Review of Income and Wealth
Series 52, Number 4, December 2006

ACCOUNTING FOR THE EFFECTS OF NEW AND DISAPPEARING GOODS USING SCANNER DATA

BY DANIEL MELSER*

Monash University, Melbourne

With the "discovery" of scanner data by statistical agencies and researchers comes a wealth of new information upon which price index calculations can be based. Old problems, such as the appearance and disappearance of goods over time, are likely to be an important feature of such data. However, given that scanner data includes the prices and quantities of the population of transactions we have more information than is traditionally available to deal with the new and disappearing goods problem. We adopt a recently developed approach using the Constant Elasticity of Substitution cost function to provide a detailed empirical analysis of the effects of new and disappearing goods for an Australian scanner data set of supermarket products. Our results indicate that the failure to account for new and disappearing goods in the cost-of-living index leads to a significant upward bias.

1. INTRODUCTION

In this article we discuss and quantify the effects of new and disappearing goods on the cost-of-living index. This is done in the information-rich context of scanner data (sometimes called barcode or point-of-sale data). The availability of this new data source has the potential to greatly improve the way price change is measured as scanner data records the *population* of sales of items in a given store over a given time period. This means that both price and quantity data is available to index practitioners often at a very disaggregated level. This has led many authors to emphasize the advantages of scanner data over the data that is conventionally used by statistical agencies to compute price indexes (Diewert, 1993; Silver, 1995; Bradley *et al.*, 1997; Dalén, 1997; Richardson, 2000; Schut, 2002; Silver and Webb, 2002).

As well as being of great benefit for the compilation of official statistics, scanner data can also be used to investigate enduring economic problems associated with index numbers. One such problem is the effect of new and disappearing goods on price indexes. In this article we undertake a detailed empirical investigation, using a large scanner data set, of the effects of non-matched goods on the cost-of-living index. The fact that we have both price and quantity data at a

Note: I am grateful to the Australian Bureau of Statistics who generously provided the data for this project. Also much appreciated were comments from Bert Balk, Robert Hill, Kevin Fox, Lorraine Ivancic, Samara Zeitsch, Carmet Schwartz and Iqbal Syed as well as participants at the UNSW CAER Conference in December 2003 and the SSHRC Conference in Vancouver in 2004. I am also very appreciative of the efforts of the editor of this journal and gratefully acknowledge the comments received from two anonymous referees on the article.

*Correspondence to: Daniel Melser, Department of Economics, Caulfield Campus, Monash University, 900 Dandenong Road, Caulfield East, VIC 3145, Australia (daniel.melser@buseco.monash.edu.au).

disaggregated level allows us to accurately estimate the differences between indexes which properly account for the effects of new and disappearing goods from those which do not.

In the next section we discuss the basic problem of quality change and new and disappearing goods in the context of the "supermarket products" which are the focus of this study. Little research has been undertaken on quantifying the effects of new and disappearing goods on this product area. However, a number of methods have been suggested in the economics literature to account for non-matched items. We briefly discuss three main approaches; estimating reservation prices, hedonic regression and an approach using the Constant Elasticity of Substitution (CES) cost function. We primarily use the last of these methods, which is outlined in greater detail in Section 3. As well as this primary approach we use a simplified version of the reservation price method to provide a cross-check on our results. In Section 4 we apply these two methods to a large scanner data set and discuss the major results. Section 5 concludes.

2. The Quality Change and New and Disappearing Goods Problem

One of the enduring problems of economic measurement is how to deal with changes in the quality and availability of goods over time. In fact this debate ranges back at least to Alfred Marshall in 1887 (Marshall, 1887, p. 209) who advocated the use of chained indexes to mitigate the effects of new and disappearing goods.

From an economic perspective the ideal measure of price change is the cost-of-living index which compares the minimum cost of obtaining a given level of utility under two price regimes. If there are differences in the quality or availability of goods under the two price regimes then this has an effect on utility which must be accounted for in the cost-of-living index (Gordon and Griliches, 1997). Given this goal of economic price measurement, it seems important to have an idea of the influence of new and disappearing goods on welfare. In this article we advance empirical research in this area.

2.1. *Estimates of the Bias from New and Disappearing Goods*

The most comprehensive project quantifying the biases in official price indexes was that undertaken by the Boskin Commission (Boskin *et al.*, 1996; Gordon and Griliches, 1997) who looked at the US Consumer Price Index (CPI). The Boskin Commission estimated that quality change and new goods constituted the largest source of bias in the US CPI. In total they estimated that the US CPI was overestimated by 0.6 percentage points in 1996 due to the failure to adequately account for quality change and new goods.

In this article we focus on one particular area of the CPI. We look at what we term "supermarket products," in particular: *Biscuits*, *Bread*, *Butter*, *Cereal*, *Coffee*, *Detergent*, *Frozen Peas*, *Honey*, *Jams*, *Juices*, *Margarine*, *Oil*, *Pasta*, *Pet Food*, *Soft Drinks*, *Spreads*, *Sugar*, *Tinned Tomatoes* and *Toilet Paper*. These products provide a selection of the goods available in supermarkets and mainly comprise "processed food" products. The Boskin Commission did not look at this product area in particular detail; however, they concluded that the "Food at Home other than Produce" category, which covers most of the products above, had an annual upward bias of 0.3 percentage points from 1967 to 1996. The justification given for this bias estimate by Boskin *et al.* (1996) is interesting. They write:

> How much would a consumer pay to have the privilege of choosing from the variety of items available in today's supermarket instead of being constrained to the much more limited variety available 30 years ago? A conservative estimate of the value of extra variety and convenience might be 10 percent [approx. 0.3 percent annualized] for food consumed at home other than produce . . .

The noteworthy aspect of the quote from the Boskin Commission is that the primary reason they give for the upward bias of official indexes is the failure to properly account for change in the *variety* of available products. What is important then is the fact that the range of products available in supermarkets has increased substantially over recent decades. As noted by Koskimäki and Ylä-Jarkko (2003, p. 11), this increase in the range of products is likely to be a consequence of monopolistic competitors endeavoring to produce differentiated products so that substitution occurs within brands rather than between brands. The result of this behavior is that an increasingly large set of niche-marketed products are available to consumers, which has an influence on their welfare and cost-of-living. Hausman (2003, p. 28) called this the "invisible hand of imperfect competition." In the following sections we briefly discuss various ways of measuring these effects.

2.2. *Estimating Reservation Prices*

A diverse range of approaches have appeared in the economics literature for dealing with new and disappearing goods. The classic approach to the problem is derived from Hicks (1940) who saw it as one of missing prices. His solution for *new goods* was to estimate the reservation (or choke) price which would have driven demand for the good to zero in the period prior to its introduction. The reservation price can be used either in a conventional price index framework, or in a parametric framework, to look at the effect on welfare of the introduction of the good. An analogous approach can be used for disappeared goods.

This "reservation price" method is very appealing and has a rigorous economic justification. Hausman (1997) adopts this approach and econometrically estimates a demand system for the introduction of a new brand of cereal in the U.S. Hausman (1997) finds that the official price index for cereals was too high by 20–25 percent due to the effect of new brands.[1]

While this approach is attractive it has the major disadvantage that it is technically very difficult to implement, involving complex econometric estimation. These estimation methods are also contentious and as emphasized by Bresnahan (1997), the discussant on Hausman's (1997) article, the assumptions made in motivating the estimation can be important in influencing the results. This has led

[1]This means that if the price change was 2.0 percent per year then the index would be overestimated by 0.4–0.5 percentage points per year.

to some suspicion of this approach. For example, the recent National Research Council report, *At What Price?* (Schultze and Mackie, 2002, p. 159) noted that:

> . . . there is no clearly acceptable technique for consistently estimating demand curves for new goods or services in such a way that choke prices can be confidently ascertained.

It appears that at present this method is quite controversial and not widely accepted. For this reason we will not adopt this version of the approach.

Recently, however, Hausman (2003) has suggested an alternative *approximate* reservation price method. This method is far simpler than the full econometric method and requires only the estimation of the price elasticity of demand, $\varepsilon_i^t \equiv -d\ln\left(x_i^t\right)/d\ln\left(p_i^t\right) = -\left(dx_i^t/dp_i^t\right)\left(p_i^t/x_i^t\right)$, where p_i^t and x_i^t are the price and quantity consumed of good i in period t. Here, instead of using the compensated demand curve which is the theoretically correct approach, we take a linear approximation to the market demand curve, i.e. $p_i^t = \hat{p}_i^t + \left(dp_i^t/dx_i^t\right)x_i^t$, where $\hat{p}_i^t$ is the estimated reservation price. It can easily be seen that, by substituting ε_i^t into this relationship and manipulating, we can estimate the reservation price using the following formula:

$$\hat{p}_i^t = p_i^t\left(1 + 1/\varepsilon_i^t\right). \tag{1}$$

Hausman (2003, p. 27) argues that this estimate of the reservation price provides a reasonable approximation; however, as we typically expect the demand curve to be convex to the origin then (1) will underestimate the true reservation price. In the empirical section we apply this method.

2.3. *Hedonic Regression*

Another popular approach to dealing with changing varieties of products is hedonic regression. The hedonic approach regards goods as being "packages" of various utility-yielding characteristics which determine price. A hedonic regression exploits the market relationship between the prices and characteristics of the good (Rosen, 1974). This approach is useful as it is often the case that the characteristics of a good are more stable across time than the various varieties (i.e. bundles of characteristics). The hedonic function can be used to estimate the price of a good for any particular combination of characteristics and hence there are a number of ways in which it can be used to calculate price indexes (Silver, 1999; Diewert, 2003).

Hedonic methods have most frequently been applied to areas where prices have changed rapidly due to technological factors, such as computers (Berndt *et al.*, 1995; Berndt and Rappaport, 2001). It has not (to the best of my knowledge) been applied to supermarket commodities like those listed above. The reason for this is that hedonic methods do not measure the effects of changes in variety. But this is just the aspect of the problem we are interested in, as emphasized by the Boskin Commission.

The hedonic regression approach to quality change and new and disappearing goods focuses entirely on how changes in prices relate to changes in characteristics where the characteristics are stable across time. However, in our case, as emphasized by the quote from the Boskin Commission above, it is not a problem of accounting for *improvements* in the characteristics of products but rather one of accounting for the expansion in the *range* of available characteristics or the way a given set of characteristics are configured. Hedonic methods as presently constituted are not able to reflect these changes. To see this consider a case where prices for different varieties and characteristics are unchanging through time but an ever expanding variety of characteristics configurations is available. As long as some of these new varieties are desirable then the cost-of-living index should fall even though prices for the characteristics have not changed. The hedonic method will clearly not account for these changes. For this reason we will not explore this method further here and will instead turn to our primary method of accounting for new and disappearing goods. We outline this approach in more detail in the following section.

3. The CES Cost-of-Living Index with New and Disappearing Goods

In this article we adopt a method of more recent vintage than the two discussed above. This method was initially proposed by Feenstra (1994), and developed, extended and refined by Nahm (1998) and Balk (1999). It is able to rigorously account for the effects of new and disappearing goods in a relatively simple framework. There have been only limited applications of this approach in the literature (Feenstra and Shiells, 1994; de Haan, 2001; Opperdoes, 2001). Let us outline this method.

We consider the case of two periods, $t = 0,1$, where we denote the index set of goods available in each period by I^0 and I^1. We will also make use of the index set of goods which are common to both periods, $I^{1,0} \equiv I^1 \cap I^0$, $\bar{U}$ is some reference utility level, p^1 and p^0 are the price vectors, b_i are quality or taste parameters in the consumers utility function and σ is the elasticity of substitution, $\sigma \equiv -d\ln\left(x_i^t/x_j^t\right)/d\ln\left(p_i^t/p_j^t\right)$ for some goods i and j. The elasticity of substitution is the same for each good and represents the extent to which consumers change their relative consumption of goods as relative prices change. It must be non-negative in order for consumers' (compensated) demand curves not to slope upwards.[2] With this terminology we can introduce the CES cost function over a changing domain of goods.

(2) $$C\left(p^t, \bar{U} \middle| I^t\right) = \left(\sum_{i \in I^t} b_i \left(p_i^t\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \bar{U}, \quad t = 0, 1.$$

Note that when we adopt (2) the cost-of-living index will reflect not only price change but also changes in the availability of goods or consumption opportunities represented by I^0 and I^1. What is important is that we can represent the cost-of-living index exactly for the CES cost function over a changing set of goods. As

[2]To see this consult the Appendix.

Feenstra (1994) and Balk (1999) demonstrated, the cost-of-living index has the following form where $\hat{P}$ is a price index over matched goods, $I^{1,0}$.[3]

$$(3)\qquad \frac{C(p^1,\bar{U}|I^1)}{C(p^0,\bar{U}|I^0)} = \left(\frac{1-\sum_{i\in I^1, i\notin I^{1,0}} s_i^1}{1-\sum_{i\in I^0, i\notin I^{1,0}} s_i^0} \right)^{\frac{-1}{1-\sigma}} \times \hat{P}, \quad s_i^t \equiv \frac{p_i^t x_i^t}{\sum_{i\in I^t} p_i^t x_i^t}, \quad t=0,1.$$

The form of the cost-of-living index is relatively straightforward; it is calculated as a matched-goods price index, $\hat{P}$, which is multiplied by a factor reflecting relative expenditure on new and disappeared goods, and the elasticity of substitution. The intuitive explanation for the form of the adjustment factor is that the expenditure shares for new and disappeared goods reflect their *importance* to consumers. The adjustment factor then compares the relative gain from new goods and the loss from disappeared goods and adjusts this ratio using the elasticity of substitution. It is interesting to note that no adjustment to the matched-goods price index occurs when, either, the expenditure shares on new and disappeared goods are equal, indicating that relative gains in consumption opportunities were equivalent to the losses, or as $\sigma \to \infty$, in which case all goods are very close substitutes and whether new goods appear or old goods disappear does not matter in terms of consumption opportunities.

Balk (1999) showed that the matched-goods price index, $\hat{P}$, has various representations. We will use three of these price indexes below. The first representation of $\hat{P}$ is the well known Lloyd–Moulton or Base-Period Weighted Price Index (P^{BW}).

$$(4)\qquad \hat{P} = \left(\sum_{i\in I^{1,0}} \hat{s}_i^0 \left(\frac{p_i^1}{p_i^0} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \equiv P^{BW}, \quad \hat{s}_i^t \equiv \frac{p_i^t x_i^t}{\sum_{i\in I^{1,0}} p_i^t x_i^t}, \quad t=0,1.$$

This index dates back to Lloyd (1975) and has attracted attention for its ability to reflect consumers' substitution behavior while only requiring knowledge of base period expenditure shares.[4] It can be seen that the Lloyd–Moulton Index is equal to the Laspeyres Index when $\sigma = 0$. Furthermore, if we regard the Lloyd–Moulton Index as a function of independent variables then it can be shown that it is monotonically decreasing in σ (Hardy *et al.*, 1952, p. 26, Th. 16) and by appropriate choice of σ it can produce any number between and including the maximum and minimum price relatives.

The second representation of $\hat{P}$ that we use is the equivalent current-period weighted expression to (4). We call this index the Current-Period Weighted Price

[3]See the Appendix for a derivation of this result, and those that follow, and a slightly different representation of the cost-of-living index.

[4]This index was used by Shapiro and Wilcox (1997), for U.S. data, to investigate whether a real-time price index could be calculated which reflected substitution. Shapiro and Wilcox (1997, p. 123) concluded that using the Lloyd–Moulton Index, " . . . it is possible to produce an approximation to the Törnqvist Index that is both feasible in real time and quite accurate."

Index (P^{CW}), again first discussed by Lloyd (1975) and defined in (5) below. Note that when $\sigma = 0$ this index is equal to the Paasche Price Index.

$$(5) \qquad \hat{P} = \left(\sum_{i \in I^{1,0}} \hat{s}_i^1 \left(\frac{p_i^1}{p_i^0} \right)^{-(1-\sigma)} \right)^{\frac{-1}{1-\sigma}} \equiv P^{CW}.$$

Finally, the Sato–Vartia Price Index (P^{SV}) can also be derived from the CES functional form and is shown in (6). The weights for the Sato–Vartia Price Index are rather complex and involve the normalized logarithmic mean of the expenditure shares in each period.[5]

$$(6) \qquad \hat{P} = \prod_{i \in I^{1,0}} \left(\frac{p_i^1}{p_i^0} \right)^{\hat{w}_i^{1,0}} \equiv P^{SV}, \quad \hat{w}_i^{1,0} = \frac{L(\hat{s}_i^1, \hat{s}_i^0)}{\sum_{i \in I^{1,0}} L(\hat{s}_i^1, \hat{s}_i^0)}.$$

Interestingly, note that the Sato–Vartia Price Index does not depend on the elasticity of substitution, σ. This is important for later purposes.

3.1. *A Restriction on the Elasticity of Substitution*

A vital point to note regarding this approach to calculating the cost-of-living index is that the elasticity of substitution must be greater than one, $\sigma > 1$, for this approach to be reasonable. Balk (1999) showed this by considering an example where $p_i^1 = p_i^0 \ \forall i \in I^{1,0}$ and where we have some newly appeared goods but no disappearing goods. Then using (3), and noting that under these assumptions the matched price index will equal one, the cost-of-living index for this particular case is shown below.

$$(7) \qquad \frac{C(p^1, \bar{U} | I^1)}{C(p^0, \bar{U} | I^0)} = \left(1 - \sum_{i \in I^1, i \notin I^{1,0}} s_i^1 \right)^{\frac{-1}{1-\sigma}}.$$

But this index must be no larger than 1 as the consumer now has a greater range of goods to choose from. It can be seen that this implies that we must have $\sigma > 1$.

Why do we have this restriction on the elasticity of substitution? Consider the following optimization argument. The cost function, by definition, is the minimum expenditure required to achieve a given level of utility. However, in looking at the effect of new and disappearing goods we are defining a *restricted* cost function where the consumption of some goods is constrained to zero in some periods. We can then write the modified cost function for period $t = 0,1$ in the following way:

[5]The logarithmic mean $L(a,b)$ is defined as $L(a,b) = (a - b)/(\ln a - \ln b)$ where a $\neq$ b and $L(a,a) = a$. Clearly we must have $a,b > 0$.

$$C\left(p^t,\bar{U}\middle|I^t\right)=\min{}_{x}\sum_{i\in I^1\cup I^0}p_i^t x_i \tag{8}$$
$$s.t.\quad U(x)\geq\bar{U}$$
$$x_i=0,\forall i\notin I^t.$$

However, this definition may cause problems if there are some goods i which are essential to consumption but are not common to both periods (i.e. where $i\notin I^{1,0}$). In this case it may be impossible to reach the reference utility level without some consumption of these goods and the constraints in the optimization problem may define a feasible set which is empty. To see that this is indeed the case for the CES functional form we can derive the CES utility function which is dual to the CES cost function.

$$U\left(x^t\middle|I^t\right)=\left(\sum_{i\in I^t}b_i^{\frac{1}{\sigma}}\left(x_i^t\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}},\quad t=1,0. \tag{9}$$

From inspection of (9) we can see that if $\sigma\leq 1$ then every good is essential to consumption—utility explodes as $x_i^t\to 0$. This was noted by Feenstra (1994). It is only when $\sigma>1$ that the consumption of a good can equal zero without utility being either undefined or zero.

Economically this means that if $\sigma>1$ then consumers can be compensated for the restricted (zero) consumption of some goods by increases in the consumption of other goods. This ability to compensate the consumer for the loss of some goods is vital in obtaining sensible answers to the effect of new and disappearing goods on the cost-of-living. If no compensation is possible then the cost-of-living index will be infinite if one of these "essential" goods is lost. It seems reasonable that at the elementary level of aggregation, where we will apply this theory, that all goods are effectively replaceable. This is clearly not so plausible at higher levels of aggregation. Consider for example the goods "food" and "clothing." Clearly if our consumption of these goods were restricted to zero then this certainly would be catastrophic for utility.

3.2. *Estimating the Elasticity of Substitution*

As the adjustment for new and disappearing goods shown in (3) depends on the elasticity of substitution we need to estimate this parameter to implement this approach in practice. Fortunately Balk (1999) outlined various ways in which the elasticity of substitution could be easily estimated. The basic idea of his approach is that all the CES matched-goods price indexes, (4)–(6), should be equal. This gives us three methods for estimating the elasticity of substitution. Because estimation is undertaken at the aggregate level, no sampling error is produced as there is a single unique solution to each equation.[6]

[6]The estimate from each method differs because they use different information (i.e. just as estimates would differ in regression estimation if we excluded various observations).

The first method used to obtain $\hat{\sigma}$, an estimate of σ, is to find the value of $\hat{\sigma}$ which makes the Base-Period and Current-Period Weighted Price Indexes equal as in (10) below.

$$(10)\qquad \hat{P}^{BW} \equiv \left(\sum_{i \in I^{1,0}} \hat{s}_i^0 \left(\frac{p_i^1}{p_i^0} \right)^{1-\hat{\sigma}} \right)^{\frac{1}{1-\hat{\sigma}}} = \left(\sum_{i \in I^{1,0}} \hat{s}_i^1 \left(\frac{p_i^1}{p_i^0} \right)^{-(1-\hat{\sigma})} \right)^{\frac{-1}{1-\hat{\sigma}}} \equiv \hat{P}^{CW}.$$

We will call this approach the *Current v Base* method. It is particularly appealing as $\hat{\sigma}$ will be positive as long as the Laspeyres Index exceeds the Paasche Index. This is because, when $\hat{\sigma} = 0$, the LHS of (10) is equal to the Laspeyres Index while the RHS is equal to the Paasche Index. To lower the LHS and raise the RHS of (10) we increase $\hat{\sigma}$ (Hardy *et al.*, 1952, p. 26, Th. 16) until equality is obtained.

The second and third methods suggested by Balk (1999) are to equate the Base and Current-Period Weighted Price Indexes, which both include the elasticity parameter, to the Sato–Vartia Price Index, which is independent of the elasticity of substitution.

$$(11)\qquad \hat{P}^{BW} \equiv \left(\sum_{i \in I^{1,0}} \hat{s}_i^0 \left(\frac{p_i^1}{p_i^0} \right)^{1-\hat{\sigma}} \right)^{\frac{1}{1-\hat{\sigma}}} = P^{SV}$$

$$(12)\qquad \hat{P}^{CW} \equiv \left(\sum_{i \in I^{1,0}} \hat{s}_i^1 \left(\frac{p_i^1}{p_i^0} \right)^{-(1-\hat{\sigma})} \right)^{\frac{1}{1-\hat{\sigma}}} = P^{SV}.$$

Though unlikely, these two methods could potentially produce a negative estimate of σ even when the Laspeyres Index is greater than the Paasche Index as the Laspeyres and Paasche Indexes do not bound the Sato–Vartia Index.

To see how these three methods are related, consider Figure 1 which depicts the three indexes in σ-space.

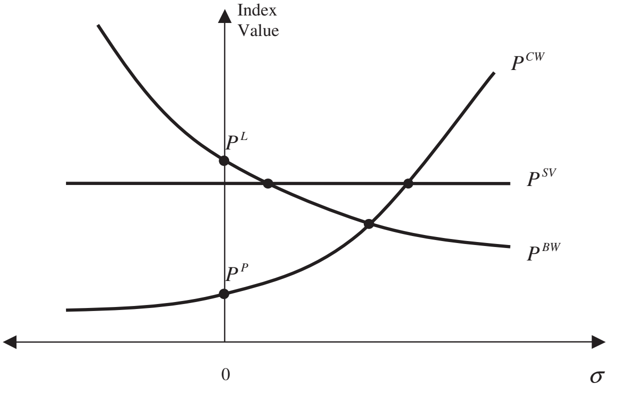

Figure 1. Estimation of the Elasticity of Substitution

Interestingly, it can be seen that the two methods that use the Sato–Vartia Index give an estimate of σ which lies either side of that from the *Current v Base* method. Hence, it seems advisable to take an average of the two Sato–Vartia methods. However, the form of the average may influence the resulting estimate of σ. For this reason, and the fact that $\hat{\sigma}$ will be positive as long as the Laspeyres Index exceeds the Paasche Index, we prefer the *Current v Base* method, though we will consider both in the empirical section that follows.

4. An Empirical Application Using Scanner Data

Now that we have discussed the theory surrounding our primary approach to new and disappearing goods we can proceed to the application of these ideas to our scanner data set. The data used in this study was purchased by the Australian Bureau of Statistics (ABS) for the purpose of investigating the use of scanner data in the Australian CPI. The data set includes observations from the start of February 1997 to the end of April 1998, 65 weeks of data in total, for 19 product categories as listed above: *Biscuits*, *Bread*, *Butter*, *Cereal*, *Coffee*, *Detergent*, *Frozen Peas*, *Honey*, *Jams*, *Juices*, *Margarine*, *Oil*, *Pasta*, *Pet Food*, *Soft Drinks*, *Spreads*, *Sugar*, *Tinned Tomatoes* and *Toilet Paper*. These products represent a selection of the goods available in supermarkets and mainly comprise processed food items. The data set includes 100 stores belonging to four supermarket chains in one of the major cities of Australia.[7] These stores accounted for around 80 percent of grocery sales in this city (Jain and Caddy, 2001, p. 4). The total value of sales for these products over the 65 week period was just over AU$600 million.

4.1. *Aggregation Methods and Other Issues*

The basic form of the data was weekly unit-value prices, and the corresponding sales volume, for a *product code* in an *outlet*.[8] In order to ensure the robustness of the results, various aggregation approaches were applied to derive the prices and quantities to be used in the index formulae. This is in the context of much research on scanner data which has shown that the method of aggregation is often very important to the results (Dalén, 1997; Reinsdorf, 1999; Jain and Caddy, 2001; Silver and Webb, 2002; Triplett, 2003; Koskimäki and Ylä-Jarkko, 2003). Note also that in applying the CES cost function approach we require each of the goods to be substitutes for one another. This makes it essential that we do not aggregate across products, say biscuits and coffee, which may be complements, but instead examine each set of goods separately.

[7]Note that while we know that the stores were from four supermarket chains, for commercial sensitivity reasons we were not able to determine which supermarket came from which chain. This reduced the range of aggregation approaches that could be pursued.

[8]Formally the *product code* is called the Australian Product Number (APN) which is the equivalent of the Universal Product Code (UPC) in the U.S. or the European Article Number (EAN).

Both *quarterly* and *monthly* aggregation of the weekly unit values was undertaken.[9] Additionally, we used both unique *product code* and *outlet* combinations as the definition of a good as well as aggregating *product codes* across *outlets*. It was anticipated that this latter approach would increase the extent of matching and mitigate the effects of new and disappearing goods (Reinsdorf, 1999, p. 153). These various aggregation methods give four different approaches in total. At a monthly frequency we have *Month (Prod. Code, Outlet)*, which uses unique *product code* and *outlet* combinations and *Month (Prod. Code)* which uses only the *product code* as the definition of a good and aggregates over *outlets*. The corresponding quarterly indexes are *Quarter (Prod. Code, Outlet)* and *Quarter (Prod. Code)*. For each of these aggregation methods we implement the approach above using chained indexes. The primary reason for this is that chained indexes are more likely to mitigate the effects of new and disappearing goods because there is greater overlap in the products available for time-periods which are adjacent than those that are more distant.[10] With these details out of the way we can proceed to the results of the empirical application. We start by discussing the estimation of the elasticity of substitution.

4.2. *Estimating the Elasticity of Substitution*

The results of the estimation of the elasticity of substitution are shown in Table 1. We focus on the *Current v Base* method with the arithmetic mean and standard deviation of the estimated elasticity of substitution shown for each product category and aggregation method. However, what is interesting is that the difference between the *Current v Base* method and the *Average Sato–Vartia* method is relatively minor. This can be seen in the second part of Table 1 showing the average of absolute deviations between these two methods. These differences are small compared with the volatility of the elasticity across time represented by the standard deviations of the estimates.

One interesting aspect of estimating the elasticity of substitution is the effect of aggregation. As we would have expected a priori, when we increase the level of aggregation the elasticity of substitution falls. What is notable, however, is that aggregation across time, from monthly to quarterly indexes, led to a far greater reduction in the elasticity than did aggregation across outlets—indicating that shifts in purchases between outlets is more important than substitution across time. In a somewhat contradictory result a larger number of *negative*, and hence implausible, estimates of the elasticity of substitution occurred when we aggregated across outlets. We now move onto the effects on new and disappearing

[9]Producing price indexes at a quarterly and monthly frequency is standard international practice and it is consistent with the new ILO CPI Manual (ILO, 2004, p. 358): " . . . it is recommended that the index number time period be at least 4 weeks or a month." It should be noted, however, that one of the potential benefits of scanner data is that price indexes could be calculated more frequently. Feenstra and Shapiro (2003) have made some progress in this regard.

[10]This is emphasized in ILO (2004, pp. 129–30). However, it should be noted that if there are seasonal products then adjacent periods may not be more similar than say periods separated by a year. Fortunately our data set does not include any products with strong seasonal patterns. Also, note that the use of chaining in high frequency data can lead to chain drift if prices "bounce" (Szulc, 1983; Feenstra and Shapiro, 2003). While this is important for the level of the indexes it is not likely to be vital in isolating the effect of new and disappearing goods on the cost of living index.

TABLE 1

Estimating the Elasticity of Substitution

Aggregation Method	Current v Base Method Average and Standard Deviation of Estimated Elasticity: Month (Prod. Code, Outlet)		Month (Prod. Code)		Quarter (Prod. Code, Outlet)		Quarter (Prod. Code)		Average Sato–Vartia Method Average Absolute Deviations from Current v Base Method: Month (Prod. Code, Outlet)	Month (Prod. Code)	Quarter (Prod. Code, Outlet)	Quarter (Prod. Code)
Statistic	Avg.	Std. Dev.	Avg.	Std. Dev.	Avg.	Std. Dev.	Avg.	Std. Dev.				
Biscuits	3.46*	0.99	2.96+	1.63	2.63*	1.02	1.63+	2.77	0.05	0.06	0.16	0.18
Bread	3.39	0.69	3.58+	1.05	2.30	0.42	2.45	0.55	0.02	0.05	0.05	0.04
Butter	4.01	0.45	3.87	0.94	3.43	0.76	3.70	1.05	0.07	0.05	0.05	0.10
Cereal	3.25	0.23	2.70	0.53	2.50	0.20	2.10	0.43	0.03	0.03	0.04	0.08
Coffee	5.66	0.46	4.66	0.65	3.73	0.41	2.63	0.46	0.06	0.02	0.03	0.02
Detergent	3.88	0.34	3.41	0.45	2.18	0.13	1.35*	0.49	0.04	0.03	0.06	0.08
Frozen Peas	3.44	0.30	3.36	0.58	2.28	0.46	1.68*	0.85	0.02	0.02	0.01	0.04
Honey	4.04	0.79	4.31	2.10	3.03	0.85	2.13+	2.20	0.08	0.07	0.03	0.11
Jams	3.17	0.58	2.72+	1.17	1.65	0.47	0.70*	0.91	0.06	0.04	0.07	0.19
Juices	3.14	0.32	3.38	0.56	2.53	0.30	2.50	0.36	0.05	0.03	0.08	0.10
Margarine	4.63	0.42	3.99	0.61	3.63	0.39	3.60	0.75	0.03	0.02	0.01	0.03
Oil	4.59	0.71	5.47	0.83	3.13	0.38	2.88	0.51	0.20	0.52	0.04	0.03
Pasta	2.16	0.32	2.51	0.61	1.43	0.17	1.00*	0.84	0.02	0.03	0.03	0.06
Pet Food	3.64	0.35	3.29	0.42	2.95	0.17	1.75*	0.97	0.04	0.04	0.01	0.09
Soft Drinks	4.17	0.65	4.06	0.65	3.28	0.46	2.35	0.70	0.11	0.04	0.06	0.06
Spreads	3.71	0.70	3.61	0.86	2.75	0.60	2.33*	1.54	0.03	0.14	0.05	0.15
Sugar	4.16	1.58	3.31+	1.92	2.00*	1.13	0.73+	1.02	0.26	0.04	0.38	0.03
Tinned Tomatoes	3.95	0.41	3.97*	1.02	2.70	0.14	1.98+	1.73	0.03	0.06	0.18	0.01
Toilet Paper	5.21	0.91	4.91	1.36	4.05	0.61	3.25	0.90	0.09	0.06	0.05	0.11
Average	3.88	0.59	3.69	0.94	2.73	0.44	2.14	1.00	0.07	0.07	0.09	0.08
Average (Exp. Wgt.)	3.81	0.57	3.61	0.84	2.63	0.71	2.23	0.92	0.06	0.05	0.07	0.08

Note: *(+) indicates that the elasticity of substitution fell below one (zero) in at least one period.

goods on the cost-of-living index but first discuss the estimate of the elasticity of substitution which we have used in the results in the following section.

In applying the adjustment for new and disappearing goods discussed above we used an estimate of the elasticity of substitution derived from the *Current v Base* method and estimated each time period to maximize the flexibility of the CES functional form. However, when the elasticity of substitution fell below one we instead used the average estimate over all time periods. In the unusual case where the average estimated elasticity of substitution over all time periods was less than one, for the *Current v Base* method, we did not undertake the adjustment for new and disappearing goods.[11] With an estimate of the elasticity of substitution in hand we can now examine the effects of new and disappearing goods on the cost-of-living index.

4.3. *The Effect of New and Disappearing Goods on the Cost-of-Living*

The results of the application of the CES cost function approach to the problem of new and disappearing goods are surprising and shown in Table 2. For each of the aggregation methods and for almost all of the goods, in 69 of the 73 possible cases, the matched-goods price index lies above the price index which reflects new and disappearing goods. The extent of the bias differs by aggregation method and product category. On average, over all goods and aggregation methods, the matched-goods price index was upwardly biased by 2.3 percentage points. The range of bias for the different aggregation methods varied from around 1.5 to 3 percent over the 65 week period or around 1.2 to 2.4 percent annually. This is significantly larger than the estimate by the Boskin Commission (Boskin *et al.*, 1996, Tab. 2) mentioned earlier of an upward bias of 0.3 percentage points each year for the "Food at Home other than Produce" category.

The interesting feature of these results is that they imply a sizeable bias for the matched-goods method despite there being a large overlap of expenditures on common goods. As can be seen in Table 3, the average proportion of expenditure on *new* and *disappeared* goods is relatively small, usually less than 1 or 2 percent of total expenditure but depending on aggregation method. For the matched-goods price index to be upwardly biased it must be the case that expenditure on new goods is consistently larger than expenditure on disappeared goods. Indeed this seems to be a very strong empirical regularity in our data set. In the RHS section of Table 3 we show the percentage change in an index of the relative expenditure between current and base periods on those goods which are common to both periods.[12] For all but 3 of our 76 comparisons, these indexes fell, and often quite significantly. It is this empirical regularity which is driving our estimate of an upward bias from omitting new and disappearing goods. However, while this is a strong feature of the data used in this study it may not arise in all such data sets. For example, in Reinsdorf's coffee data (Reinsdorf, 1999, p. 155, Tab. 3) there seems to be no systematic difference between the expenditure on new and disappeared varieties of coffee. In contrast in a scanner data study by Dalén (1997, p. 2, Tab. 1), which included data for four products categories, we do in fact see strong

[11]This only occurred for *Jams*, *Pasta* and *Sugar* for the *Quarter (Prod. Code)* aggregation method.
[12]Note that this is a chained index of the adjustment in (3) without the elasticity exponent.

TABLE 2

THE EFFECTS OF NEW AND DISAPPEARING GOODS ON THE COST-OF-LIVING INDEX

	CES Cost Function Method Adjustment for New and Disappearing Goods over 65 Weeks (%)				Approximate Reservation Price Method Adjustment for New and Disappearing Goods over 65 Weeks (%)			
Aggregation Method	Month (Prod. Code, Outlet)	Month (Prod. Code)	Quarter (Prod. Code, Outlet)	Quarter (Prod. Code)	Month (Prod. Code, Outlet)	Month (Prod. Code)	Quarter (Prod. Code, Outlet)	Quarter (Prod. Code)
Biscuits	2.08	4.19	0.51	5.52	–1.57	1.18	–0.53	1.09
Bread	0.61	1.46	3.69	3.22	0.35	0.46	0.90	0.83
Butter	0.88	1.81	–1.74	0.85	0.29	0.61	–0.37	0.22
Cereal	6.08	5.89	3.77	5.62	1.72	1.61	0.90	1.11
Coffee	1.03	0.70	1.14	1.07	0.34	0.23	0.37	0.29
Detergent	1.83	1.32	0.93	6.07	0.59	0.42	0.22	0.79
Frozen Peas	1.71	1.46	3.14	5.49	0.51	0.41	0.53	0.62
Honey	0.06	0.35	0.53	1.52	0.05	0.12	0.09	–0.07
Jams	1.44	6.60	–4.27	#	0.42	1.09	0.22	1.35
Juices	3.27	1.99	2.90	3.69	0.87	0.54	0.67	0.84
Margarine	1.22	1.25	1.67	0.76	0.33	0.38	0.49	0.25
Oil	2.14	0.73	2.36	1.17	0.70	0.26	0.65	0.31
Pasta	3.35	3.22	3.89	#	0.71	0.48	0.47	1.17
Pet Food	0.99	2.78	1.79	5.37	0.33	0.81	0.52	1.48
Soft Drinks	3.97	1.30	–0.40	0.54	1.14	0.40	–0.10	0.11
Spreads	8.74	1.92	0.83	1.42	2.40	0.60	0.30	0.70
Sugar	–0.58	0.19	0.50	#	–0.15	0.04	–0.07	0.10
Tinned Tomatoes	5.31	2.33	3.92	5.17	1.63	0.87	1.03	0.97
Toilet Paper	4.07	2.82	1.59	2.28	1.23	0.67	0.51	0.70
Average	2.54	2.23	1.41	[3.11]	0.63	0.59	0.36	0.68
	[2.75]	[2.02]	[1.66]		[0.68]	[0.60]	[0.39]	[0.64]
Average (Exp. Wgt.)	2.77	2.50	1.65	[3.30]	0.58	0.69	0.35	0.75
	[2.83]	[2.45]	[1.70]		[0.59]	[0.70]	[0.36]	[0.74]

Note: #indicates that the average elasticity was less than one so the CES Approach is invalid. The average excluding these goods is in square brackets.

TABLE 3

Expenditure Shares on New and Disappeared Goods

Aggregation Method	Average Proportion of Current and Base Expenditure Shares (%)								Changes in Current Relative to Base Expenditure Shares Change over 65 Weeks (%)			
	Month (Prod. Code, Outlet)		Month (Prod. Code)		Quarter (Prod. Code, Outlet)		Quarter (Prod. Code)		Month (Prod. Code, Outlet)	Month (Prod. Code)	Quarter (Prod. Code, Outlet)	Quarter (Prod. Code)
New or Dis. Goods	Dis.	New	Dis.	New	Dis.	New	Dis.	New				
Biscuits	1.66	2.20	0.06	0.76	3.97	4.58	0.29	2.38	–7.36	–9.41	–2.43	–8.14
Bread	0.50	0.77	0.11	0.36	0.78	2.12	0.34	1.66	–3.70	–3.57	–5.31	–5.20
Butter	0.59	0.77	0.01	0.41	1.47	1.05	0.12	0.43	–2.54	–5.52	1.73	–1.27
Cereal	1.15	2.07	0.01	0.78	1.67	2.95	0.08	1.49	–12.33	–10.43	–5.11	–5.55
Coffee	0.40	0.75	0.01	0.17	0.66	1.47	0.02	0.51	–4.86	–2.31	–3.24	–1.93
Detergent	0.75	1.15	0.02	0.28	1.82	2.13	0.04	0.72	–5.48	–3.66	–1.26	–2.68
Frozen Peas	0.75	1.04	0.02	0.23	1.57	2.14	0.10	0.70	–4.03	–2.89	–2.33	–2.41
Honey	0.53	0.61	0.02	0.12	1.08	1.15	0.11	0.57	–1.08	–1.34	–0.28	–1.84
Jams	1.33	1.56	0.03	0.54	2.46	2.85	0.08	1.33	–3.25	–6.84	–1.59	–4.90
Juices	0.81	1.24	0.02	0.29	1.73	2.63	0.17	1.32	–5.85	–3.68	–3.64	–4.55
Margarine	0.70	0.99	0.01	0.24	0.92	1.88	0.18	0.71	–4.07	–3.18	–3.85	–2.11
Oil	1.80	2.32	0.05	0.30	1.47	2.59	0.18	0.69	–7.23	–3.40	–4.50	–2.03
Pasta	1.12	1.40	0.08	0.25	2.23	2.68	0.27	0.90	–3.93	–2.46	–1.85	–2.47
Pet Food	1.14	1.37	0.09	0.52	2.27	3.17	0.23	1.74	–3.26	–5.97	–3.65	–5.91
Soft Drinks	2.03	2.75	0.02	0.26	2.02	1.93	0.09	0.24	–9.83	–3.34	0.34	–0.60
Spreads	0.91	2.17	0.01	0.37	1.21	1.83	0.10	0.83	–16.54	–5.05	–2.51	–2.93
Sugar	0.34	0.29	0.00	0.02	0.47	0.47	0.00	0.12	0.71	–0.26	–0.03	–0.46
Tinned Tomatoes	1.33	2.31	0.02	0.54	3.06	4.64	0.16	2.12	–13.27	–7.12	–6.38	–7.63
Toilet Paper	1.37	2.58	0.01	0.53	2.25	3.49	0.06	1.51	–16.09	–7.11	–4.99	–5.72
Average	1.01	1.49	0.03	0.37	1.74	2.41	0.14	1.05	–6.53	–4.61	–2.68	–3.60
Average (Exp. Wgt.)	1.18	1.72	0.04	0.42	1.93	2.65	0.16	1.21	–7.29	–5.26	–2.92	–4.12

evidence that the proportion of expenditure on new goods is larger than that on disappeared goods.

An interesting question is, what is causing the disparity between expenditure on new and disappeared goods? One explanation is that there are an ever increasing number of products so that the number of newly introduced goods exceeds the number of goods withdrawn from the market. If this is the case then we would typically expect expenditure shares to follow a similar pattern. Table 4 compares the number of products available in each product category in the first and last of the time periods for each aggregation method. The results show that for most of the product categories the product range increased over time. However, this appears to be only a partial explanation. For example, consider the case of *Soft Drinks* where there were sizeable reductions in the number of varieties over the period despite an upward bias in the matched-goods index for 3 out of 4 aggregation methods. However, in Table 3, we see that for *Soft Drinks* generally the proportion of expenditure on new goods was larger than that on disappeared goods. This may indicate that complex factors, such as consumers' desire for variety, are at play.

4.4. *A Comparison with the Approximate Reservation Price Method*

As outlined in Section 2.2, an alternative method for determining the influence of new and disappearing goods on the cost-of-living index is to estimate reservation prices. It is interesting to compare the results from the *CES Cost Function Approach* with the *Approximate Reservation Price Method*. In order to apply this latter method we require an estimate of the price elasticity of demand, ε_i^t, from (1). To ensure comparability with the *CES Cost Function Approach* we used an estimate of ε_i^t derived from the CES functional form, $\hat{\varepsilon}_i^t = \hat{\sigma}\left(1 - s_i^t\right)$, where $\hat{\sigma}$ is the estimated elasticity of substitution and s_i^t is the expenditure share of the good.[13] Then in order to estimate the reservation price of a good i which is new in period 1 and hence absent in period 0 we use (1) to obtain $\hat{p}_i^1$ and then note that if ε_i^t is fixed over time (i.e. $\varepsilon_i^1 = \varepsilon_i^0$) then it can be shown that $\hat{p}_i^0 = \hat{p}_i^1 / \left(p_i^1 / p_i^0\right)$. However, the good i is new so p_i^0 does not exist in which case we use the overall price index to represent p_i^1 / p_i^0.[14] This price is then used in an index formula in a conventional fashion. An analogous method is used for goods which were available in period 0 and disappeared in period 1.

In determining the effect of new and disappearing goods using this approximate reservation price method we compare a matched-goods Törnqvist Price Index (P^T), shown in (13), with an Augmented Törnqvist Price Index (P^{AT}) which reflects new and disappearing goods, (14).

$$P^T \equiv \prod_{i \in I^{1,0}} \left(\frac{p_i^1}{p_i^0} \right)^{\frac{\hat{s}_i^0 + \hat{s}_i^1}{2}} \tag{13}$$

[13]To see this is indeed correct for the CES functional form consult the Appendix.
[14]That is, we use $\hat{p}_i^0 = \hat{p}_i^1 / P^{AT}$ to derive $\hat{p}_i^0$ where P^{AT} is defined in (14).

TABLE 4
Number of Products

Aggregation Method	Month (Prod. Code, Outlet)			Month (Prod. Code)			Quarter (Prod. Code, Outlet)			Quarter (Prod. Code)		
Time Period	First	Last	Chg. (%)	First	Last	Chg. (%)	First	Last	Chg. (%)	First	Last	Chg. (%)
Biscuits	38,408	39,220	2.11	918	883	–3.81	41,777	42,537	1.82	992	1,009	1.71
Bread	11,587	11,048	–4.65	308	335	8.77	12,085	12,175	0.74	325	355	9.23
Butter	3,174	3,422	7.81	61	65	6.56	3,321	3,495	5.24	64	66	3.13
Cereal	17,769	19,533	9.93	389	476	22.37	18,577	20,613	10.96	399	491	23.06
Coffee	7,958	8,937	12.30	152	168	10.53	8,412	9,408	11.84	162	176	8.64
Detergent	7,310	6,956	–4.84	147	142	–3.40	7,514	7,430	–1.12	151	154	1.99
Frozen Peas	8,163	8,300	1.68	189	186	–1.59	8,651	8,636	–0.17	201	196	–2.49
Honey	3,841	3,887	1.20	97	90	–7.22	4,018	4,013	–0.12	102	93	–8.82
Jams	10,857	10,329	–4.86	312	282	–9.62	11,626	11,125	–4.31	328	303	–7.62
Juices	41,800	42,426	1.50	862	894	3.71	43,577	44,769	2.74	899	963	7.12
Margarine	4,514	4,872	7.93	80	89	11.25	4,625	5,128	10.88	81	90	11.11
Oil	8,220	9,116	10.90	232	251	8.19	9,049	9,900	9.40	250	273	9.20
Pasta	16,256	18,207	12.00	465	544	16.99	17,453	19,552	12.03	523	578	10.52
Pet Food	39,697	41,605	4.81	786	876	11.45	41,229	44,628	8.24	806	935	16.00
Soft Drinks	37,277	32,043	–14.04	774	681	–12.02	40,067	34,758	–13.25	812	732	–9.85
Spreads	4,191	4,615	10.12	84	89	5.95	4,668	4,924	5.48	89	93	4.49
Sugar	4,190	4,200	0.24	96	103	7.29	4,355	4,439	1.93	101	106	4.95
Tinned Tomatoes	3,646	3,867	6.06	90	91	1.11	3,931	4,210	7.10	98	97	–1.02
Toilet Paper	6,564	6,835	4.13	129	134	3.88	6,835	7,138	4.43	135	138	2.22
Average	–	–	3.39	–	–	4.23	–	–	3.89	–	–	4.40
Average (Exp. Wgt.)	–	–	0.99	–	–	3.99	–	–	2.20	–	–	5.52

Numbers of Products in First and Last Time Periods and the Percentage Change

$$(14) \qquad P^{AT} \equiv \prod_{i \in I^{1,0}} \left(\frac{p_i^1}{p_i^0} \right)^{\frac{s_i^0 + s_i^1}{2}} \times \prod_{i \in I^0, i \notin I^{1,0}} \left(\frac{\hat{p}_i^1}{p_i^0} \right)^{\frac{s_i^0}{2}} \times \prod_{i \in I^1, i \notin I^{1,0}} \left(\frac{p_i^1}{\hat{p}_i^0} \right)^{\frac{s_i^1}{2}}.$$

The difference between the two indexes is that the Augmented Törnqvist Price Index includes the effect of new and disappearing goods through the use of the estimated reservation prices.

The results of this exercise are informative and help to reinforce our strong suspicion that the matched-goods price index is upwardly biased. The point estimate of the bias from this approach, over all product categories and aggregation methods, is 0.6 percentage points over the 65 weeks. The estimate ranges between 0.3 to 0.8 percentage points for the different aggregation methods.

As suspected the linearization of the demand curve has led to an estimate of the bias from omitting the effects of new and disappearing goods which is significantly less than that obtained for the CES Cost Function Approach. Nevertheless, the "conservative" estimate, based upon linearization, still implies significant bias, of the order of 0.5 percentage points annually. More work must be undertaken on increasing the flexibility of the functional which are used to assess the effects of new and disappearing goods. While the assumption of homotheticity inherent in the CES functional form is unlikely to be pivotal, as income effects are likely small over such a short time period, other factors, like the requirement of a constant elasticity of substitution for each good, may be influencing our results. Linearizing the market demand is not ideal, however, as it will tend to underestimate the bias, first because of the linearity and second due to fact that the market demand curve is flatter than the compensated demand curve. While the use of a compensated demand curve is theoretically correct, the assumption of CES preferences may impose too much convexity on the curve which could lead to an overestimate of the effects of new and disappearing goods. Further research is required to determine the extent of the various factors at play here before any approach can be applied in a statistical agency environment.

5. Conclusion

The main purpose of this article has been to quantify the effects of new and disappearing goods on the cost-of-living index using a scanner data set. To this end we primarily adopted a particular approach to the measurement of this effect based on the CES cost function. The advantage of this approach, as opposed to alternative methods such as hedonic regression or the estimation of reservation prices, is that very little has to be estimated. Using the *CES Cost Function Approach*, all that we require is an estimate of the elasticity of substitution which can be relatively easily obtained. These estimates were then used to determine the effects of new and disappearing goods on the cost-of-living index.

Most significantly, our results show that the matched-goods price index is upwardly biased due to the systematically larger expenditure on new goods than on disappeared goods. This upward bias appears to be larger than previously thought and, on average, is between 1.5 and 3 percent over the 65 week period under study.

In annualized terms this amounts to an upward bias of 1.2–2.4 percent. Our use of an approximate reservation price method confirmed these results though they indicated a smaller, but nevertheless still significant, upward bias for the matched-goods price index. Bias of this magnitude is too large to ignore and shows that the matched-goods approach is inadequate in a dynamic economic environment where the range and variety of products are constantly changing.

Appendix

The Elasticity of Substitution and the Demand Function

Here we briefly show that, for the CES functional form, we must have $\sigma \geq 1$ for the compensated demand function not to slope upwards. We also derive the form of the price elasticity of demand, ε_i^t. Using Shephard's Lemma we can derive the compensated demand curve:

$$x_i^t(p^t, \bar{U}|I^t) = \frac{\partial C(p^t, \bar{U}|I^t)}{\partial p_i^t} = b_i(p_i^t)^{-\sigma}\left(\sum_{i \in I^t} b_i(p_i^t)^{1-\sigma}\right)^{\frac{\sigma}{1-\sigma}} \bar{U}. \tag{15}$$

Let us differentiate $x_i^t(p^t, \bar{U}|I^t)$ in order to determine the slope of the compensated demand function. With a bit of manipulation and using some of the definitions outlined above we obtain the following expression:

$$\frac{\partial x_i^t(p^t, \bar{U}|I^t)}{\partial p_i^t} = \sigma \frac{x_i^t}{p_i^t}(s_i^t - 1). \tag{16}$$

Given that x_i^t, p_i^t and s_i^t are positive with $s_i^t \leq 1$ we see that for the derivative to be non-positive, that is for the demand curve not to slope upwards, we must have $\sigma \geq 0$. Also from (16) we can easily see the form of the price elasticity of demand for the CES functional form.

$$\varepsilon_i^t \equiv -\frac{\partial x_i^t(p^t, \bar{U}|I^t)}{\partial p_i^t}\frac{p_i^t}{x_i^t} = -\sigma(s_i^t - 1). \tag{17}$$

Deriving the CES Cost-of-Living Index

The cost-of-living index over changing domains of goods for the CES cost function has the form shown below:

$$\frac{C(p^1, \bar{U}|I^1)}{C(p^0, \bar{U}|I^0)} \equiv \left(\frac{\sum_{i \in I^1} b_i(p_i^1)^{1-\sigma}}{\sum_{i \in I^0} b_i(p_i^0)^{1-\sigma}}\right)^{\frac{1}{1-\sigma}}. \tag{18}$$

Following Balk (1999), let us briefly show how the exact CES cost-of-living index can be calculated. We will make use of Shephard's Lemma which, when applied to the CES cost function, gives the relationship shown in (19) between observable expenditure shares and the unobservable parameters of the cost function.

$$s_i^t \equiv \frac{p_i^t x_i^t}{\sum_{i\in I^t} p_i^t x_i^t} = \frac{b_i(p_i^t)^{1-\sigma}}{\sum_{i\in I^t} b_i(p_i^t)^{1-\sigma}}, \quad i \in I^t, \quad t = 0, 1. \tag{19}$$

Let us also define the expenditure shares over the set of matched goods, denoted by $\hat{s}_i^t$. Using (19) above it can be shown that the following relationship holds between the expenditure shares for matched goods and the parameters of the cost function.

$$\hat{s}_i^t \equiv \frac{p_i^t x_i^t}{\sum_{i\in I^{1,0}} p_i^t x_i^t} = \frac{b_i(p_i^t)^{1-\sigma}}{\sum_{i\in I^{1,0}} b_i(p_i^t)^{1-\sigma}}, \quad i \in I^{1,0}, \quad t = 0, 1. \tag{20}$$

Using these equations we are now able to derive the exact cost-of-living index over a changing domain of goods.

$$\frac{C(p^1, \bar{U} | I^1)}{C(p^0, \bar{U} | I^0)} = \left(\frac{\sum_{i\in I^1} b_i(p_i^1)^{1-\sigma}}{\sum_{i\in I^0} b_i(p_i^1)^{1-\sigma}} \right)^{\frac{1}{1-\sigma}} \tag{21}$$

$$= \left(\frac{\sum_{i\in I^1} b_i(p_i^1)^{1-\sigma}}{\sum_{i\in I^{1,0}} b_i(p_i^1)^{1-\sigma}} \frac{\sum_{i\in I^{1,0}} b_i(p_i^1)^{1-\sigma}}{\sum_{i\in I^{1,0}} b_i(p_i^0)^{1-\sigma}} \frac{\sum_{i\in I^{1,0}} b_i(p_i^0)^{1-\sigma}}{\sum_{i\in I^0} b_i(p_i^0)^{1-\sigma}} \right)^{\frac{1}{1-\sigma}} \tag{22}$$

$$= \left(\frac{\sum_{i\in I^{1,0}} s_i^1}{\sum_{i\in I^{1,0}} s_i^0} \right)^{\frac{-1}{1-\sigma}} \left(\frac{\sum_{i\in I^{1,0}} b_i(p_i^1)^{1-\sigma}}{\sum_{i\in I^{1,0}} b_i(p_i^0)^{1-\sigma}} \right)^{\frac{1}{1-\sigma}}. \tag{23}$$

Equation (21) is the definition of the cost-of-living index and in (22) we simply multiply and divide the CES cost-of-living index by the same expressions. We can eliminate the first and third fractions in (22) by using (19) and we are left with a representation of the cost-of-living index which is a function of observable expenditure shares and unobservable parameters of the cost function relating to matched-goods. The second term on the RHS is equal to the Base-Period Weighted and Current-Period Weighted Price Indexes by substituting in equation (20) for $t = 0$ and $t = 1$ respectively. For the derivation of the Sato–Vartia Price Index, see Feenstra (1994) or Balk (1999).

References

Balk, Bert, "On Curing the CPI's Substitution and New Goods Bias," Paper presented at the Fifth Meeting of the Ottawa Group, Reykjavik, Iceland, 1999.
Berndt, Ernst R. and Neil J. Rappaport, "Price and Quality of Desktop and Mobile Personal Computers: A Quarter-Century Historical Overview," *American Economic Review*, 91(2), 268–73, 2001.
Berndt, Ernst R., Zvi Griliches, and Neil J. Rappaport, "Econometric Estimates of Price Indexes for Personal Computers in the 1990s," *Journal of Econometrics*, 68(1), 243–68, 1995.
Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale Jorgenson, *Toward a More Accurate Measure of the Cost of Living*, Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index, US, 1996.
Bradley, Ralph, Bill Cook, Silvia G. Leaver, and Brent Moulton, "An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI," Paper presented at the Third Meeting of the Ottawa Group, Voorburg, Netherlands, 1997.
Bresnahan, Timothy F., "Comment on Valuation of New Goods under Perfect and Imperfect Competition," in Timothy F. Bresnahan and Robert J. Gordon (eds), *The Economics of New Goods*, University of Chicago Press, Chicago, 237–47, 1997.
Dalén, Jörgen, "Experiments with Swedish Scanner Data," Paper presented at the Third Meeting of the Ottawa Group, Voorburg, Netherlands, 1997.
de Haan, Jan, "Generalized Fisher Price Indexes and the Use of Scanner Data in the CPI," Paper presented at the Sixth Meeting of the Ottawa Group, Canberra, Australia, 2001.
Diewert, W. Erwin, "The Economic Theory of Index Numbers: A Survey," in W. Erwin Diewert and Alice O. Nakamura (eds), *Essays in Index Number Theory 1*, North-Holland, Amsterdam, 1993.
———, "Hedonic Regressions: A Review of Some Unresolved Issues," Paper presented at the Seventh Meeting of the Ottawa Group, Paris, France, 2003.
Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84(1), 157–77, 1994.
Feenstra, Robert C. and Clinton R. Shiells, "Bias in U.S. Import Prices and Demand," NBER Working Paper, No. 4841, 1994.
Feenstra, Robert C. and Matthew D. Shapiro, "High Frequency Substitution and the Measurement of Price Indexes," in Robert C. Feenstra and Matthew D. Shapiro (eds), *Scanner Data and Price Indexes*, Studies in Income and Wealth Vol. 64, The University of Chicago Press, Chicago, 123–46, 2003.
Gordon, Robert J. and Zvi Griliches, "Quality Change and New Products," *American Economic Review*, 87(2), 84–8, 1997.
Hardy, G. H., J. E. Littlewood, and G. Polya, *Inequalities*, 2nd edition, Cambridge University Press, Cambridge, 1952.
Hausman, Jerry, "Valuation of New Goods under Perfect and Imperfect Competition," in Timothy F. Bresnahan and Robert J. Gordon (eds), *The Economics of New Goods*, University of Chicago Press, Chicago, 209–37, 1997.
———, "Sources of Bias and Solutions to Bias in the Consumer Price Index," *Journal of Economic Perspectives*, 17(1), 23–44, 2003.
Hicks, John R., "The Valuation of Social Income," *Economica*, 7(2), 105–24, 1940.
ILO (International Labour Organisation), *Consumer Price Index Manual: Theory and Practice*, ILO, Hong Kong, 2004.
Jain, Malti and Joanne Caddy, "Using Scanner Data to Explore Unit Value Indexes," Room Document at the Sixth Meeting of the Ottawa Group, Canberra, Australia, 2001.
Koskimäki, Timo and Mari Ylä-Jarkko, "Segmented Markets and CPI Elementary Classifications," Paper presented at the Seventh Meeting of the Ottawa Group, Paris, France, 2003.
Lloyd, Peter, "Substitution Effects and Biases in Nontrue Price Indices," *American Economic Review*, 65(3), 301–13, 1975.
Marshall, Alfred, "Remedies for Fluctuations of General Prices," in Arthur C. Pigou (ed.), *Memorials of Alfred Marshall*, Kelley and Millman, New York, 188–211, 1887.
Nahm, Daehoon, "Incorporating the Effect of New and Disappearing Products into the Cost of Living: An Alternative Index Number Formula," *Seoul Journal of Economics*, 11(3), 261–82, 1998.
Opperdoes, Eddy, "Some Empirical Experiments with CES Functions," Statistics Netherlands, Paper presented at the Sixth Meeting of the Ottawa Group, Canberra, Australia, 2001.
Reinsdorf, Marshall, "Using Scanner Data to Construct CPI Basic Component Indexes," *Journal of Business and Economic Statistics*, 17(2), 152–60, 1999.
Richardson, David H., "Scanner Indexes for the CPI," Paper presented at the Sixth Meeting of the Ottawa Group, Canberra, Australia, 2000.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34–55, 1974.

Schultze, Charles and Christopher Mackie (eds), *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*, National Academy of Sciences, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Academy Press, Washington DC, 2002.

Schut, Cecile M., "Using Scanner Data to Calculate Consumer Price Indices," Statistics Netherlands, Netherlands, 2002.

Shapiro, Matthew D. and David W. Wilcox, "Alternative Strategies for Aggregating Prices in the CPI," *Federal Reserve Bank of St. Louis Review*, 79(3), 113–25, 1997.

Silver, Mick, "Elementary Aggregates, Micro-Indices and Scanner Data: Some Issues in the Compilation of Consumer Price Indices," *Review of Income and Wealth*, 41(4), 427–38, 1995.

———, "An Evaluation of the Use of Hedonic Regressions for Basic Components of Consumer Price Indices," *Review of Income and Wealth*, 45(1), 41–56, 1999.

Silver, Mick and Bruce Webb, "The Measurement of Inflation: Aggregation at the Basic Level," *Journal of Economic and Social Measurement*, 28(1–2), 21–35, 2002.

Szulc, Bohdan J., "Linking Price Index Numbers," in W. Erwin Diewert and C. Montmarquette (eds), *Price Level Measurement*, Statistics Canada, Ottawa, 537–66, 1983.

Triplett, Jack E., "Using Scanner Data in Consumer Price Indexes: Some Neglected Conceptual Considerations," in Robert C. Feenstra and Matthew D. Shapiro (eds), *Scanner Data and Price Indexes*, 151–64, 2003.