RETAIL FOOD PRICES: SOME INSIGHTS FROM SUPERMARKET SCANNER DATA

Tim Lloyd
School of Economics, University of Nottingham

Revised version of paper presented 12 March 2013.

Abstract

The advent of the barcode and laser scanning technology provides a potentially rich source of data (so-called, scanner data) on the purchases of the nation's food consumers. Using scanner data obtained from all the major UK supermarkets, this paper offers a glimpse at the prices of some of the purchases that make up the nation's shopping trolley. The price data analysed belong to over 500 barcode-specific products recorded at weekly intervals over a two and a half year sample period in the largest seven national retail chains, giving nearly a quarter of a million prices in all. Characteristics of the prices are reviewed with an eye to their dispersion over time and across retailers and this provides insights into pricing strategy and the nation's cheapest retailer. The data also allow the extent, magnitude and duration of promotional discounting ('sales') to be explored in greater detail than has hitherto been possible in the UK, and a summary of findings will be provided, some of which may even inform the way you shop.

Introduction

We all eat food. Some of us even buy it, and when we do it is likely that we do so at one of the national grocery retail chains that have come to dominate the landscape of food retailing in the UK; supermarket chains now accounting for around 80% of retail expenditure on food. Accompanying this concentration of the nation's food purchases has been the introduction of barcode laser scanning,

a technology that records precisely what we buy at supermarket check-outs. While primarily used for stock control and in-house marketing, the base data are also made available to market research organisations such as AC Nielsen, who process, summarize and analyse them for all manner of clients, and it is a sample of these Nielsen data, that this paper explores. Specifically, the dataset contains the prices of around 500 everyday food products, recorded at weekly intervals over a two-and-a-half year sample period in the largest seven national retail chains. The list includes (manufacturer) branded products and (supermarket) private labels and amounts to around a quarter of a million prices in all. Large though this is, the data merely offer a glimpse at the prices confronting the modern food shopper, a typical supermarket stocking in excess of 25,000 products.

Despite the current ubiquity of the barcode, economists have only quite recently had access to scanner data in the UK and elsewhere. The data are not simply novel, they represent a new kind of data since they relate to highly detailed (barcode-specific) prices of products that consumers actually buy, rather than the broad aggregates that have traditionally been made available from authoritative sources, such as the Cost of Living and Food Survey (Office of National Statistics 2011). For example, a seemingly homogenous 'Bread' category, which is actually amalgamated into 'Bread, rice and cereals' in the CLFS, is actually represented by 583 separate time series in the dataset used here.

After a summary of the dataset, the paper highlights some key characteristics of the barcode-specific prices, with an eye to their dispersion over time and across retailers, a review which provides some insights into pricing strategy and the nation's cheapest retailer. One of the most interesting findings is the extent to which the price varies across the national retailers, despite the physically identical nature of the product, a fact that is at odds with textbook notions of the 'law of one price' for homogenous goods. The data also allow the extent, magnitude and duration of promotional discounting ('sales') to be explored in greater detail than has hitherto been possible. The reliance on sales is the most striking of the findings and while all retailers use them, they do so to varying degrees, thereby underlining that as far as food retailing is concerned, standard models comprising representative firms are not well suited to UK food retailing. These and other findings are presented to highlight new avenues of research and may even inform the way you shop.

A Dataset of Scanner Prices

The prices analysed in this paper have been obtained from *AC Nielsen*, a leading global market research company, who at the time of the sample (8 September

2001 to 17 April 2004) collated data at weekly intervals on products sold in the UK's seven largest supermarkets. As a group, these retailers accounted for around three-quarters of all food sold (independent retailers and smaller supermarket chains accounting for the remainder). The data set identifies products at a highly detailed level. In general, two products are distinct if they have different barcodes, so that 100 gram and 200 gram jars of the same brand of instant coffee are different products for which separate prices are recorded. Furthermore, many of the products are national brands that are sold by all retail chains, so the data set contains retailer-specific prices of identical products. We identify each retailer-product combination with a Unique Product Code (UPC), so that, for example, a 100 gram jar of Nescafe 'Gold Blend' instant coffee stocked by Tesco and Sainsbury are two separate UPCs each with their own time series of weekly prices. In all there are 1,704 such UPC price series, the distribution of which is summarised in Table 1. Products included in the sample are from 15 categories of food (Table 1). Data (percentage of data set) are most prevalent in the bread (34%), soup (18%), coffee (8%) and orange juice (6%) categories, each of which contains in excess of 100 UPCs. The least populated categories, such as frozen fish fingers (1%) and frozen pizza (1%), contain 20 UPCs each. As is evident from these figures, the data set does not fully reflect consumer spending on food (fresh fruit and vegetables are not part of the dataset since they do not carry unique barcodes) but the range of categories is nevertheless broad, spanning beverages and foods across a range of formats namely fresh, chilled, ambient and frozen.

As Table 1 also shows, seven categories contain products in both branded and private label forms. In the UK, where sales of private label products now account for around half of the total consumer spend on food, this dimension of the data set offers potential insights in to any differences between the pricing of manufacturer- and retailer-branded products. Private label products account for nearly one-fifth of the products listed in the data set.

One of the most interesting aspects of the data set is that prices are available for all the major grocery chains. These include the market leader, Tesco; the other mainstream retailers, Sainsbury and ASDA; some soft discounters, Safeway, Somerfield and Kwik Save and a luxury retailer, Waitrose. As a group they represent the spectrum of supermarket retailers in the UK during the sample period.[1] While not every product is stocked by all retailers, 64% (325/507) are sold in at least 2 retailers, and 18% sold in all seven. Overall, the distribution of products by supermarket and brand status is sufficiently even to prevent results from any one classification dominating the findings (See Lloyd et al., 2011 for details).

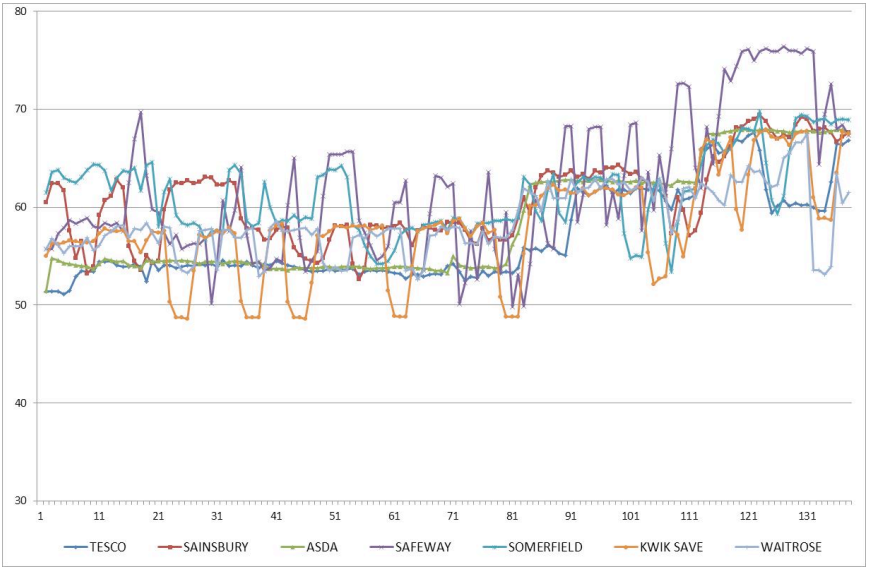
Table 1: Distribution of Unique Product Codes (UPCs) in the Sample by Category

Category	Brands	Private Label	All	% of total
Orange Juice	57	51	108	6.34
Instant Coffee	111	27	138	8.10
Tinned Tuna	51	0	51	2.99
Tinned Tomatoes	50	0	50	2.93
Tinned Soup	237	71	308	18.08
Oven Chips	83	0	83	4.87
Corned Beef	25	5	30	1.76
Frozen Peas	34	0	34	2.00
Fish Fingers	20	0	20	1.17
Breakfast Cereal	66	0	66	3.87
Tea Bags	59	8	67	3.93
Yoghurt	65	4	69	4.05
Wrapped Bread	488	95	583	34.21
Jam	33	44	77	4.52
Frozen Pizza	20	0	20	1.17
Total	**1,399**	**305**	**1,704**	**100.00**

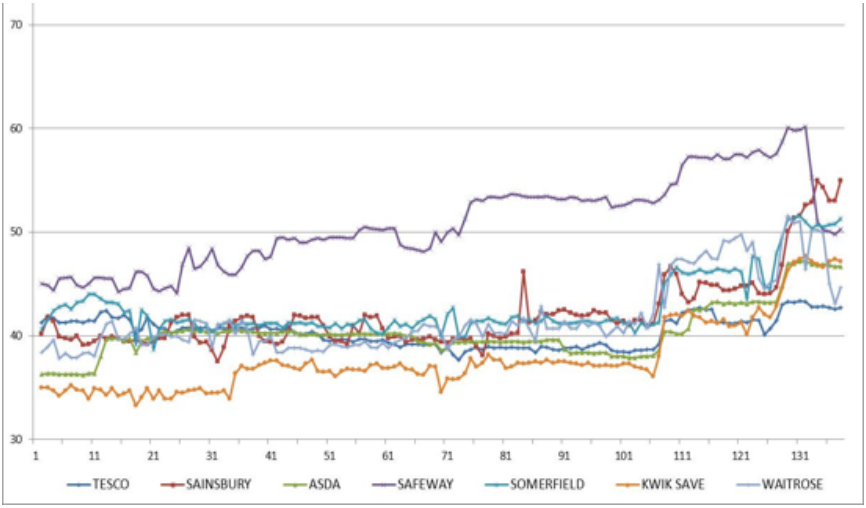
To give a flavour of the data, Figure 1 illustrates the prices of two products (a) a premium branded large white sliced loaf and (b) an everyday private label version, both of which are sold in all seven national retail chains, so in total the graphs display the prices of 14 UPCs. While representing just two of the 507 products in the dataset they exhibit a number of features that are common more broadly. In particular, sales appear to punctuate the price series, albeit with a frequency and intensity that varies by retailer (discounters using sales more frequently than others) and brand status (brands being promoted more than private labels). When not on sale, the price of each UPC tends to coalesce around particular levels, changing at discrete points, possibly reflecting shifts in raw material costs. Prices of the branded product command higher prices than the private label version (here 59p versus 39p) due to things like quality differences and packaging. However, it is also apparent that even the prices of an identically bar-coded product exhibit persistent and substantial differences, whether these be private label or the branded product. These features and the implications are analysed in more detail below.

Figure 1: The Price of a Loaf

(a) Kingsmill Everyday Topgrade 800g Medium Sliced White Loaf



(b) Private Label Everyday 800g Medium Sliced White Loaf



Some Stylised Facts

1) Scanner prices are noisy

In principle, prices may be expected to change due to changes in factors such as production costs and promotional activity. Indeed, one of the major attractions of this new kind of data is that it allows investigation of the importance of sales in the variability of food prices. Retail prices in general are thought to be 'sticky' on account of the significant costs (whether this be psychological or administrative) of amending price lists and aisle labels, particularly so in a supermarket that may contain anywhere between 25,000 and 40,000 products. So what we find is rather surprising. Based on all the prices in the dataset, statistical analysis reveals that the price of a UPC remains unchanged for just 2.4 weeks. Furthermore, the overwhelming majority of price changes are small; in fact, a little over half the price changes are less than one penny and declines are only slightly less common than price rises (47% versus 53%). In sum, prices tend to be 'noisy', that is to say they are characterised by small, transitory movements that are typically reversed. This observation is at odds with the precision of the EPOS monitoring (in which 100% of transactions are recorded) and the National pricing strategies that are widely held to be the norm in in the major supermarket chains in the UK (Competition Commission, 2000). To understand what is going on, we need to know how the data – the scanner prices – are constructed.

Owing to the vast number of individual purchases actually made, scanner prices rely on base data that has been averaged in some way. The Nielsen prices used in this study are no exception and are actually 'unit values' (or 'average revenue' prices) meaning that the price recorded in any given week for a specific (barcoded) product in a particular retailer represents the ratio of the total value of transactions for the product to the total quantity sold over the preceding week in that retailer. It turns out that this represents the average price weighted by the proportion of the units sold at each price. To the extent that retail chains operate national pricing strategies they do so within a store format, meaning there are separate price lists for convenience, supermarket and hypermarket formats, rather than a universal price across all outlets a retailer operates (Eales, 2012). The upshot is that while average revenue prices neatly reflect the relative importance of the prices that the product was sold at in the overall average, they are likely to vary when the prices and/or the composition of units sold at different prices change. In fact, average revenue prices may change even if the shelf prices in all of a retailer's stores do not, since all that is required to induce a change in the average revenue price is a change in the composition of purchases of the product in question. While this has the advantage of taking proper account of the range in prices that consumers face it also means that if

the frequency of a price change is the object of interest –as it is in the 'sticky prices' literature (*e.g.* Nakamura and Steinsson, 2013) it will be necessary to remove this noise from the price series prior to the analysis, and this is something that we now turn to.

ıı) Reference prices better reflect underlying movements

Prices convey information. In order to extract this information from scanner prices it is necessary to remove the noise, which as discussed above, is an artefact of their construction. Using a simple algorithm it is possible to generate artificial series, called *reference prices* that track the central tendency of the actual price data, thereby revealing the essential price behaviour that is of keen interest to economists. To illustrate, consider Figure 2 which homes-in on just one of the price series shown in Figure 1(a) - the ASDA price of the Kingsmill 800g medium sliced white loaf. Of the (100 or so) price changes during the sample only two are substantive, and it is these that form the reference price series for this UPC that is superimposed in the figure.

Figure 2: Reference Prices (The ASDA Price of Kingsmill 800g loaf)

Table 2 summarises the duration of actual and reference prices across various classifications of the dataset. Whereas actual prices 'last' (*i.e.* remain unchanged for) just over a fortnight, reference prices (which exclude price changes due to sales and the averaging process discussed above) last around four months. While

the inertial quality of reference prices is self-evident, what is more surprising is the variation in price durations that are observed across different classifications of the dataset. As the results in Table 3 illustrate, reference prices last longer for private label products than for national brands; frozen and tinned good last longer than perishable foods and most interesting of all is that reference prices of some supermarket chains are markedly longer than others. Specifically, Asda and Tesco prices last around twice as long as their counterparts in Somerfield and Safeway. Clearly, some supermarkets are better able to control their costs and keep prices stable than others.

Table 2: The Implied Duration of Actual Scanner Prices and Reference Prices (weeks)

	Actual Prices	Reference Prices
Overall	2.4	13.9
Retailer		
Asda	3.6	20.8
Tesco	2.9	20.8
Sainsbury	2.5	15.6
Kwik Save	2.2	13.9
Waitrose	2.0	15.6
Somerfield	1.9	11.4
Safeway	1.9	8.9
Brand Status		
Private Label	3.0	17.9
Brand	2.3	13.8
Product format		
Frozen	3.3	20.8
Tinned	3.0	20.8
Chilled	2.4	12.5
Ambient	2.2	14.6
Fresh	1.9	10.4

The implied duration of a price is the reciprocal of the frequency of price changes for the median UPC over the sample period in each of the various categorisations of the data. The reference prices used here are based on the modal price in a 12 week rolling window. Other commonly used algorithms suggest even greater inertia in reference prices. See Lloyd *et al.* (2013) for details.

iii) Sales

One potentially important source of price change is promotional discounting (what we refer to here as 'sales'). Products are not typically declared as being 'on-sale' in scanner datasets, so sale prices are identified from the price data alone. Given the clearly recognisable episodes of discounting evident in Figure 1, spotting sale prices is, in the main, a straightforward exercise. Table 3 reports summary statistics of the sales defined according to 10, 25 and 35 per cent thresholds. It shows that nearly 8% of prices are classed as 'on sale' using the 10% threshold, a figure that drops to 3.5% and 1.4% using the larger discounts. Thus while sales are clearly the exception to the normal rule of pricing, only very deep sales are rare. Interestingly, brands are promoted almost twice as frequently as private labels.

Table 3 : Summary Statistics of the Sales Data

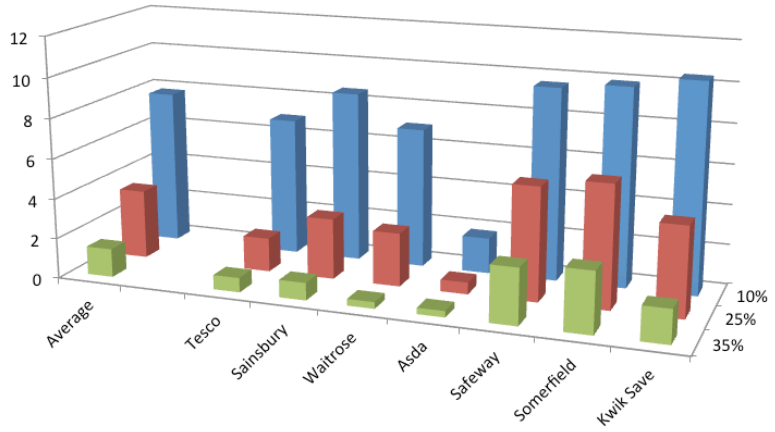
	Sale Threshold								
	10%			25%			35%		
	All	Brands	Private Label	All	Brands	Private Label	All	Brands	Private Label
Frequency (%)	7.8	8.5	4.6	3.5	3.8	1.9	1.4	1.5	0.9
UPCs (%)	63.0	66.9	44.9	36.8	49.6	23.9	20.1	21.4	13.8
Duration (weeks)	4.5	4.5	4.4	4.4	4.4	4.5	4.2	4.1	4.7

Table 3 also reports the proportion of time series that contain at least one sale episode and here the incidence of sales is more evenly distributed. Specifically, two-thirds of all UPCs have been on a 10% sale, one-fifth experiencing a deep (35%) sale. Taken together, the statistics suggest that sales are unusual but commonly applied across products. Of course, this characteristic is a familiar one, reflecting the role of sales in encouraging consumers to try new products. Interestingly though, around one-third of the series are never discounted. The figures in Table 3 also suggest that sales tend to be around four weeks long, irrespective of their depth.

Turning to differences in sales activity by retailer, Figure 3 shows the proportion of each retailer's prices that are sales prices under the three thresholds. There are quite marked differences across retailers: Asda uses sales rarely (almost one-tenth of the average) consistent with its reputation for 'every-day low-prices';

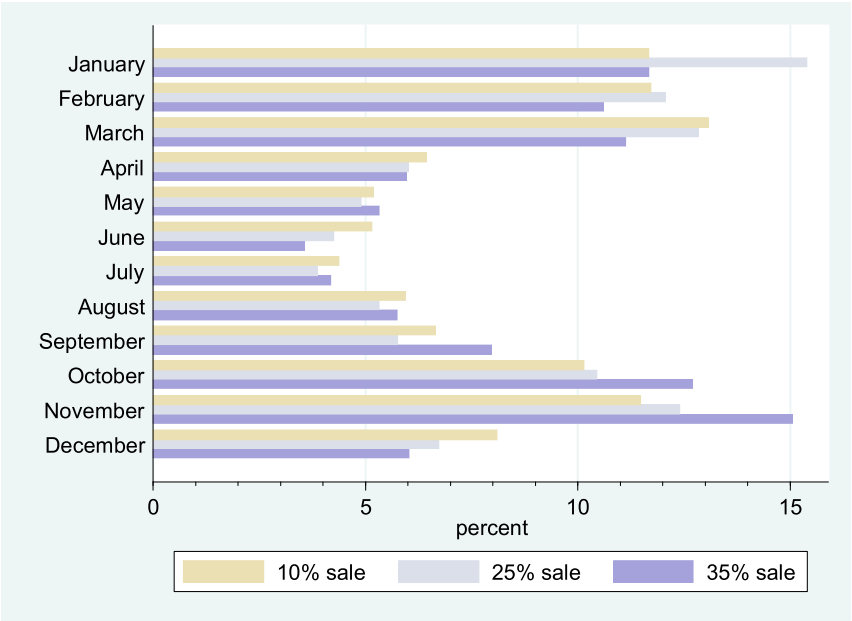
Safeway, Somerfield and Kwik Save form a group of discounters, in that the use of sales is above average; while the more mainstream chains, Tesco, Sainsbury and Waitrose are typical users of promotional sales activity. This classification is consistent across the depth of sales but becomes increasingly apparent the deeper the sale. All supermarket chains use deep sales sparingly, this being even less common for Tesco, Sainsbury Waitrose and Asda. As far as marketing strategy is concerned, the heterogeneity that is observed across retailers does not sit comfortably with the notion of a market comprising representative retailers. While this will come as no surprise to the shoppers among you, it is more easily overlooked in economic models of the food sector.

Figure 3: The Percentage of Prices that are Sales by Retailer and Sale Depth



Finally, I would like to mention one other characteristic of price dynamics - that is common across food retailing - that being the seasonality of promotional activity; sales being more prevalent in winter than in summer. Interestingly, a more nuanced picture (see Figure 4) emerges upon closer inspection, whereby the acceleration in the frequency of promotions, particular deep sales, in the months preceding Christmas is halted in December itself, when the use of sales is low. Promotional intensity resumes with the 'January sales' and remains high until Easter when it decline sharply, eventually bottoming-out in July. What seems to be suggested here is that having attracted price-conscious shoppers into the stores prior to each festival, sales are withdrawn, just when many food purchases are likely to be made. Retailers are nothing, if not canny.

Figure 4: The Seasonal Pattern of Sales
(thresholds of 10%, 25% and 35%)



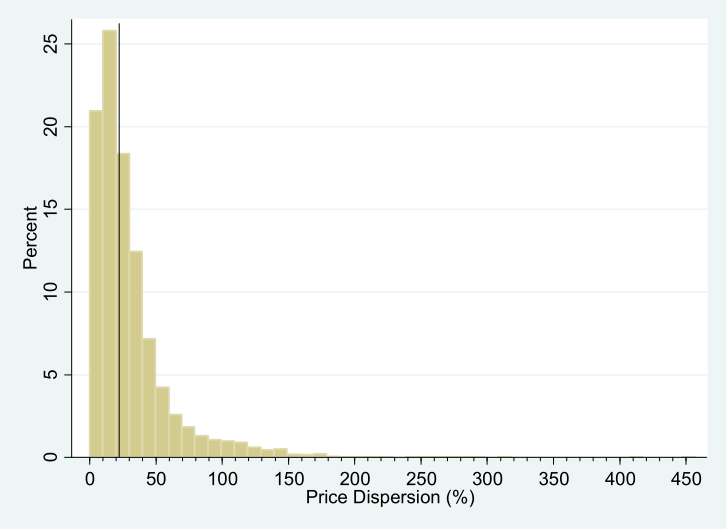
3.4 Price dispersion

One of the principal features of the data highlighted in Figure 1 is that prices for the same (*i.e.* identically barcoded) products differ markedly. While the most eye-catching examples in the dataset are due to sales - the largest range in the price of a product in a particular week being 457% - these are both infrequent and temporary occurrences. Of greater curiosity, given that the 'law of one price' is one of the basic tenets of textbook economics, is the observation that prices of identically-barcoded products are *persistently* different.

To gauge the typical range in price we observe consider Figure 5 which plots the distribution of price dispersion, defined here as the median weekly difference between highest and lowest prices for UPCs within the same product code over the sample period. The average of these, which represents a median of medians, turns out to be 22%, suggesting that the typical price range observed for barcoded products is close to one-quarter of a product's price. Such is the skew of the distribution, that for some products the typical range in price confronting the consumer is much greater than this, and almost 10% of products have a price

dispersion in excess 50%. Using references prices (which removes the influence of sales) has little effect, the typical dispersion still being 18%. Given that we are talking about identically barcoded products, this might suggest that at the barcode level at least, the law of one price appears to break down. Of course, while products are identical, retailers are less so, and it is the characteristics of the 'shopping experience' (customer service, in-store ambience, free parking and the like) that accounts for such persistent price differences. Nevertheless, what is clear is that persistently large differences in the prices of identical products is an enduring feature of food retailing, an empirical reality that sits somewhat awkwardly with the perception, promoted by the supermarkets themselves in their marketing campaigns, of widespread price-matching.

Figure 5: Distribution of Price Dispersion (median range in prices within each product code)



The Cheapest Retailer

The existence of persistent price differences begs the question whether prices are consistently lower (higher) in one particular supermarket, or more evenly spread. To investigate this issue the average prices of the products stocked in each retailer have been calculated in Table 4. As mentioned above, not every product is stocked in each retailer, so the table reports prices based on two classifications of products: (a) those products stocked in each retailer and (b) a subset of 92 products (comprising 78 branded 14 own-label) that are stocked in all seven retailers.

Retail chains that focus more on basic products of standard quality may be expected to have a lower average price than retailers that stock more premium (and thus more expensive) products. Hence, a price comparison based on the products stocked in each retailer (*i.e.* category (a)) sends a signal more about the product mix available than on pricing *per se*. In contrast, a comparison of the prices of commonly stocked products (*i.e.* those in category (b)) more accurately indicates which of the retailers is best on (the albeit narrow metric of) price.[2] Being stocked by all retailers these common products tend represent 'must-have' market leaders, major national brands and popular own-label products. Table 4 details the average prices along with the rank (1 denoting the cheapest supermarket) within these product groups. One retailer leads the rankings in both classifications, and that supermarket is ASDA.[3]

Table 4: Average Prices by Retailer for two groups of products

	All Products			
	(a) Stocked by retailer		(b) Common to all retailers	
	Price	Rank	Price	Rank
Tesco	128	4	128	2
Sainsbury	132	5	131	3
Asda	117	1	127	1
Safeway	142	7	140	6
Somerfield	127	3	140	6
Kwik Save	121	2	131	3
Waitrose	140	6	137	5
Average	**130**		**133**	

In terms of the average price for those products sold in each individual retailer, the influence of product mix is evident, in that the budget retailers (such as Kwiksave and Somerfield) are among the cheapest whereas more luxury retailers (Sainsbury and Waitrose) are among the most expensive. Results for the common set of products reflect size and buyer power more, with the mainstream retailers heading the rankings. The range in prices is however fairly modest, ASDA being 5% cheaper than the market average. Even the most expensive retailers (Safeway and Somerfield) are only 5% more expensive that the market average. Interestingly, it is the cheapest supermarket that uses sales the least, whereas the most expensive are among the most intensive users of discounts.

Concluding Comments

In this paper I have attempted to illustrate the richness of scanner data and indicate its potential for researchers, who like me, study food prices and the economic behaviour they represent. Few things we buy now are not read by a barcode scanner, so the technology provides a breadth and depth of information that was unthinkable a couple of decades ago. Indeed, scanner data is not only a new source of data but a new kind of data. As we have seen, prices at the barcode level do not behave like the traditional – market level - data which obliterate the distinction between retailers let alone regular and sale prices, both of which characterise barcode data.

While scanner data open-up many new avenues of academic pursuit, I also hope that they can inform supermarkets shoppers everywhere on a more practical level. So remember, to take advantage of sale prices you'll need to shop around - something of course the supermarkets know we are loathed to do in an era when 'Convenience is King'. Nevertheless, a reputation for sales doesn't mean your shopping basket will be any cheaper. Don't be fooled; price-matching is rare, despite what the slogans might try to portray. And whatever you do, don't leave Christmas shopping to the last minute. Finally, try to enjoy the shopping experience because while sales may draw us in, most products you place in the trolley are going to be cheaper elsewhere . . . if you only had the time to look.

Endnotes

1 Discounters such as Lidl, Netto and Aldi did not submit data to Nielsen at the time of the sample, but together accounted for less than 3% of market share. Marks and Spencer did not sell branded goods at that time and are excluded for this reason. Note that Safeway, Kwik Save and Somerfield ceased trading in 2005, 2007 and 2011 respectively.

2 As any marketeer will tell you, it is value rather than price that is key, but since we only have data on prices, we can leave the vexed question of which supermarket offers the best 'value for money' to others to debate.

3 While this may come as something of a surprise to some, ASDA has received the award of Britain's Cheapest Supermarket since 1997 in an annual competition undertaken by the trade magazine *The Grocer*.

References

Competition Commission, (2000). *Report on the Supply of Groceries from Multiple Stores in the United Kingdom*, London HMSO.

Eales, T. 'An Overview of the Market and Brand Performance in the Downturn' Paper Presented at the annual conference of the Institute of European and Comparative Law, St Catherine's College, University of Oxford, 2012.

Lloyd, T., McCorriston, S., Morgan C.W., Zgovu, E. (2011) 'Do Sales Matter: Evidence from UK Food Retailing' *School of Economics Working Paper 11/01* University of Nottingham.

Lloyd, T.A., S. McCorriston, C.W. Morgan (2013) 'Retailer Heterogeneity and Price Dynamics' *TRANSFOP Working Paper No.8*, University of Exeter.

Nakamura, E. and J. Steinsson (2013) "Price rigidity: microeconomic evidence and macroeconomic implications" *Working Paper 18705*, National Bureau of Economic Research,.

Office of National Statistics (2010) *Family Spending 2009 A report on the Living Costs and Food Survey*, Crown Copyright, London.