Room document at the Ottawa Group Meeting, 2011

Non-linear Pricing and Price Indexes: Evidence and Implications from Scanner Data

Kevin J. Fox*
School of Economics and CAER
University of New South Wales
Sydney, NSW 2052
Australia
Email: k.fox@unsw.edu.au

Daniel Melser
School of Finance and Economics
University of Technology, Sydney
Sydney, NSW 2008
Australia
Email: daniel.melser@uts.edu.au

February 2011

Abstract

Non-linear pricing, the fact that prices do not necessarily change in proportion to size, is a ubiquitous phenomenon. However, it has been neither particularly well understood nor well measured. Non-linear pricing is of practical importance for statistical agencies who, in constructing price indexes, are often required to compare the relative price of a product-variety of two different sizes. It is usually assumed that prices change one-for-one with package and pack size (e.g. a 1-liter cola costs half as much as a 2-liter bottle). We question the wisdom of such an assumption and outline a model to flexibly estimate the price-size function. Applying our model to a large US scanner data set for carbonated beverages, at a disaggregated level, we find very significant discounts for larger-sized products. This highlights the need to pursue methods such as those advocated in this paper.

Keywords: Hedonic regression, non-linear pricing, quantity discount, local regression.
JEL Classification Codes: C43, C50, D00, E31.

*Corresponding author. We are appreciative of the financial support of the Australian Research Council (LP0667655) in undertaking this research and useful comments we received from participants at the UNSW Economic Measurement Group Workshop. Any remaining errors are our own.

1 Introduction

In a modern economy there is a dizzying array of consumer-products. These products are invariably available in a wide range of flavors, colors, brands, grades as well as package sizes—from mini-cans, bottles, boxes or bags to extra large and family sizes—and pack sizes—singles, doubles, half-dozens, dozens and mega-packs. This proliferation in the available product varieties has posed problems for the construction of price indexes on a number of fronts. In particular, as the product varieties change and evolve so must the indexes which purport to measure fluctuations in their prices. Comparing items on a quality-adjusted basis has become a major challenge for statistical agencies. This quality adjustment problem, of comparing two different items, is most often encountered in high-tech sectors where new and improved products replace older outdated varieties. However, it is also apparent in the more mundane product categories stocked by supermarkets and drug stores. Indeed, while much of the academic literature on quality change and price indexes has focussed on the introduction of new goods, (e.g. Bils, 2009; Hausman, 2003; Pakes, 2003; Diewert, 1998) one of the most common quality adjustment problems faced by statistical agencies is in comparing the prices of the same good sold in different sizes. For example, if the price of a 2-liter bottle of cola is collected in one period, but in the following period a 1-liter price quote is available. How should these prices be compared and hence inflation estimated in such a case? Should the price of the former be halved? This raises the question of what the market relationship is between products of different sizes.

There is general interest, at both a theoretical and practical level, about the relationship between product price, package and pack size. The ubiquitousness of lower unit-prices for larger-sized products has spawned a large literature. Recent work in this area, such as Cohen (2008) and McManus (2007), has focused upon estimating a structural model of consumer and producer interaction. Much of the previous literature viewed the existence of size discounts through the prism of price discrimination. Our goal is somewhat different. We propose to focus on the reduced form of the hedonic function, as is common in the price index literature, and to empirically document the price-size nexus. This provides both a contribution to our understanding of size discounts but also outlines a novel and flexible approach to estimating the hedonic function which can be actively pursued by statistical agencies in improving the accuracy of their size adjustments and hence of their price indexes.

Our empirical focus is on the carbonated beverages product category. We use a large, highly granular, US scanner data set from Information Resources Inc. (IRI) to examine the relationship between price, package and pack size. This data set covers 50 markets (metropolitan areas) and the 24 months from the beginning of 2005 to the end of 2006. Carbonated beverages are interesting for a number of reasons. They are sold in a variety of package sizes, such as small 8oz and 12oz cans all the way up to large 2-liter and 3-liter bottles, while also being available in different pack sizes, such as 6-packs, dozens and the like. This is evident in our data. Table 1 shows the breakdown of the value of sales, as a proportion of total sales, by various pack and package sizes. Total sales over the two-year period amounted to $159.5 million. As can be seen, much of the sales are focused around the single 1-liter and 20oz packages. However, larger pack sizes, such as half-dozens and dozens, were also very widely purchased. This distinction between package and pack size, which is clear with regard to carbonated beverages, has not been made clear in the literature thus far but is an important one for statistical agencies as the adjustments required along these two size dimensions may differ. The carbonated beverages product category is also interesting because it is a relatively significant expenditure item for consumers. It constitutes around one-third of one percent of the US CPI.[1] Because carbonated drinks is an important expenditure category it is often the focus of competitive pricing strategies amongst supermarkets such as being a loss leader (Lal and Matutes, 1994; Huang and Lopez, 2009).

Table 1: Value of Sales by Package and Pack Size (% of Total)†

	Pack Size						
Package Size	1	4	6	12	18	24	Total
8oz/237ml	0.1013	0.0406	1.4549	0.0666	0.0174	0.0011	1.6819
12oz/355ml	0.6125	0.3548	2.7882	48.5913	0.0471	8.6243	61.0182
14oz/414ml	0.1094	0.0000	0.0032	0.0184	0.0023	–	0.1333
16oz/473ml	0.0134	0.0223	0.6107	0.0009	–	–	0.6473
20oz/591ml	6.1494	0.0070	3.1745	0.2166	–	0.0000	9.5475
24oz/709ml	0.0796	–	3.9681	0.0040	–	–	4.0517
33.8oz/1l	2.0529	0.0000	–	0.0263	–	–	2.0792
50.7oz/1.5l	0.4553	–	–	–	–	–	0.4553
67.6oz/2l	19.6175	0.0032	–	–	–	–	19.6207
101.4oz/3l	0.7650	–	–	–	–	–	0.7650
Total	29.9563	0.4279	11.9996	48.9241	0.0668	8.6254	100.0000

†The values for pack and package size indicate intervals with respect to the previous value. For example for package size the intervals are; 0oz to 8oz, greater than 8oz and less than 12oz, etc.

In the following section we outline the problems faced by statistical agencies in quality-

[1] The US city average weights for December 2009 give the carbonated drinks expenditure category a weight of 0.294% for the CPI for All Urban Consumers and 0.380% for the CPI for Urban Wage Earners and Clerical Workers.

adjusting for changes in pack and package sizes. Section 3 outlines some of the theory around size discounts and formulates a flexible local hedonic model for estimating these effects. In section 4 we apply these methods to the IRI scanner data set on carbonated beverages and outline the results. We conclude by discussing some of the implications of these results for the construction of price indexes.

2 Price Indexes and Size Adjustments

The standard approach used by statistical agencies when comparing the same product of different size is to make the prices comparable by multiplying one of them by the appropriate size ratio (see for example, Armkneckt and Maitland-Smith, 1999; Triplett, 2004; ILO, 2004). There are at least three situations where this approach may be applied.

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 recorded. When this is not available it is often the case that another size is recorded and the price is automatically size-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 product. For example, a manufacturer may decide to increase the size of a bottle of cola from 2-liters to 2.25-liters. If the price of the 2-liter variety was originally collected then this poses a comparability problem with the price in the later period. A common approach adopted by statistical agencies is to scale the price of the new item by the difference in package sizes—in our example the scale factor would be 0.8889=2/2.25. This adjusted price is then directly compared with the price for the 2-liter product in the earlier period in estimating inflation.

Finally, retailers sometimes undertake promotional offers such as 'buy-one-get-one-free' (BOGOF). 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 by statistical agencies to 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 essentially treated as being free. A second, more common, approach in this case is for the statistical agency to halve the price of the good in the later period. This 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 upon price indexes. Implicit in statistical agencies adjustments for changes in pack and package size is the assumption of a linear relationship between these factors and price. As noted by Triplett (2004), it is certainly 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... (Triplett, 2004, Chapter 1, p. 20,)

If price-per-unit varies then assuming a linear relationship will give incorrect, and potentially biased, measures of price inflation if there are systematic changes in size over time. This raises the important question of how this problem should be dealt with in the construction of official price indexes.

The ILO manual on consumer price indexes (CPIs) (ILO, 2004) echoes Triplett (2004) and cautions against the mechanical application of size-based 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 characteristic. 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. (ILO, 2004, p. 29)

While this is certainly sound advice, it is unclear what type of 'common sense' adjustment rules should be applied in practice.

Given the current practices adopted by statistical agencies, the equivocal stance of the ILO CPI manual, and the fact that very little empirical research has been undertaken in this area, there is clearly scope for further investigation. Of primary interest is determining the relationship between price and size. We proceed to do exactly this in our empirical application below but first outline a model of the relationship between price, pack and package size.

3 Modeling Price, Pack and Package Size

In developing a model of the price-size relationship it is important to have an idea of the likely drivers of nonlinear pricing. These will be important in determining the structure of the model, and the degree of flexibility that is required in isolating the function of interest. The literature on non-linear pricing has stressed price discrimination as a primary cause of such pricing policies. The first paper to rigorously address the issue was Spence (1976) but the literature has grown considerably since this time. If the producer has some sort of pricing power, and consumers differ, then it may be possible that a non-linear pricing schedule will separate the consumers and lead to higher firm profits. Maskin and Riley (1984) showed, for a simple utility function and consumers with heterogenous preferences, that the optimal pricing function for a monopolistic firm will have declining unit costs.

Size discounts may also arise from differences in the marginal costs of production. Clements (2006) emphasized, amongst other factors, that the cost of the packaging may be lower for larger-sized products. Consider the example of bottled drinks. A 2-liter bottle has twice the volume of a 1-liter bottle but the surface area of the package is likely to increase by only 40-50%. Yet the cost of packaging is unlikely to be large enough to have a major influence upon price. Another area where there may be cost economies of size is in retailing, transport, advertising and storage. However, it is likely that the costs of selling an item—processing it at the checkout—as well as transporting, storing, and stocking the shelves with it, will be almost independent of pack or package size.

If we take the consumer perspective, rather than that of the firm, we note that there may be greater storage or transport costs from purchasing a larger package or pack size. Hence consumers may need to be compensated for this by paying a lower price per unit. 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 bottle of carbonated softdrink then they may only consume part of the contents in the first sitting. There is a risk that the remainder will spoil. This would lower the value of larger package sizes to consumers. Yet contrary to this intuition, Gertsner and Hess (1986) have argued that consumers will save shopping time, and could have lower overall transport costs, by purchasing in bulk. Hence it is possible that we may see premiums, rather than discounts, for larger products.

Recent work by McManus (2007) and Cohen (2008), focusing upon coffee and paper towels respectively, has taken a more unified approach and specified and estimated a structural model of both consumption and production. Consumers make discrete choices over different products of various size and producers sell at some markup over marginal costs. This approach helps to understand the structural factors generating non-linear pricing regimes but goes much further, and makes many more assumptions, than is required for price index construction. What is needed by statistical agencies is some estimate of the relative prices of products of different package sizes so that prices can be appropriately adjusted if necessary.

3.1 A Flexible Hedonic Model of the Price-Size Function

We propose to examine the question of non-linear pricing within a hedonic framework. A hedonic pricing model can be justified on a consumer basis, as under certain conditions representing consumer preferences (see Diewert, 2003a). Or alternatively it can be motivated from a producer perspective as reflecting the marginal costs of production as well as markups (Pakes, 2003). More generally, as Rosen (1974) noted in his seminal work on the subject, the hedonic function is a reduced form that is likely to reflect all aspects of the strategic interaction between buyers and sellers in a market incorporating both the effects of cost, technology, preferences and market power. From a price index perspective this is desirable. Our objective is not necessarily to estimate either producer technology or consumer preferences but to provide a simpler, more basic, representation of the budget constraint which prevailed at a specific time and/or place and which can be used in producing a quality-adjusted price index. We propose a flexible nonparametric hedonic regression method which represents the relationships of interest succinctly but with a suitable degree of fidelity to the data.

In the empirical investigation which follows our highly granular scanner data set enables us to control for the attributes of each product and isolate the impact of pack and package size on price. In particular, we have data on prices p_{itm} for varieties $i = 1, 2, \ldots, I$ over time periods $t = 1, 2, \ldots, T$ and markets $m = 1, 2, \ldots, M$. The attributes which are likely to determine the price of this product are its pack (x_i^A) and package size (x_i^B)—the primary objects of interest—the market and time period in which it was sold and the particular features of the specific product—taste, flavour, colour, content and so forth.

Turning to the latter price-driver first, we define a brand-type indicator variable such that $d_{i\iota} = 1$ when product i is of brand-type $\iota = 1, 2, \ldots, L$ and zero otherwise. A brand-type is a combination of a particular brand and a product type. For example, 'Pepsi' is a brand but 'caffeine-free Pepsi', 'diet Pepsi'and 'Pepsi vanilla' are some examples of brand-types. Including a dummy variable for each such brand-type tightly controls for any price effects due to the nature of the product.

In terms of the time and market effects, we estimate a separate regression for each combination of time and market. This is because the preceding discussion, particularly with regard to the price discrimination motive for non-linear pricing, has emphasized just how much the price-size function may depend upon 'local' factors. Moreover, the stability of this function across time and space is of some practical importance to statistical agencies in terms of identifying the extent to which common quality adjustment ratios can be used or whether individual markets and time periods require tailored adjustments. Given this, and denoting the general pack size and package size function as $s(\ln x_{itm}^A, \ln x_{itm}^B)$, adding a random error term, e_{itm}, and hypothesizing a log-linear functional form, as advocated by Diewert (2003b), we have the following hedonic model:

$$\ln p_{itm} \quad = \quad \sum_{\iota=1}^{L} \delta_\iota d_{i\iota} + s(\ln x_{itm}^A, \ln x_{itm}^B) + e_{itm}, \qquad i = 1, 2, \ldots, I. \tag{1}$$

We want to model the size function flexibly while also aiding interpretability. Our solution is to use a simple functional form but to estimate it 'locally' by way of a local regression estimator (Cleveland, 1979). In particular we assume that $s(\ln x_{itm}^A, \ln x_{itm}^B)$ is a linear function of its components but we estimate this function for each point in the product space. This helps to reveal the local behavior of the price-size surface. At each data point a weighted regression is estimated with the weights determined by the size-distance from the reference point. Hence points which are neighboring to the point of interest in terms of pack and package size get the highest weights and far away points get much lower weights. The exact weights used in our application are those of the tricube function. Here, if itm is the reference observation then the weight for observation jtm, represented by $w_{jtm|itm}$, is proportional to $w_{jtm|itm} = \left(1 - \left(\frac{d_{jtm|itm}}{d^*_{itm}}\right)^3\right)^3$. Here $d_{jtm|itm}$ is the distance of the observation from the reference point and d^*_{itm} is the maximum such distance. The extension to two dimensions is straightforward following standardization.

The advantage of this approach is that we can approximate a potentially very complex function by using a simple functional form and estimating it locally. It also enables us to easily obtain estimates of the primary object of interest, the gradient of the relationship between price and package size, and price and pack size:

$$\frac{\partial \ln p_{itm}}{\partial \ln x^C_{itm}} = \frac{\left(\frac{\partial p_{itm}}{p_{itm}}\right)}{\left(\frac{\partial x^C_{itm}}{x^C_{itm}}\right)} = \frac{\partial s(\ln x^A_{itm}, \ln x^B_{itm})}{\partial \ln x^C_{itm}}, \qquad C = A, B. \tag{2}$$

This is a particularly informative quantity as if the price-size effect were linear we would expect the price-size elasticity to be equal to one. In the empirical section below we will examine this quantity across different markets and time periods and draw some conclusions about the extent of non-linear pricing.

4 An Empirical Examination

We apply our model to the carbonated beverages scanner data set made available for academic research purposes by IRI (Bronnenberg, Kruger and Mela, 2008). The data set is extremely rich and detailed. In our analysis we focus on data across 50 markets—essentially a mix of both large and small metropolitan areas[2]—across more than 2000 stores from 116 chains in 2005 and 2006.

It has unit prices for each Universal Product Code (UPC, i.e. barcode)—of which there are 5,304—by store at a weekly frequency. We aggregate across weeks to create monthly price observations, and across stores to create chain average prices for each market. The

[2]The markets are: Atlanta, Birmingham/Montgomery, Boston, Buffalo/Rochester, Charlotte, Chicago, Cleveland, Dallas, Des Moines, Detroit, Eau Claire, Grand Rapids, Green Bay, Harrisburg/Scranton, Hartford, Houston, Indianapolis, Kansas City, Knoxville, Los Angeles, Milwaukee, Minneapolis/St. Paul, Mississippi, New England, New Orleans, New York, Oklahoma City, Omaha, Peoria/Springfield, Philadelphia, Phoenix, Pittsfield, Portland, Providence, Raleigh/Durham, Richmond/Norfolk, Noanoke, Sacramento, Salt Lake City, San Diego, San Francisco, Seattle/Tacoma, South Carolina, Spokane, St. Louis, Syracuse, Toledo, Tulsa, Washington DC, West Texas/New Mexico.

standard calculation frequency for price index construction is monthly so this a natural time unit in our context. Aggregation across stores within a market, to chain-level average prices, seems reasonable given the homogeneity of chain pricing found by Hwang, Bronnenberg and Thomadsen (2010), who examined the same data as we use. Our aggregation leads to a data set of manageable proportions, though it still yields a total of 2,528,533 observations spread over the 50 markets and 24 months. This gives an average of 2,107 observations per market and month. Using the available brand descriptions we define 351 different brand-types for which dummy variables are included in the model.

Using the nonparametric local regression approach we estimate the effects of pack and package size on price. In this estimation we are required to choose the smoothing level, in our case the proportion of observations which will fall within the local regression window. The fact that the data are tightly centered around particular pack and package sizes, rather than scattered evenly across size-space, means that slight changes in location can lead to large changes in weights. After some experimentation the most stable and reliable estimates were obtained when we maximized the span, including all the data and let the weighting function down-weight observations which are further away, in terms of size, and up-weight those observations which are closer. This leads to a degree of localization in the estimated slopes without introducing undue volatility.

The outcome of our estimation is a pricing surface in package size and pack size for each of the 1,200 (=50×24) market-months. As a way of summarizing these results we construct prices normalized on a single 1-liter bottle and average across all the different markets and time periods. The results are shown in Table 2, for certain pack and package sizes of interest, and are depicted in Figure 9 graphically.

These results indicate very significant discounts for size particularly along the package size dimension. For example, an 8oz bottle of soda costs 89.32% of a 1-liter despite being less than a third of the size, while a 2-liter bottle of soda costs only 1% more than a 1-liter on average. These indicate very significant discounts for quantity along the package size dimension. Moreover, there is the tantalizing result that, on average, a 3-liter bottle actually costs *less* than an equivalent 2-liter bottle. Not only is the package size function not linearly increasing, it is sometimes not even increasing.

Along the pack size dimension there are also large discounts available, though not quite as significant. A 6-pack of the widely sold 12oz package size costs, on average, only 1.95 times more than a single serving while a 24-pack costs just 3.60 times as much as a single serving. From a statistical agency perspective our estimated price-size function implies that linear price adjustment is likely to get the quality adjustment wrong, and significantly so.

In terms of implementing appropriate quality adjustment for changes in size Table 2

Table 2: Price, Pack and Package Size†

	Pack Size					
Package Size	1	4	6	12	18	24
8oz/237ml	0.8932	1.5515	1.8377	2.4680	2.9470	3.3468
12oz/355ml	0.9485	1.6455	1.9490	2.6303	3.1580	3.6006
14oz/414ml	0.9633	1.6825	1.9948	2.6978	3.2417	*3.6982*
16oz/473ml	0.9742	1.7055	2.0247	2.7482	*3.3101*	*3.7816*
20oz/591ml	0.9867	1.7420	2.0763	2.8363	*3.4229*	3.9131
24oz/709ml	0.9926	*1.7721*	2.1230	2.9092	*3.5041*	*3.9986*
33.8oz/1l	1	1.8015	*2.1663*	2.9843	*3.6001*	*4.1084*
50.7oz/1.5l	1.0064	*1.8142*	*2.1801*	*2.9950*	*3.6029*	*4.1006*
67.6oz/2l	1.0101	1.8231	*2.1911*	*3.0077*	*3.6134*	*4.1069*
101.4oz/3l	1.0097	*1.8314*	*2.2042*	*3.0292*	*3.6370*	*4.1288*

†Italics indicate that a pack-package size combination was unavailable in our data.

potentially provides a way forward. The relative prices in this table give some indication of the appropriate scalar for linking products of different size into the index. However, an examination of the more disaggregated results is instructive.

First to package size. In figures 1a to 3b we depict the price function over various package sizes, and its derivative, for a pack containing a single product for a selection of cities—Chicago, Indianapolis and San Diego—in each of the 24 time periods. We are only interested in relative prices, so the price-size function is normalized so as to be one for a 1-liter bottle in each period. It is clear that there is significant variability in the function even within a given city across time. This points towards the particular competition dynamics in a city, and in a given time period, having a major role in determining the momentary size pricing function. Nevertheless, there are some features which do stand out. In particular the fact that the price-package size function is steeper (i.e. has a larger derivative) at lower values. Though in almost every case we examined the derivative was significantly less than one. The prevalence of negative derivatives is also surprising.

Alternatively, if we hold time fixed and look across markets at the package size function then we get an impression of the significant diversity in city pricing functions. The last month in our data set, December 2006, is illustrative. Figures 7a and 7b show this diversity in pricing functions and the large number of markets where there are significant downward sloping portions of this function. Indeed, comparing the various charts, there appears to be somewhat more heterogeneity across cities for a given time period than across time for a given city.

We turning now to pack size and illustrate our results similarly. In Figures 4a to 6b we depict the pack size function for the markets, Atlanta, Boston and Los Angeles, across time. In each of these figures we fix the package size at 12oz. This was the most popular package size overall—see Table 1—and it was also available in every possible pack size. Figures 8a and 8b illustrate the diversity across markets by plotting the pack size function in January 2005, the first month in our data set, for all of the 50 markets.

The biggest contrast between the package size and pack size functions is that the latter never has a negative slope. That is the cost of a large pack is always more than a smaller one. But this slope is certainly far from linear. Though it is generally closer to one than was the case for package size. There is also less apparent heterogeneity in the pricing function for packs, compared with packages. These results are likely to be the result of the greater scope that consumers have to 'unbundle' multipacks compared with different package sizes. This may lead to a higher degree of substitution between different pack sizes than between different package sizes and hence a higher degree of pricing homogeneity for the former.

5 Conclusion

Our primary objective has been to investigate empirically the relationship between package size, pack size, and price with reference to the practices of statistical agencies. Using a flexible hedonic regression framework our empirical results indicate that prices, package and pack size do not exhibit a simple one-for-one relationship. That is, the price of a product does not double if the package size doubles or if 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. In the case of package size it was sometimes even negative.

This has important implications for the methods which are used to construct official price indexes. While the sign of the error introduced into the index will depend on whether package and pack sizes are increasing or decreasing, it is clear that the difference can potentially be large enough to contaminate index comparisons. For this reason the recommendations of this paper are that statistical agencies discontinue 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 in constructing the index. In some cases, where there are widespread package size changes introduced by a major manufacturer, exclusion may not be possible as it would leave the agency with too few observations. Second, hedonic regression methods could be used to inform package and pack size adjustments. Our results indicate that the hedonic function needs to be estimated flexibly across different markets and time periods. This article has outlined a methodology for pursing this second option and demonstrated that it is certainly viable.

6 References

Armknecht, P. A. and F. Maitland-Smith (1999), "Price Imputation and Other Techniques for Dealing with Missing Observations, Seasonality and Quality Change in Price Indices," International Monetary Fund, Statistics Department, WP/99/78.

Bils, M. (2009), "Do Higher Prices for New Goods Reflect Quality Growth or Inflation?," *Quarterly Journal of Economics* 124(2), 637-675.

Bronnenberg, B. J., M. W. Kruger and C. F. Mela (2008), "Database PaperThe IRI Marketing Data Set," *Marketing Science* 27(4), 745-748.

Clements, K. W. (2006), "Pricing and Packaging: The Case of Marijuana," *The Journal of Business* 79, 20192044.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association* 74(368), 829836.

Cohen, A. (2008), "Package Size and Price Discrimination in the Paper Towel Market," *International Journal of Industrial Organization* 26(2), 502-516.

Diewert, W. E. (1998), "Index Number Issues in the Consumer Price Index," *Journal of Economic Perspectives* 12(1), 47-58.

Diewert, W. E. (2003a), "Hedonic Regressions: A Consumer Theory Approach," in Scanner Data and Price Indexes, *Review of Income and Wealth Conference*, Vol. 64, Feenstra, R. C. and Shapiro, M. D. (eds.), National Bureau of Economic Research, University of Chicago Press, 317-348.

Diewert, W. E. (2003b), "Hedonic Regressions: A Review of Some Unresolved Issues," Paper Presented at the Seventh Meeting of the Ottawa Group, Paris, France.

Gertsner, E. and J. D. Hess (1986), "The Effect of Consumer Heterogeneity on Package Size and Unit Price Variations," Working Paper No. 66, Department of Economics and Business, North Carolina State University.

Hausman, J. (2003), "Sources of Bias and Solutions to Bias in the Consumer Price Index," *Journal of Economic Perspectives* 17(1), 23-44.

Huang, R. and R. A. Lopez (2009), "Are you are what you eat? Overweight Status and Soft Drink Choices," Selected Paper prepared for presentation at the Agricultural and Applied Economics Associations 2009 AAEA and ACCI Joint Annual Meeting, Milwaukee.

Hwang, M., B. J. Bronnenberg and R. Thomadsen (2010), "An Empirical Analysis of Assortment Similarities Across U.S. Supermarkets," *Marketing Science* 29(5), 858-879.

ILO (2004), *Consumer Price Index Manual: Theory and Practice*, compiled by ILO in conjunction with the Technical Expert Group on Consumer Price Indices.

Lal, R. and Matutes, C. (1994), "Retail Pricing and Advertising Strategies," *The Journal of Business* 67(3), 345-370.

Maskin, E. and J. Riley (1984), "Monopoly with Incomplete Information," *Rand Journal of Economics* 15(2), 171-196.

McManus, B. (2007), "Nonlinear Pricing in An Oligopoly Market: The Case of Speciality Coffee," *The RAND Journal of Economics* 38(2), 512-532.

Pakes, A. (2003), "A Reconsideration of Hedonic Price Indices with an Application to PC's," *American Economic Review* 93(5), 1578-1596.

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

Spence, M. (1976), "Nonlinear Prices and Welfare," *Journal of Public Economics* 8(1), 1-18.

Triplett, J. E. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, Prepared for: Directorate for Science, Technology and Industry, OECD.

Figure 1: Package Size Function (Pack=1) – Chicago

(a) Price Level (b) Derivative

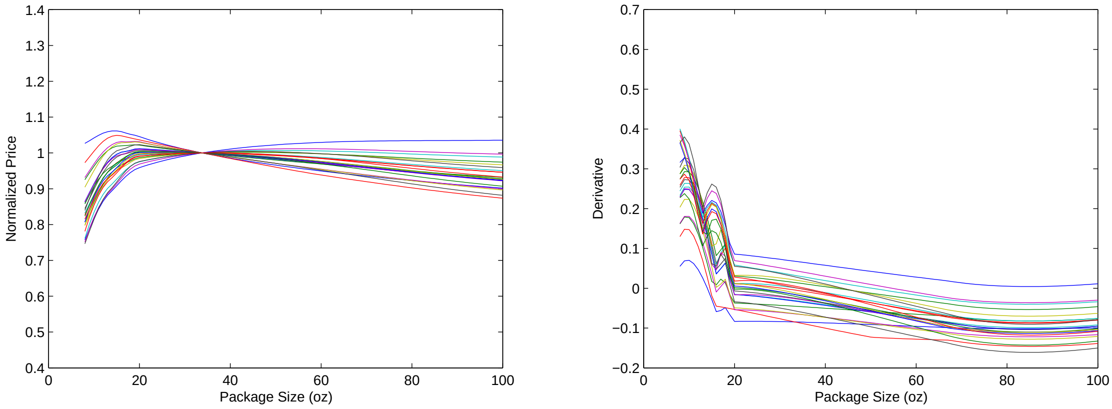


Figure 2: Package Size Function (Pack=1) – Indianapolis

(a) Price Level (b) Derivative

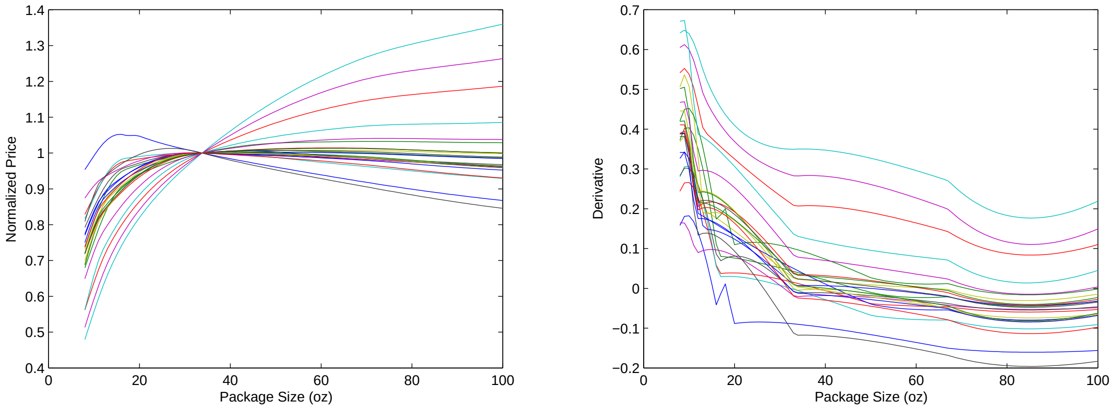


Figure 3: Package Size Function (Pack=1) – San Diego

(a) Price Level (b) Derivative

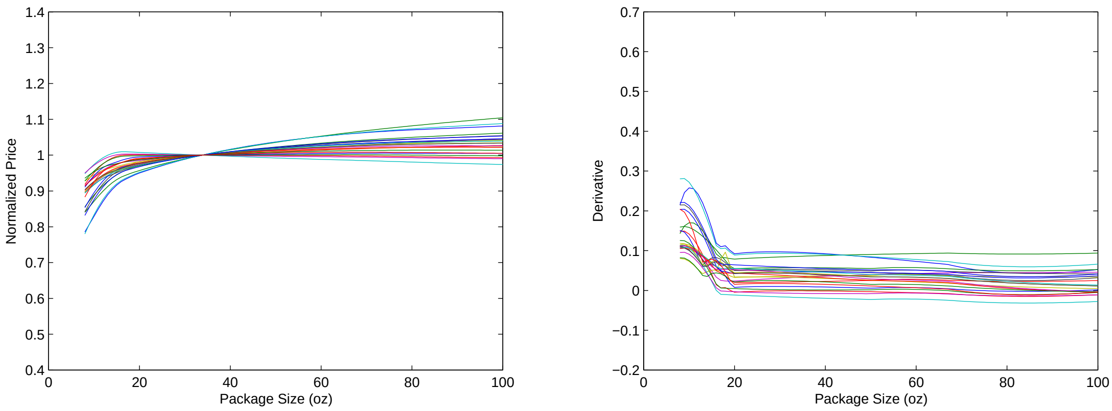


Figure 4: Pack Size Function (Package=12oz) – Atlanta

(a) Price Level

(b) Derivative

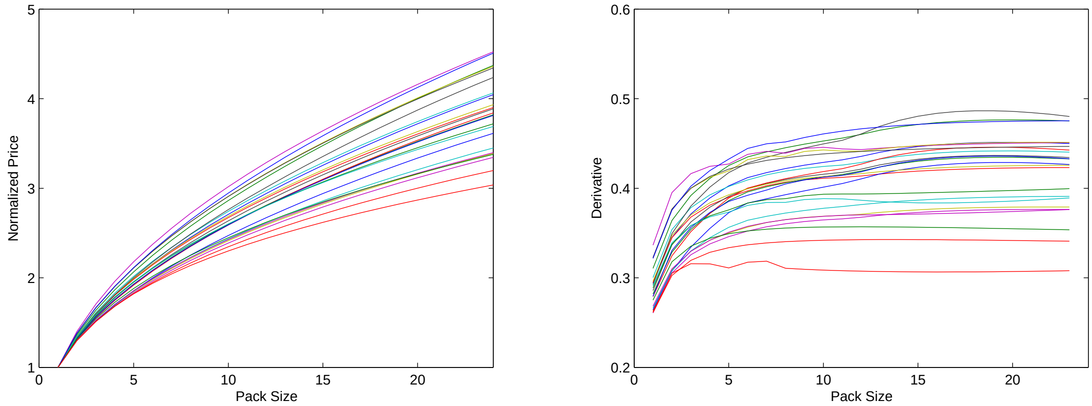


Figure 5: Pack Size Function (Package=12oz) – Boston

(a) Price Level

(b) Derivative

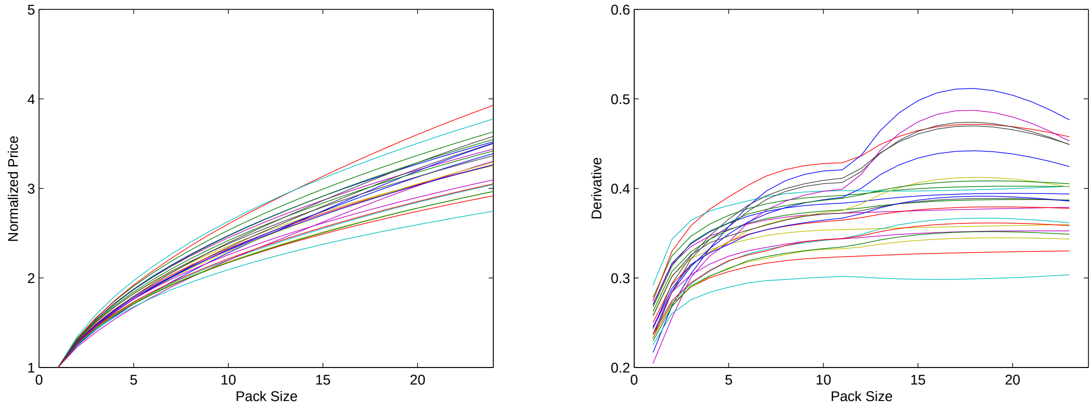


Figure 6: Pack Size Function (Package=12oz) – Los Angeles

(a) Price Level

(b) Derivative

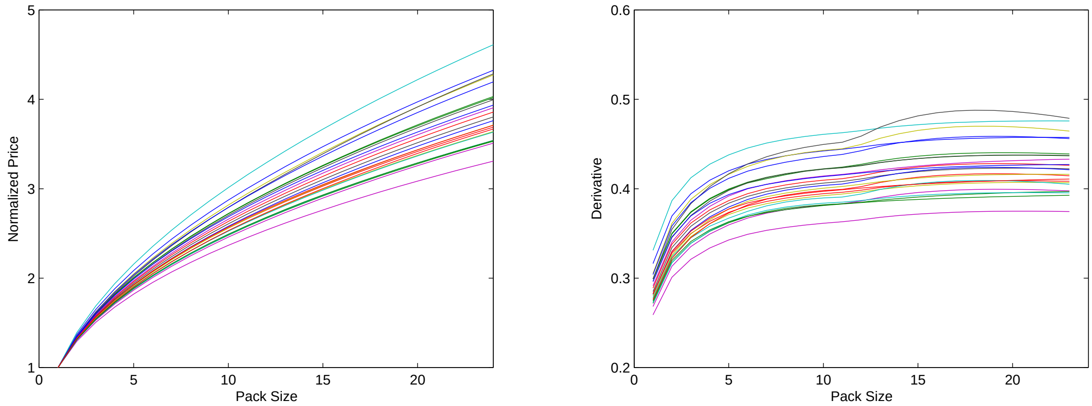
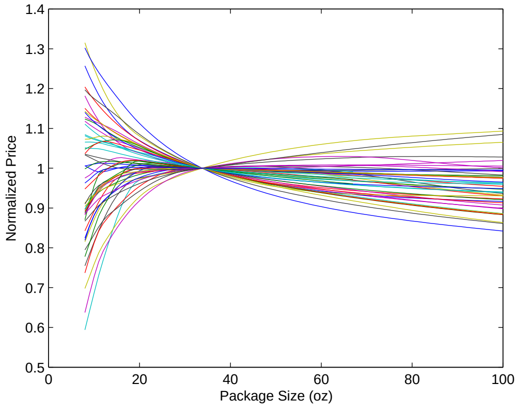


Figure 7: Package Size Function (Pack=1) – December 2006

(a) Relative Price



(b) Derivative

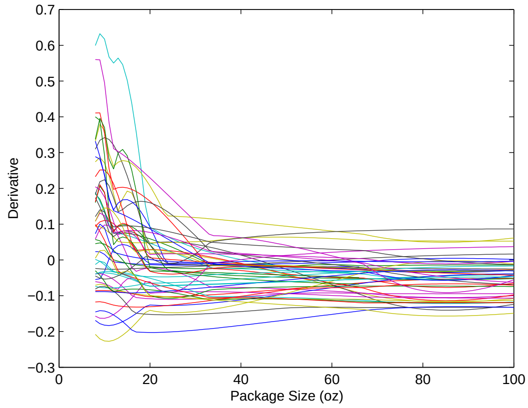
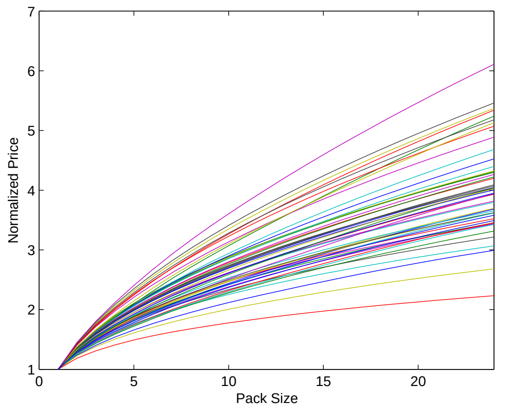


Figure 8: Package Size Function (Package=12oz) – January 2005

(a) Relative Price



(b) Derivative

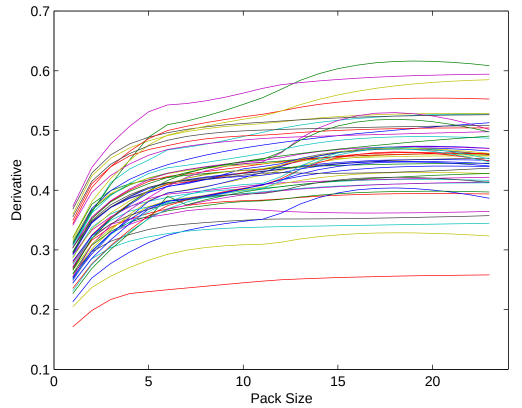


Figure 9: Size Surface – Average of All Time Periods and Markets