added that where the consumer had no choice whether the product was delivered or not (like ordering books and CDs from amazon.com), the price to be followed should include the delivery charge. Furthermore, the question of who arranges for the delivery, the consumer or the retailer, might be worth considering in this respect.

For Germany it was preferable to assign delivery charges to COICOP group 07.3.6 (08.1.0) if

- a) the invoice that accompanied the delivered product showed the price of the product and separately the delivery charges; and*
- b) the delivery charge formed a large part of the total price to be paid for the full service (product plus delivery).*

Furthermore, Germany suggested applying the following general principle: If the delivery charge is a fixed charge regardless of the number of items bought, then it should be assigned to 07.3.6 (08.1.0). If, in contrast, the delivery charge was proportional to the product itself, e.g. a delivery charge of € y per box of water, then the delivery charge should be assigned to the COICOP group the delivered product was assigned to.”

Eurostat's opinion at the time was as follows:

"a) Concerning the question whether the HICP should take account of delivery charges or not:

Firstly, delivery charges form part of consumption expenditure and should be covered by the HICP. Secondly, representative items should take account of delivery services if the products are usually bought including delivery. For those products where delivery is currently gaining importance, like the delivery of books and CDs, it might be worth adding price representatives that account for delivery.

b) Concerning the question how the delivery charges should be covered and which COICOP groups they should be assigned to:

As a first criterion it seems to be reasonable to ask the question whether delivery is optional or obligatory.

- As a general rule, delivery charges should be covered together with the delivered product(s) where delivery is obligatory.*
- Where the consumer has the possibility to arrange for delivery himself (optional delivery) the delivery charges should be covered in COICOP 07.3.6 (08.1.0 respectively) depending on the underlying service of delivery (probably depending on the size and weight of the delivered product).*

From a practical point of view, there would not be any problem if a consumer ordered only one product. Where a consumer orders several products belonging to several COICOP groups, one would have to apportion the delivery charge across COICOP headings. The discussion in Athens showed that most MSs were not in favour of this solution.

In order to provide a practical solution also in this case, a second criterion should be considered, i.e. whether the delivery charge is paid as a proportional charge (per unit, percentage of the price of the delivered product) or a fixed charge regardless of the number of items bought.

- If the delivery charge is to be paid as a proportional charge, it should be assigned to the same COICOP group as the delivered product. The same applies to the delivery of several products if the delivery charge is proportional, in which case the delivery charge should be split accordingly.*
- If the delivery charge is a fixed charge regardless of the number of items bought, it should be assigned to either COICOP 07.3.6 or COICOP 08.1.0. This rule should always be applied, also when the delivery charge forms a large part of the total price to be paid for the full service (product plus delivery)."*

The annex gives a draft Eurostat guideline that has so far not been taken forward.

However, this is not the full extent of the issue. Even with a general principle in place there are practical issues that must be considered. One question is how to deal with sites that do not set delivery charges per item, but rather either set charges as a fixed fee for delivery up to a set number of items, and then an added fee per item, or have a scale for delivery, where the unit cost changes depending on the number ordered. Both of these cause problems for price collection - most specifically in deciding whether to cost delivery on a single item or as an average across several. For example if a bookseller charges £7.50 for the delivery of up to 3 books, then do we treat the delivery charge as £7.50 (the cost if you ordered 1 book), or £2.50 (assuming that consumers would only pay the minimum delivery charge per book). Clearly a major challenge is to obtain detailed information from retailers (or customers) on the make up of individual orders and average charges for delivery. This type of information is not readily forthcoming in the UK. The samples used for the Expenditure and Food Survey is generally too small for the purpose and Internet companies are reluctant to supply such information.

One reason for including delivery charges with the item to which they refer is to ensure that the price is on a comparable basis to that for the same good purchased at a conventional outlet. This is clearly valid for a large durable good which is normally delivered free of charge from a conventional outlet. But consider the case of a household's weekly shop at a

supermarket. The “delivery” charges associated with this kind of purchase – the cost of using a motor vehicle, or possibly taxi fares, are included elsewhere in the index, not as part of the price of the items purchased. So comparability would not be aided by allocating the delivery charge across the items purchased.

Furthermore, in cases where the consumer purchases a heterogeneous set of items simultaneously from one supplier (common with purchases of book/CDs/computer software) a decision to add the delivery charge to the item price, can cause problems. We would first need to arrive at an average basket of items, and then decide how to apportion the delivery charge (i.e. per item, or in proportion to value, or by weight). In these cases a more appropriate approach might be to produce a separate index for delivery charges, based on an average basket. Information from household budget surveys about purchasing patterns may be useful for determining how delivery charges are applied in practice, though there are survey design and response issues to be overcome.

In summary, there is no definitive answer to the question of delivery charges. At present ONS thinking, based on pragmatism and an earlier discussion held in the context of the HICP, is as follows:

- Include internet delivery charge in item price where this is also the case for items purchased through other retail outlets (e.g. furniture) and for any other large items where delivery from conventional outlets is normally “free”;
- Also include these charges for single purchases these cannot be consumed unless delivered by the retailer or their agent;
- For bulk internet purchases of a range of goods (e.g. supermarket purchases) construct separate indices for delivery charges based on typical baskets of goods. This will allow changes in delivery charge regimes to be more easily accommodated.

5.2 Differentiation between prices for consumers and for businesses

Clearly the Retail Prices Index is just that, an index based on retail prices. Therefore any prices that are only available to businesses do not fall within the scope of the index, and should be excluded from the calculation. So the question is how to identify these prices so that they can be excluded.

There is no definitive, or easy, answer to this, however we can produce several indicators that a web-site is only offering prices to businesses, though these will need to be applied with appropriate knowledge of the market. Factors considered so far are:

- Clear statements that the web-site is trade only;
- Prices that are quoted for large numbers of a good (e.g. Processor chip prices of £100 per unit, if 1000 or more purchased);
- Prices quoted without VAT;
- Web-sites for known wholesalers, who are known to exclude the public.

Of these we would need to be most careful in applying the restriction on prices quoted without VAT. For some goods, such as Personal Computers, this is standard practice, and indicates nothing more than the prices are available for both retail and small businesses.

5.3 Treatment of overseas websites

The treatment of overseas websites is a complicated issue, not least because of the difficulty in identifying whether a website is UK or based overseas, and whether the transaction is deemed to take place there, at the purchasers home or elsewhere. Originally you would be able to identify the source of a website in one of two ways, either the website address (.uk suffix means in the UK), or from the currency in which transactions take place. However this has changed, and neither of these can definitively be used to determine the location of a web company.

In the first instance website names (or addresses) can be sold, and are not regulated by geographic location. The largest example of this is the .tv suffix - which is popular with any company who has a relationship with the visual media. This is nominally the suffix related to the Island of Tuvalu, however the addresses have been sold world wide - with several companies in the UK using this suffix. Therefore identification purely on this suffix will mean that we exclude UK based websites (and potentially include overseas ones).

In the second case companies do not always restrict sales to take place in the local currency. Websites frequently ask what currency you wish to pay in, and use an on-line currency converter to produce worldwide prices from the local currency price. As these are normally rounded to produce "normal looking" prices the original currency cannot be discerned. Therefore the use of this information to establish location of a website is unreliable. Even a combination of these two sources of information does little to improve the situation, as the companies most likely to buy other Countries' web addresses, are also those most likely to be using multiple currency formats.

This leaves us in a position where it is difficult to identify whether all websites are UK or overseas. In this case we must ask the question whether this makes much difference to the price indices, or if it does, whether the definition of prices should be restricted to UK websites only or allowed to include others. The question is perhaps more pertinent to the Harmonised Index of Consumer Prices (HICP) where consistency in geographical coverage is important for the calculation of the Monetary Union HICP from the figures for individual members states. But again this is unlikely to be an issue of great consequence at current levels of online retailing.

The easy (and perhaps pragmatic) solution would be to say that it is not necessary to exclude overseas websites. This can be seen as a reasonable solution for several reasons:

- The prices are available to consumers in the UK (unlike other prices in other Countries), therefore are applicable to the UK market;
- It could be argued that the transaction took place at the purchaser's home;
- These prices are available world wide, and so would be likely to influence sales specifically in the UK, therefore price movements are unlikely to be significantly different from UK websites;

- There is no automatic link between the site of a website, and where goods are shipped. As a company can be regarded as from the UK if the shipping is done from within the UK, the overseas distinction becomes even more difficult to define.

All of these factors suggest including in the sampling frame all websites that are either clearly UK, or for which there is a possibility that they could be UK based.

5.4 Choice of sampling frame and the effect of volatility in the market

The final two areas are closely related, and so are being considered together. A major concern in the production of an Internet price index is to obtain price quotes from a representative sample of outlets. For traditional stores this is relatively straightforward, as it is possible to visit shops to determine what they actually sell (as only those goods that people purchase will be on display), and there is only a finite area to search within a given location. For the Internet this is different in several ways:

- The potential number of outlets for each good is extremely large, and will range from a small home based company selling a small number of a very limited set of items, to large supermarket chains selling many items;
- It is not easy to determine sales figures, or identify which goods are most important to the outlet;
- It is not possible to know whether a good is stock, or being advertised in the anticipation of stock becoming available - this is important as we can only price a good in the RPI if it is actually available in the store or ordered from known stock;
- There is, currently, a significant possibility of outlets disappearing without warning, causing major problems of continuity in the RPI.

These factors make the identification of a relevant sampling frame more problematical than for traditional outlets, and will lead to the need for innovative solutions.

Work on this area is not as advanced as for those listed above. A start has been made on linking the collection with the traditional selection, in that the large stores represented in the main selection, and known to have a web-outlet (e.g. Tesco, Sainsbury), would always be included in the sample. However, it is less easy to determine how to choose from web-only retailers (e.g. Amazon, Jungle). For the goods already included in the RPI (books, non-perishables at supermarkets), this is done judgementslly, using known web-retailers, and perhaps this will continue to be the best solution in the short term. However, as internet sales become more important, it will also be more important to ensure some form of random sampling in order to maintain the representativity of the index.

Concerns also extend to the current volatility in web-retailers. At the moment many sellers via the internet do not make profits, and this leads to a situation where they are always in danger of collapse, and ceasing trading. We have already experienced one example in the UK where a major online retailer has been introduced into the RPI calculation only to disappear some months later as a result of bankruptcy. We need to consider whether the effect of this

can be minimised by controlling initial sampling, and produce procedures to deal with the situation when it does arise. We will also need to consider the mechanisms by which the sample is updated.

6. Credit Card Surveys: a workable solution?

As part of its methodological development programme, ONS assisted in a pilot project on Internet shopping led by a major credit card company in partnership with a market research company. From June 2001 online diaries were kept by a sample of about 800 UK internet users recording detailed online purchasing behaviour including cost, nature of item, site purchased from, delivery cost and method of payment. This was designed to provide, at the very least, a useful indication of Internet sales volumes but its ultimate purpose was to compile an “E-tail Price Index” for internet purchases gathering data on online purchasers and the value and volume of the goods they obtain online.

6.1 Methodology

The survey consisted of a number of stages:

- *Recruitment of respondent panel*

A number of potential methods were investigated:

- Recontacting respondents from a current Internet User Profile Survey (IUPS). This consists of a telephone survey of 1600 internet users every six months on internet behaviour and associated attitudes. Those that indicate that would be willing to be re-contacted for further projects would provide a contact database and allow ready access respondents for whom data on shopping frequency and behaviour is already available. However, this source was discounted for two reasons: the small sample size and the six-month timelag. The latter can be problematic and an area that can be fast changing. 2001 results from this survey, based on recollection rather than on the keeping of a diary, suggested that very approximately 2% of the estimated £650 billion total retail spend in the UK is online.
- Utilising an existing online panel where email addresses are gathered on an ongoing basis from respondents completing a survey known as Telebus (telephone omnibus survey). All these respondents will have agreed to be contacted via e-mail for the purpose of completing online surveys. The sample has the advantage both of size (11,000 respondents) and of being demographically representative of internet users but again suffers from out-of-dateness with some respondents having been recruited over a year ago.
- Using pre-screened contacts from a list broker such as Experian, who compile targeted lists from the results of “lifestyle” surveys. Such sources of more often than not less up-to-date than the market research company’s own survey lists. It was also felt not to be so cost-effective, perhaps requiring lists from more than one source to produce a representative and large enough survey.
- Random-digit-dialling (RDD) where people are contacted from a random list of telephone numbers and asked if they would be willing to take part in research on internet access and purchasing behaviour. Experience suggests that only about 40% of numbers contacted are “live” residential households.

In the event RDD was used on the basis that it was most likely to generate a truly random of online shoppers.

- *The telephone interview*

During the initial “recruitment” telephone interview respondents were asked:

- Whether they use the internet
- Whether they *ever* purchase online
- Details of past online shopping experiences.
- Expectations of future shopping experiences.
- Demographic details.

The recruitment targets for the “panel” of online shoppers were:

- 400 who had purchased online in the past three months.
- 50 who had purchased online longer ago.
- 50 who had not purchased online but intended to do so in the next three months.

It was recognised that this would over-represent heavier online purchasers but it was also important to maximise the volume of sales data gathered. The results would be re-weighted at the analysis stage to the true incidence in the target population using information from the previously mentioned IUPS. Quotas were also set for age, gender and working status.

- *The Diary*

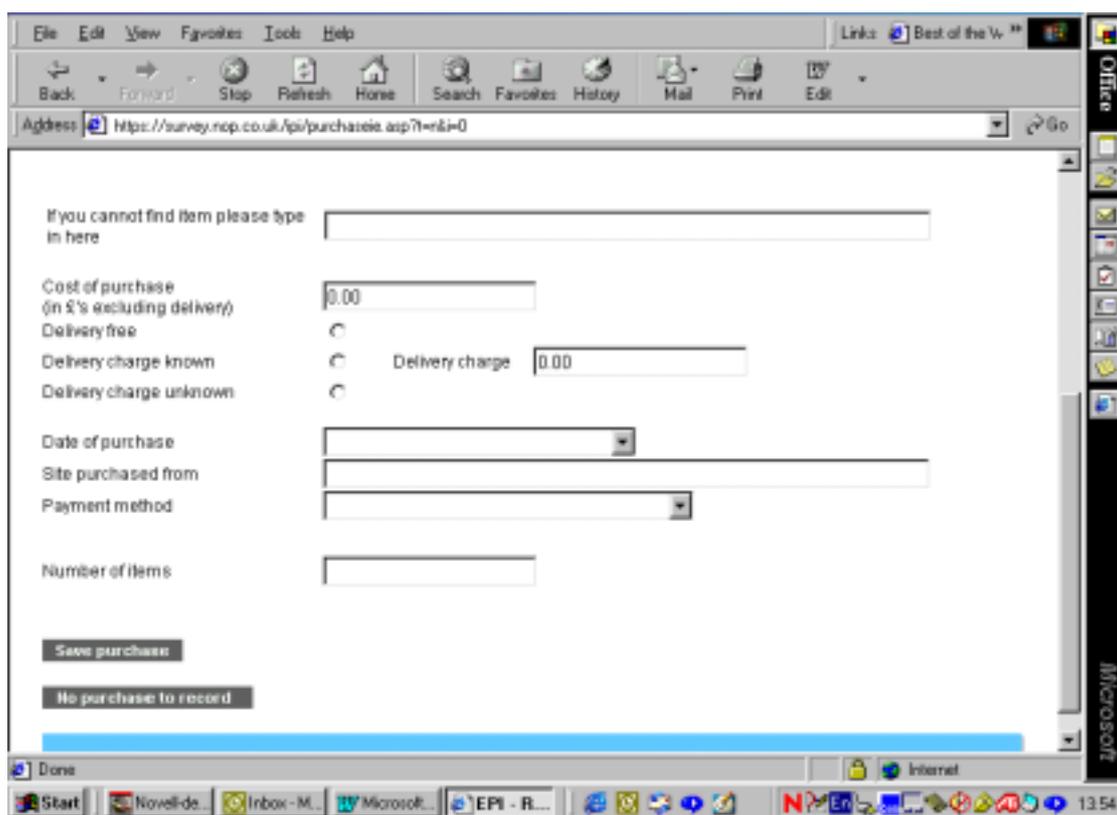
Respondents were asked to complete online diaries- given that they would all have online access this seemed to be the most sensible and cost-effective approach for everybody including the respondents themselves. It also allowed real-time data monitoring by market research staff and also respondents to monitor their own responses (allowing them to quality assure their inputs more effectively). In addition as the diary is web-based, respondents were able to record purchases from whichever terminal the purchase was made, e.g. at home or at the office. The cost of data cleaning was much reduced by the online approach e.g. querying online in real time odd responses. Incentive payments were also made.

A copy of the diary page is given below.

6.2 Evaluation of pilot

Despite the hard efforts of all involved the response rate was very disappointing. The assumption at the planning stage was that 35% of recruited respondents would actually proceed to the diary completion stage in the first month. In the event, whilst the data collected was of good quality, the response rate was much lower than expected and the attrition rate high in the subsequent months. In consequence, the survey did not proceed past the pilot stage.

The question arises, whether the survey design could have been improved to produce satisfactory data. The alternative option of using the Telebus Survey was considered but in the event this was not pursued.



7. Other options

Obtaining price and sales data direct from online retailers is the main alternative option that the ONS has not pursued. This may well provide a partial solution in the longer-term or at least help to plug some of the current data gaps, but there are a number of measurement problems to be resolved even if companies agreed to co-operate in principle. These include those confronted already in our attempts to get sales data, such as the difficult of finding a reliable sampling frame and distinguishing sales to private households from sales to other commercial companies. Indeed, there may be a lot to be learnt from experiences gained in the use of scanner data. Both are administrative sources but there doesn't appear at present to be a company that is actively engaged in bringing together into one database information on online retail purchasing.

8. Conclusions

The ONS has been expanding the scope of price collection to incorporate into the Retail Prices Index (and the Harmonised Index of Consumer Prices) goods and services sold on the Internet. A strategy is in place for identifying which items should be included. Sales information is obtained from the Expenditure and Food Survey, supplemented by other market research information, and prices are obtained from the corresponding websites. But there are a number of conceptual and data collection issues that need to be resolved and that would benefit from the sharing of national experiences. It is also an area that is lacking internationally agreed guidelines but that will be increasing in importance and greater use is made of online sales in the future.

Annex: Eurostat proposals on the coverage of delivery charges

Purchase of	Delivery
	<i>Obligatory/Optional Proportional charge/Fixed charge</i>
Individual product	Same COICOP as delivered product(s) COICOP 07.3.6 or 08.1.0 Weighting information is needed. COICOP 07.3.6 or 08.1.01
Several products, single COICOP group	
Several products, several COICOP groups	Apportion delivery charge to the COICOP groups the delivered products belong to

For information: Extract from Commission Regulation (EC) No 1749/1999

COICOP 07.3.6 Other purchased transport services (S)

- funicular, cable-car and chair-lift transport,
- removal and storage services,
- services of porters and left-luggage and luggage-forwarding offices,
- travel agents' commissions, if separately priced.

Excludes: cable-car and chair-lift transport at ski resorts and holiday centres (09.4.1).

COICOP 08.1.0 Postal services (S)

- payments for the delivery of letters, postcards and parcels,
- private mail and parcel delivery.

Includes: all purchases of new postage stamps, pre-franked postcards and aerogrammes.

Excludes: purchase of used or cancelled postage stamps (09.3.1); financial services of post offices (12.6.2).