Nonlinear Pricing and Inflation Measurement: Evidence and Implications from Scanner Data

By Kevin J. Fox and Daniel Melser
Nonlinear Pricing and Inflation Measurement: Evidence and Implications from Scanner Data

Kevin J. Fox*  
School of Economics and CAER  
University of New South Wales  
Sydney, NSW 2052  
Australia  
Tel: +(61 2) 9385 3320  
Fax: +(61 2) 9313 6337  
Email: K.Fox@unsw.edu.au

Daniel Melser  
Department of Economics  
Monash University  
Melbourne, VIC 3145  
Australia  
Tel: +(61 2) 9270 8118  
Fax: +(61 2) 9252 3181  
Email: Daniel.Melser@buseco.monash.edu.au

April 2007

Abstract: Using supermarket electronic point-of-sale “scanner” data on soft drinks for the US and Australia, we find significant discounts available for larger package sizes and multi-packs. Besides its interest from the point of view of microeconomic analysis, such nonlinear pricing is of practical importance for statistical agencies that construct inflation measures over goods with a multitude of package and pack sizes. We argue that the assumption of linear pricing should be avoided if possible, and that alternative methods such as those described in this paper should be adopted.

Keywords: Package Size, Hedonic Regression, Nonlinear Pricing, Quantity Discount.

JEL Classification Codes: C43, C50, D00, E31.

* We gratefully acknowledge the provision of data by the Australian Bureau of Statistics and the University of Chicago, James M. Kilts Center, Graduate School of Business. We are also appreciative of the financial support of the Australian Research Council. Useful comments on this paper were received from Shiji Zhao, Keith Woolford, Tue Gørgens and participants at the Economic Measurement Group Workshop, December 2005, UNSW. Any remaining errors are the responsibility of the authors.
1. Introduction

The proliferation of varieties of products poses multiple problems for the measurement of inflation. In a modern economy there is a dizzying diversity of consumer-products available. Supermarkets stock goods in a range of flavours, brands and types, as well as package sizes, from ‘mini’ cans, bottles, boxes or bags to extra large and family sizes. Goods are also often provided in a range of different pack sizes such as singles, doubles, half-dozens and dozens. This paper examines methods for dealing with these complexities in price index construction.

Dealing with the sheer numbers of new products has meant sampling issues have become increasingly important. As products change so must the indexes which purport to measure fluctuations in their prices – linking items into and out of the index as products turn over has become a major undertaking. This quality adjustment problem of comparing two dissimilar items is most often encountered in relatively high-tech sectors where new and improved products replace older-dated varieties; see e.g. Silver and Heravi (2005). However, it is also apparent in the more mundane goods categories of supermarket and convenience store items. One of the most common quality adjustment problems in constructing measures of inflation is in comparing the same good sold in different package sizes or when there is a change in the number of packs available. For example, it is unclear how the price for a 2-litre bottle of cola should be compared with that for a 1-litre bottle, or a dozen cans of cola compared with singles.

Lower unit-prices for buying larger package sizes are ubiquitous in supermarkets. A recent visit to a major UK supermarket found that there were five different varieties available of a particular range of Heinz Tomato Ketchup. The package (bottle) sizes available were 460gm, 570gm, 700gm, 910gm, and 1.2kg, where the corresponding prices were £1.19, £1.38, £1.65, £1.98, and £2.59. In this case there was a discount of over 16.5% per unit from purchasing the 1.2kg bottle instead of the 460gm variety.

The primary aim of this paper is to explore the empirical relationship between price, package and pack size. To this end, we develop an hedonic regression model, with applications to supermarket electronic point-of-sale (“scanner”) data sets for the US and Australia. The commodity group chosen for study is soft drinks, given the relatively large variety of package and pack sizes available to consumers. The results reveal a substantial degree of nonlinear pricing, where ‘size’ does not lead to a one-for-one change in prices. This has implications for any statistical agencies which assume linear
pricing.\footnote{Leading statistical agencies, such as the U.S. Bureau of Labor Statistics and the Australian Bureau of Statistics, use the assumption of linear pricing, but only sparingly.} Also, differences in the degree of nonlinear pricing between the two countries suggest that market structure is a significant determinant of the pricing function.

In the next section we briefly review some of the perspectives on nonlinear pricing in the economics literature. In Section 3, we outline the package and pack size quality adjustment problem. Section 4 develops a flexible hedonic regression model for package and pack size based upon consumer optimization. In Section 5 we implement our approach empirically and quantify the extent of quantity discounts in the soft drink market for the US and Australia. We are also interested in the extent to which the use of the linear adjustment method in price index construction may introduce error. We explore some specific examples. Section 6 concludes.

\section{Nonlinear Pricing}

There is considerable interest in nonlinear pricing in a variety of fields, such as in the industrial organization literature; see e.g. Wilson (1993). The first paper to rigorously address the issue was Spence (1976) but the literature has grown significantly since this time. If the producer has some sort of pricing power and consumers differ then it may be possible that a nonlinear pricing schedule will separate the consumers and lead to maximum profits. Maskin and Riley (1984) showed, for a simple utility function and a continuum of consumers, that the optimal pricing function for a monopolistic firm will have declining unit costs. These early results have been extended in a number of ways into different contexts emphasizing that there are strong reasons for quantity discounts based upon price discrimination.

Another perspective on the phenomenon of declining unit prices is that it relates not just to price discrimination but also to the costs of production. Clearly, if costs-per-unit are lower for larger package or pack sizes then, all else equal, this would lead to quantity discounts. Why might marginal costs fall for larger sized items? One reason could relate to the packaging itself (Clements, 2006). Consider the example of soft drinks and the change from a 1-litre to 2-litre bottle. The volume has clearly doubled but the surface area of the packaging has increased by considerably less than double. The amount by which the area increases will depend on the exact shape of the container. Consider a simple example where we assume a cylindrical container (an approximation...}
to a bottle) and we fix the cylinder’s height. In this case it can be shown that when the volume increases by 100% the surface area of the cylinder increases by just under 42%.\(^2\) These economies of scale in packaging may lead to lower unit prices. While empirical evidence on this point is scant, the potential flaw in this argument is that packaging costs are typically too small a component of overall costs to lead to the kind of quantity discounts we see in practice. Moreover, we also observe discounts in the pack-size dimension, e.g. buying a dozen cans of cola is cheaper per-unit than buying a single, but the packaging-per-unit is the same in each case.

Another area where there may be economies of scale is in retailing, transport, advertising and storage. The costs of selling an item, as well as storing and transporting it from the manufacturer to the retailer, may be less per-unit for larger pack and package sizes than for smaller ones. For example, the costs incurred by the retailer in selling an item are, among other things, the cost of stacking the shelves with an item then processing it at the checkout.

So far we have viewed the quantity discount problem exclusively from the firm’s perspective. For consumers there may be greater storage or transport costs from purchasing a larger package or pack size and they need to be compensated for this by paying a lower unit price. Furthermore, there may also be a greater risk of wastage if larger package sizes (and to a lesser extent pack sizes) are purchased. For example, if a consumer purchases a large can of tomato soup, or bottle of soft drink, then they may only consume part of the contents in the first sitting and there is a risk that the remainder will go unused. This lowers the value of larger package sizes to consumers.

In contrast to this intuition, it has been argued by Gertsner and Hess (1986) that consumers will save shopping time and could have lower overall transport costs by purchasing in bulk. In this case we would expect to see *premiums* for bulk purchases. Hence, it seems that the relationship between package size, pack size and price cannot be determined from intuitive reasoning and theoretical modelling alone.

### 3. Nonlinear Pricing and Inflation Measurement

The relationship between price, package size and pack size is of considerable interest to economists for a variety of reasons. Not least of these is related to inflation measurement. A common approach used by statistical agencies when comparing the same good

---

\(^2\)More complex examples where the volume/area and height/radius are optimized give similar results.
but of different pack or package sizes is to try to make the prices comparable by multiplying them by the ratio of the respective package or pack sizes. That is, the price of a 2-litre bottle is made comparable with that of a 1-litre bottle by halving its price; see for example, Armknecht and Maitland-Smith (1999), Triplett (2004), ILO (2004). This approach is highly questionable given the fact that unit prices tend to fall as package size increases.

There are a number of different ways in which a nonlinear relationship between price, pack and package size could introduce error into price indexes. These depend on the precise survey and index calculation methods employed. We consider three examples below.

First, often the pricing specifications adopted by statistical agencies will stipulate that a particular package size be collected. For example, the specification may say that a 2kg bag of potatoes should be collected. When this is not available it is often the case that another package size is recorded and the price is automatically quantity-scaled, either by the price collector or the index calculation computer system, so that it represents the desired specification size.

Second, a comparability problem will often arise when there is a change in package size for a particular good. For example, a can of soup may increase from 400ml in one period to 460ml in a later period. If the price for the 400ml variety was originally collected then this poses a comparability problem with the price in the later period. As noted, a common approach is to scale the price of the new item by the difference in package size – in this case the scale factor would be 0.8696 (=400/460). This adjusted price is then directly compared with the price for the 400ml product in the earlier period.

Finally, retailers may sometimes undertake a promotional offer on a good such as buy-one-get-one-free (these go by the acronym “BOGOF” in the UK). For concreteness, suppose that a price for a single bottle of juice was collected in some previous period while in the current period the retailer introduced a BOGOF offer at the same price. There are two approaches that are often adopted to deal with this problem. The first is to note that it still costs the same amount to buy a bottle of juice so the price has not changed for someone who wants just a single bottle. While this is true it neglects the fact that the consumer has gained utility from the additional unit of the product which is ‘free’. A second approach in this case is to halve the price of the good in the later period which is then compared with the price in the base period.

Each of these examples illustrates the way in which changes in package size and
pack size can impact on the construction of price indexes. It is important to adjust for changes in price related to pack and package size, but implicit in many of these adjustments is a linear relationship between these factors and price. It is not clear that this is a valid assumption:

Considering that the relation between size and price is seldom linear, it is a bit surprising that statistical agencies use predominantly the simple linear form of package size adjustment... (p. 20, Chp. 1, Triplett, 2004)

If price-per-unit varies then assuming a linear relationship will give incorrect and potentially biased results. This raises the important question of how this problem should be dealt with in the construction of official price indexes.

The recent ILO Manual (2004) on Consumer Price Indexes (CPIs) echoes Triplett (2004) and cautions against the mechanical application of package size quality adjustment:

It is generally a considerable oversimplification to assume that the quality of a product changes in proportion to the size of some single physical characteristics. For example, most consumers are very unlikely to rate a refrigerator that has three times the capacity of a smaller one as being worth three times the price of the latter. Nevertheless it is clearly possible to make some adjustment to the price of a new quality or different size to make it more comparable with the price of an old quality. There is considerable scope for the judicious, or common sense, application of relatively straightforward quality adjustments of this kind. (p. 29, ILO, 2004)

While this is certainly sound advice it is not clear what type of ‘common sense’ adjustment rules should be applied in practice.

Given the practices adopted, the equivocal stance of the ILO Manual and the fact that very little research has been undertaken in this area, there is clearly scope for further investigation. First, we want to identify the nature of the relationship between price, pack and package size, and second, determine whether this relationship is robust enough to be useful in compiling indexes.

4. An Hedonic Regression Model

In our attempt to understand the effects of package and pack size on price we develop an hedonic regression model. This model is based upon an optimizing representative consumer choosing over a range of products, i.e. goods offered in a range of package
and pack sizes. Our hypothesized functional form for the price function offers a flexible framework in which to estimate the desired effects.

The theory of hedonic regression has become increasingly popular in recent years. With the modern empirical literature pioneered by Griliches (1961), the conceptual basis of hedonic regression was established by Lancaster (1966), and then outlined and developed by Rosen (1974) and Muellbauer (1974). It is now widely used in research, increasingly in relation to quality-adjusting price indexes (e.g. Melser, 2005), with many applications using electronic point-of-sale “scanner data”; see e.g. Silver and Heravi (2001, 2005). In addition, it is increasingly used in the construction of official price indexes by statistical agencies such as the US’s Bureau of Labor Statistics (Moulton, 2001), the UK’s Office for National Statistics (Ball and Allen, 2003), the Australian Bureau of Statistics (Berger, 2003) and others. Diewert (2003) has recently outlined the consumer theory justification for hedonic regressions, and this forms the basis of our approach. We briefly present it here from a somewhat different perspective.

The well-known first-order condition for cost minimization (and utility maximization), for some representative consumer, is as follows:

\[ p_{it} = \lambda_t U'_i(q_t), \quad i = 1, \ldots, I, \ t = 0, \ldots, T, \]  

(1)

where \( p_{it} \) is the price of product \( i \) in period \( t \), \( \lambda_t \) is the Lagrange Multiplier, \( q_t = [q_{it}, \ldots, q_{in}] \) is the vector of quantities and \( U'_i(q_t) \) is first-order derivative of the utility function with respect to the quantity of good \( i \).

In order to obtain an hedonic regression model we need to hypothesize a functional form for the utility function. Note that in general the marginal utility for a particular product will depend upon the characteristics of that product along with the units consumed, as well as the consumption of all other goods. The standard form of an hedonic price function does not include the quantity consumed of other products. As noted by Diewert (2003), to derive such a model from consumer theory requires strong assumptions on the form of the utility function, \( U(q_t) \), namely that it is linear in goods:\(^3\)

\[ U(q_t) = \sum_{i=1}^{I} \alpha_i q_{it}, \quad t = 0, \ldots, T. \]  

(2)

Here \( \alpha_i \) is a taste or quality parameter. Given this utility function the hedonic model

\(^3\)In an interesting debate in the literature, Hausman (2003) has objected to this assumption as too restrictive and hence favours the use of alternative methods for measuring quality change, namely the estimation of reservation prices for new and disappearing goods.
has the form,

\[ p_{it} = \lambda_t \alpha_i, \quad i = 1, \ldots, I, \quad t = 0, \ldots, T \]  

(3)

We now apply this model to the estimation of package and pack size effects on price. Let us augment our notation somewhat and note that each product \( i \) has three dimensions; characteristics, denoted by the vector \( z_i \), (e.g. colour, flavour, brand); package size, denoted by the scalar \( x_i \), (e.g. 500ml or 2lt); and pack size, denoted by the scalar \( v_i \), (e.g. 6 for a 6-pack). What we mean by a “good” is defined by the characteristics vector \( z_i \), and this good is available in various package and pack sizes, which together define a “product”. Consistent with the requirements for the application of hedonic methods outlined above, we assume that the representative consumer has the following functional form for the taste/quality parameters in the utility function:

\[ \alpha_i = f(z_i)g(x_i)h(v_i), \quad i = 1, \ldots, I, \]  

(4)

where \( f(\cdot) \), is an aggregator function over product characteristics, and \( g(\cdot) \) and \( h(\cdot) \) are functions of package size and pack size, respectively. Using (4) the relation in equation (1) has the following form:

\[ p_{it} = \lambda_t f(z_i)g(x_i)h(v_i), \quad i = 1, \ldots, I, \quad t = 1, \ldots, T. \]  

(5)

Equation (5) is useful as it formalises the relationship between the price of a product, its characteristics, pack and package size. With our assumption about the functional form of utility, the price depends multiplicatively on the characteristics function \( f(z) \), the package size effect \( g(x) \) and the multipack effect \( h(v) \). Note that these functions relate directly to elements of the consumer’s utility function.

We have viewed the hedonic function from a consumer perspective but it is just as easy to derive from the point of view of a producer. If we instead used the first order condition for a profit maximising firm, \( p_{it} = C'_i(q_t) \) where the \( C'_i(\cdot) \) is the derivative of the firm’s cost function with respect to the \( i^{th} \) product, we could proceed as above by specifying a functional form for the cost function and hence give the model a producer-side perspective. This approach is followed by Pakes (2003). The problem with such an approach is that the assumption of a price taking firm in many markets is inconsistent with reality. Imperfect competition introduces a markup into the price equation and problems with endogeneity. For a monopolist we have,
\[ p_{it} = \frac{C'_i(q_t)}{1 + \frac{\partial p_{it}}{\partial q_{it}} p_{it}}. \]  

(6)

The elasticity of substitution is simultaneously determined with price meaning that the hedonic function is difficult to estimate. We prefer the consumer-side derivation as here the assumption of a price taking individual is more reasonable and hence the hedonic function directly reflects consumer preferences. We proceed to outline some specific functional forms.

4.1. Functional Forms

In this section we specialise the theoretical model outlined above into an econometric model for estimation.

The conventional approach in the literature for defining the characteristics aggre-
gator \( f(z_i) \) in (5) is to specify a linear functional form in the key characteristics of the products. While this approach works well in most cases, particularly for high-tech goods, food and drink products are in some ways more complex to model than electrical or mechanical goods because the way in which the characteristics influence utility are complex. For example, the utility derived from a bottle of Coca-Cola may not relate simply to the physical characteristics of the good, but to certain difficult-to-measure traits such as ‘taste’, ‘texture’ and such ephemeral characteristics as the ‘style’ of the packaging. While brand dummy variables are a possibility, we instead choose a fully non-parametric approach where we include a dummy variable for each particular configuration of characteristics, excluding package and pack sizes, so that,

\[ \ln f(z_i) = \sum_{\iota=1}^{I} \eta_{\iota} m_{i\iota}, \]  

(7)

where \( m_{i\iota} = 1 \) when \( i = \iota \) and zero otherwise, and the parameters \( \eta_{\iota} \) are hence the level effects of each characteristics configuration. For example, there would be separate dummies for Pepsi Max and regular Pepsi.

Our primary focus is in estimating the package and pack size effects. We employ a flexible parametric approach which provides a Taylor series approximation in logs to the underlying function. While we also consider the lower-order models, the most flexible functional form we consider for \( h(\cdot) \) and \( g(\cdot) \) in (5) is a third degree polynomial:

\[ \log h(x_i) = \gamma_1 \log(x_i) + \gamma_2 [\log(x_i)]^2 + \gamma_3 [\log(x_i)]^3, \]  

(8)
and
\[
\log g(v_i) = \beta_1 \log(v_i) + \beta_2 [\log(v_i)]^2 + \beta_3 [\log(v_i)]^3,
\]
where the \( \gamma_j \) and \( \beta_j \), \( j = 1, \ldots, 3 \), are unknown parameters to be estimated. This functional form allows us to test whether the effect of package size, and pack size, is homogenous of degree one. The homogeneity requirement is that, for example, the price doubles if either the package size doubles or the number of packs doubles. More generally, with reference to (5), we require for some scalar \( k > 0 \) that, \( h(kx) = kh(x) \) and \( g(kv) = kg(v) \). This is of interest because the assumption of linear homogeneity underlies the use of linear adjustments to prices. From equations (8) and (9), it can be shown that for all values of \( k \), \( x \) and \( v \) linear homogeneity of \( h(\cdot) \) and \( g(\cdot) \) necessarily implies \( \gamma_1 = 1 \), \( \gamma_2 = 0 \), \( \gamma_3 = 0 \) and \( \beta_1 = 1 \), \( \beta_2 = 0 \), \( \beta_3 = 0 \), respectively.

Using equations (5), (7), (8) and (9), adding dummy variables \( d_{t\tau} \) for the time periods to represent \( \lambda_t \) (where \( d_{t\tau} = 1 \) for \( t = \tau \) and zero otherwise), and appending a random error term \( e_{it} \) to reflect optimisation error, we get the following estimating equation:

\[
\log(p_{it}) = \alpha + \sum_{\tau=2}^{T} \delta_{t\tau} d_{t\tau} + \sum_{i=2}^{I} \eta_{i} m_{iit} + \gamma_1 \log(x_{i}) + \gamma_2 [\log(x_{i})]^2 + \gamma_3 [\log(x_{i})]^3 + \beta_1 \log(v_{i}) + \beta_2 [\log(v_{i})]^2 + \beta_3 [\log(v_{i})]^3 + e_{it}, \ i = 1, \ldots, I, \ t = 1, \ldots, T.
\]

We are now in a position to apply this approach to a large supermarket scanner data set. We focus on just one commodity group where package size and multi-pack purchases are perhaps most common, soft drinks.

5. Results: Price, Package and Pack Size

We make use of two large scanner data sets on soft drinks. The first is scanner data from 111 supermarkets belonging to four chains in one of Australia’s major cities over a 65-week period in 1997 and early 1998. The data is comprehensive in giving information on the price and volume, and provides a description of the product sold including package and pack size, at a weekly frequency. The second data set is similar, but from a large US supermarket chain based in Chicago, Dominick’s Fine Foods, with 96 stores. This data set is also very detailed, having information for a seven-and-a-half-year period. The size and quality of these scanner data sets mean that they are increasingly being
used in both research as well as the construction of official price indexes (see Reinsdorf, 1999).

The defining feature of both data sets is that they are very large with around 1 million observations in the Australian data and almost 6.5 million in the US data. We include time-dummy variables at a weekly frequency which leads to 390 dummies for the US data. In the Australian data set, we have 106 goods (types of soft drinks, available in multiple package and pack sizes) and 188 in the US data set.

Our primary area of interest is in the slope and curvature in the price function in package and pack-size space. We estimate equation (10) for both data sets. We consider three alternative specifications; first, we include all parameters as shown in (10) (Model A), second, we exclude the third-order terms (i.e. $\beta_3 = 0$ and $\gamma_3 = 0$) (Model B) and third, both second- and third-order terms are set to zero (i.e. $\beta_2 = \beta_3 = 0$ and $\gamma_2 = \gamma_3 = 0$) (Model C). The results are recorded in tables 1 and 2 where regression statistics as well as the key parameter estimates are reported, along with White’s (1980) heteroscedasticity robust standard errors. The $R^2$ statistics are fairly high for panel data indicating that the model does a good job of explaining the data.

**Table 1: Regression Results, Australian Data**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-Squared</td>
<td>0.8710</td>
<td>0.8706</td>
<td>0.8659</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>41146.6</td>
<td>41496.6</td>
<td>40,296.3</td>
</tr>
<tr>
<td>Mean Squared Error</td>
<td>0.0853</td>
<td>0.0855</td>
<td>0.0886</td>
</tr>
<tr>
<td>Number of Parameters</td>
<td>176</td>
<td>174</td>
<td>172</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>1,067,069</td>
<td>1,067,069</td>
<td>1,067,069</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(Package Size)</td>
<td>-5.2668</td>
<td>0.079233</td>
<td>-1.0594</td>
<td>0.006243</td>
<td>0.5879</td>
<td>0.00064</td>
</tr>
<tr>
<td>[Log(Package Size)]^2</td>
<td>0.7092</td>
<td>0.010832</td>
<td>0.1222</td>
<td>0.000436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Log(Package Size)]^3</td>
<td>-0.0270</td>
<td>0.000487</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(Number of Packs)</td>
<td>0.7800</td>
<td>0.004382</td>
<td>0.9556</td>
<td>0.002390</td>
<td>0.9131</td>
<td>0.00078</td>
</tr>
<tr>
<td>[Log(Number of Packs)]^2</td>
<td>0.1323</td>
<td>0.003501</td>
<td>-0.0205</td>
<td>0.000948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Log(Number of Packs)]^3</td>
<td>-0.0312</td>
<td>0.000780</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Model A is represented by equation (10), Model B excludes the third-order terms and Model C exclude the second- and third-order terms.

The results for the first-order model, Model C, are the easiest to interpret. Here, if the coefficients on package size and pack size equal one then price and these variables
Table 2: Regression Results, US Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(Package Size)</td>
<td>1.8727</td>
<td>0.101691</td>
<td>-0.1894</td>
<td>0.008337</td>
<td>0.4372</td>
<td>0.00429</td>
</tr>
<tr>
<td>[Log(Package Size)]²</td>
<td>-0.2577</td>
<td>0.014764</td>
<td>0.0457</td>
<td>0.000598</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Log(Package Size)]³</td>
<td>0.0148</td>
<td>0.000711</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(Number of Packs)</td>
<td>0.7590</td>
<td>0.001826</td>
<td>0.7416</td>
<td>0.000852</td>
<td>0.7613</td>
<td>0.000304</td>
</tr>
<tr>
<td>[Log(Number of Packs)]²</td>
<td>-0.0099</td>
<td>0.001240</td>
<td>0.0047</td>
<td>0.000188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Log(Number of Packs)]³</td>
<td>0.0030</td>
<td>0.000235</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: See the note to Table 1.

move in linear step (e.g. a doubling of package size leads to a doubling in price). It can be seen that the coefficient on package size for the Australian data is 0.5879 and 0.9131 for the number of packs. The US data generally shows a flatter slope with coefficients of 0.4372 and 0.7513 respectively. The $t$-tests of each of these individual coefficients against the alternative that they are equal to one, using heteroscedasticity robust standard errors, soundly reject the hypothesis of homogeneity of degree one in package and pack size.

One interesting feature of the results is the steeper slope for pack size than package size. One explanation for this is that the larger package sizes economize on packaging, as discussed in section 2, whereas this is not the case for the packs size, i.e. a dozen cans of cola are a 12-fold replica of a single can so packaging costs change linearly.

Taken together, the results indicate a significant degree of nonlinearity in the relationship between price and the pack and package size variables, with significant discounts for larger quantities along both dimensions.

For the third- and second-order models (models A and B) the relationship between package size, pack size and price is difficult to observe just by looking at the coefficients. For Model A, from equations (5) and (8) the effect on price of a change in package ($x$)
size, all else equal, can be expressed as,
\[
\log \left( \frac{p(z, x_i, v)}{p(z, x_j, v)} \right) = \log h(x_i) - \log h(x_j) = \gamma_1 (\log(x_i) - \log(x_j)) + \gamma_2 ([\log(x_i)]^2 - [\log(x_j)]^2) + \gamma_3 ([\log(x_i)]^3 - [\log(x_j)]^3) \tag{11}
\]
and similarly for a change in pack size \((v)\) using equations (5) and (9):
\[
\log \left( \frac{p(z, x, v_i)}{p(z, x, v_j)} \right) = \log g(v_i) - \log g(v_j) = \beta_1 (\log(v_i) - \log(v_j)) + \beta_2 ([\log(v_i)]^2 - [\log(v_j)]^2) + \beta_3 ([\log(v_i)]^3 - [\log(v_j)]^3). \tag{12}
\]

We construct tables 3 and 4 using the estimated coefficients from tables 1 and 2 in equations (11) and (12), and taking exponents (with appropriate restrictions on the second- and third-order terms for models B and C, we construct tables). The tables show estimated price relatives for different package and pack sizes for the most common values relative to a 1 litre package size and a pack size of one.

Table 3: Package Size and Price

<table>
<thead>
<tr>
<th>Volume</th>
<th>Australian Data</th>
<th></th>
<th></th>
<th>U.S. Data</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model A</td>
<td>Model B</td>
<td>Model C</td>
<td></td>
<td>Model A</td>
<td>Model B</td>
</tr>
<tr>
<td>296ml/10oz</td>
<td>0.5819</td>
<td>0.5577</td>
<td>0.4888</td>
<td></td>
<td>0.6200</td>
<td>0.6247</td>
</tr>
<tr>
<td>300ml/10.1oz</td>
<td>0.5833</td>
<td>0.5602</td>
<td>0.4927</td>
<td></td>
<td>0.6231</td>
<td>0.6275</td>
</tr>
<tr>
<td>355ml/12oz</td>
<td>0.6064</td>
<td>0.5946</td>
<td>0.5440</td>
<td></td>
<td>0.6639</td>
<td>0.6644</td>
</tr>
<tr>
<td>375ml/12.7oz</td>
<td>0.6158</td>
<td>0.6072</td>
<td>0.5618</td>
<td></td>
<td>0.6778</td>
<td>0.6773</td>
</tr>
<tr>
<td>473ml/16oz</td>
<td>0.6675</td>
<td>0.6690</td>
<td>0.6439</td>
<td></td>
<td>0.7401</td>
<td>0.7368</td>
</tr>
<tr>
<td>591ml/20oz</td>
<td>0.7367</td>
<td>0.7433</td>
<td>0.7340</td>
<td></td>
<td>0.8066</td>
<td>0.8026</td>
</tr>
<tr>
<td>600ml/20.3oz</td>
<td>0.7422</td>
<td>0.7489</td>
<td>0.7406</td>
<td></td>
<td>0.8114</td>
<td>0.8074</td>
</tr>
<tr>
<td>946ml/32oz</td>
<td>0.9641</td>
<td>0.9661</td>
<td>0.9679</td>
<td></td>
<td>0.9765</td>
<td>0.9759</td>
</tr>
<tr>
<td>1lt</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td></td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>1.25lt</td>
<td>1.1689</td>
<td>1.1576</td>
<td>1.1402</td>
<td></td>
<td>1.1036</td>
<td>1.1062</td>
</tr>
<tr>
<td>1.5lt</td>
<td>1.3409</td>
<td>1.3164</td>
<td>1.2692</td>
<td></td>
<td>1.2014</td>
<td>1.2054</td>
</tr>
<tr>
<td>2lt</td>
<td>1.6911</td>
<td>1.6393</td>
<td>1.5031</td>
<td></td>
<td>1.3863</td>
<td>1.3888</td>
</tr>
<tr>
<td>3lt</td>
<td>2.4059</td>
<td>2.3114</td>
<td>1.9077</td>
<td></td>
<td>1.7354</td>
<td>1.7176</td>
</tr>
</tbody>
</table>

Note: See the note to Table 1.

The results indicate that there are discounts for larger packs and packages, or alternatively that premiums are paid for smaller sizes. To get an idea of the magnitude
Table 4: Pack Size and Price

<table>
<thead>
<tr>
<th>Number of Packs</th>
<th>Australian Data Model A</th>
<th>Australian Data Model B</th>
<th>Australian Data Model C</th>
<th>U.S. Data Model A</th>
<th>U.S. Data Model B</th>
<th>U.S. Data Model C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>1.8109</td>
<td>1.9204</td>
<td>1.8830</td>
<td>1.6860</td>
<td>1.6758</td>
<td>1.6950</td>
</tr>
<tr>
<td>4</td>
<td>3.4991</td>
<td>3.6158</td>
<td>3.5458</td>
<td>2.8323</td>
<td>2.8210</td>
<td>2.8731</td>
</tr>
<tr>
<td>6</td>
<td>5.1706</td>
<td>5.1879</td>
<td>5.1344</td>
<td>3.8391</td>
<td>3.8338</td>
<td>3.9121</td>
</tr>
<tr>
<td>15</td>
<td>11.7448</td>
<td>11.4422</td>
<td>11.8531</td>
<td>7.7038</td>
<td>7.7123</td>
<td>7.8590</td>
</tr>
</tbody>
</table>

Note: See the note to Table 1.

of the effects of package size, note that from Table 3 we would expect that the price of a 2-litre soft drink relative to a one litre would be between 50.3% and 69.1% higher for the Australian data and 35.4% to 38.8% for the US data. These are considerable deviations from the 100% increase implied by the simple linear relationship. The effect of pack size on price is also pronounced, with a dozen costing, on average, no more than 10 times as much as a single in the Australian data and just 7 times for the US data.

To further illustrate the relative results between the alternative models, Figures 1 and 2 give the results for the Australian data, and Figures 3 and 4 show the corresponding results for the US data. It can be seen that the first-order and higher order models show fairly similar patterns of price-response to package size (volume in millilitres) and pack size (units) for each data set. However, the size of the effect does appear to differ between the US and Australian data with considerably larger discounts available for quantity in the former. This indicates that, perhaps unsurprisingly, the nature of the market may be important in determining the pricing function.

For the Australian data we were able to investigate the extent to which the pattern of pricing varies over supermarket chains. There are 111 stores and four chains. We estimated the first-order model for data on each chain and coefficients on package size were 0.6095, 0.6091, 0.5805, and 0.5433. For pack size we estimated coefficients of 0.9567, 0.9514, 0.8985 and 0.8592. This spread of coefficient estimates, which is statistically significant, further implies that pricing patterns may not be systematic across retailers.

Assuming that our estimates are a close approximation to the appropriate adjust-
ment, then applying the correct quality adjustment in the case of changes in pack or package size will make a significant difference to the price relatives. Consider an example where we move from pricing a 1 litre soft drink to a 2 litre variety; denote the respective prices as $p(1)$ and $p(2)$. Then in order to construct a price relative which holds quality constant we need to either divide the price of the 2-litre variety by the estimated quality difference, $\hat{p}(2)/\hat{p}(1)$, or multiply the price for a 1-litre by the inverse of this ratio, i.e. the quality adjusted price relative is,

$$\frac{p(2)}{\{\hat{p}(2)/\hat{p}(1)\}} = \frac{p(2)}{p(1)/\{\hat{p}(1)/\hat{p}(2)\}}$$

(13)

For the Australian data, from table 3 the estimated quality difference between a 2-litre and a 1-litre bottle is in the range of 1.69 and 1.50, this compares with the linear adjustment technique which supposes that quality doubles when package size doubles. In this case the difference between our quality adjustment method and the one-for-one linear approach is that for the former the price relatives are higher by between 18.34% ($=2/1.6911-1$) and 33.1% ($=2/1.5031-1$). The equivalent figures for the US results are 44.0% and 47.7%. These are large differences and have the potential to introduce substantial errors into an index calculation. Note, however, that if package size decreases then the errors reverse sign so it is only if there is tendency for container size to change in one direction that a systematic bias will emerge.

6. Conclusion

Our primary objective has been to investigate empirically the relationship between package size, number of units per pack, and price. We outlined a model based on the hedonic regression technique and aimed at estimating the marginal utility from different package and pack sizes.

Using scanner data on soft drinks for the US and Australia, we find that prices, package and pack size do not exhibit a simple one-for-one relationship. That is, the price of a good does not double if the package size doubles or the number of units in a pack doubles. In fact we found that the relationship between price and package size was significantly flatter than this. This has important implications for the methods which are used to construct official price indexes, such as the Consumer Price Index. While the sign of the error introduced into the index will depend on whether package and pack sizes are increasing or decreasing between sampling periods, we have shown
that the difference can potentially be large. For this reason, the recommendation of this paper is that statistical agencies abandon, where possible, the procedure of scaling prices by changes in the package size or the number of packs on a one-for-one basis.

There are two main alternatives. First, the prices are excluded from the index altogether. This could reduce bias but increase variance if fewer observations are available for use. Second, econometric methods, such as those outlined in the paper, could be used to inform package and pack size adjustments, particularly if scanner data is available with regular updates. Although there are relatively small differences in results between our alternative methods, this may not always be the case. We recommend that relatively flexible functional forms are to be preferred. It is hoped that this paper has demonstrated that the econometric option is viable.
References


