Do Sales Matter? An Exploration of Price Discounting in UK Food Retailing

T.A. Lloyd, C.W. Morgan, S. McCorriston and E. Zgovu

Abstract

This paper assesses the impact of promotional activity in the prices of food products on supermarket shelves. The study analyses a unique, high frequency panel of supermarket prices consisting of over 230,000 weekly price observations on around 500 products in 15 categories of food stocked by the UK's seven largest retail chains. In all, 1,700 weekly time series are available at the barcode-specific level including branded and own label products. Prices are inclusive of promotions and thus allow the frequency, magnitude and duration of sales to be analysed in greater detail than has hitherto been possible with UK data. Using this price data, sales periods are indentified. Results show that around 8% of products are on sale at any one time, and that sales are typically four weeks in duration. The average discount is 24% of the regular price. Importantly, sales are shown to have a relatively modest role in overall price variation – less than the dispersion in prices by retailer, pointing to changes in the regular price, via general inflation and idiosyncratic shocks, as the principal cause of price variation.

Keywords: Food Retailing, pricing, sales.

JEL Classification: L16; L66; Q13.

This paper was prepared for presentation at the International Association of Agricultural Economists Conference, Beijing, China, August 16-22 2009. Tim Lloyd and Wyn Morgan are Associate Professors and Evious Zgovu research assistant in the School of Economics, University of Nottingham, UK and Steve McCorriston is Professor in the School of Economics and Business, University of Exeter, UK. Tim Lloyd is grateful to Caroline Saunders and colleagues at the Agribusiness Economics Research Unit, University of Lincoln, New Zealand where much of the work was undertaken during sabbatical leave.

1. Introduction and Motivation

How firms and retailers adjust prices is an important issue from both a macroeconomic and industrial organisation perspective. On the macroeconomic side, the issue of 'sticky' prices is a characteristic of models related to understanding the inflationary process, the impact and timing of business cycles and the potential success of macroeconomic interventions. In this setting, the importance of 'menu costs' plays a role whereby firms find it costly to adjust prices on a frequent basis, particularly if there is some uncertainty about how sustained the source of the price change (e.g. the rise or fall in commodity prices) is likely to be. Various theoretical macroeconomic models have addressed the issue of sticky prices including, *inter alia*, Calvo (1983) and Caplin and Spulber (1987). Recent research has focussed on the empirical evidence of price behaviour with the tendency to use micro-data on prices to gauge the importance of aggregate price adjustment. The conventional wisdom (at least relating to the US evidence) that had emerged was that prices adjust approximately once a year though more recent work by Bils and Klenow (2004), based on broader commodity coverage showed that price changes for manufactured and retailed goods occurred around once every 4 months. More recently, Nakamura and Steinsson (2006) have highlighted the importance of accounting for sales and discounting behaviour by retailers and once discounting is allowed for, changes in the regular price are less frequent than reported by Bils and Klenow (*op. cit.*). Recently, the issue of price behaviour from a micro-econometric perspective has been explored in a European setting in the Inflation Persistence Network coordinated by the European Central Bank and is summarised in Dhyne *et al.* (2005). On the issue on the significance of sales in understanding the pricing behaviour of firms, it is notable that Dhyne *et al.* conclude that sales behaviour is *not* a significant factor in understanding the price behaviour in Europe and that sales are less important in Europe than they are in the US. We take the issue of discounting and regular pricing as the focus of this paper on understanding the pricing behaviour of UK food retailers.

Pricing behaviour of firms is also important from an industrial organisation perspective in understanding the dynamics of competition in highly concentrated markets. Again the role of discounting is an important phenomenon. For example, Varian (1980) focussed on the role of sales as firms pursue a mixed strategy, using high prices to sell to consumers who do not compare prices and low prices to those that do. Pesendorfer (2002) focuses on inter-temporal price discrimination where firms offer periodic discounts but where prices are at a 'regular' level most of the time. Note that the models of discounting by Varian and Pesendorfer offer different predictions about price dynamics over time. Regarding empirical evidence, Hosken and Reiffen (2004) have explored price dynamics for food products in the US and have argued that neither the Varian or Pesendorfer models offer a satisfactory explanation of price dynamics in US food retailing[1].

[1] Other papers that deal with the issue of sales include Lal and Villas Boas (1998), Warren and Barsky (1995), Villas-Boas (1995) and Sobel (1984) among others.

Against this background, this paper focuses on the importance and frequency of price discounting among the 7 main food retailers in the UK. In terms of empirical strategy, the paper is closest to that of Hosken and Reiffen (*op. cit.*) though we offer several departures from recent strands in this research. Most notably, we focus on products at the barcode level rather than product groups, as this will provide detailed evidence on the potential significance of discounts that may be hidden in a broader aggregate. As the data are recorded by retail chain they facilitate the analysis of identically bar-coded products across retailers[2]. In addition, the (weekly) frequency of our data is higher than monthly data used in much recent research and it extends over a long (three year) period. As such, the data here offers the potential for deep insights into the significance of sales across multi-product retailers, and is the first analysis conducted on data from the UK[3]. Acknowledging the existence and the potential significance of heterogeneity in pricing strategies across brands and retailers is an important aspect of this and future research on price dynamics (see also Fougére *et al.*, 2005). As such, our paper is broader in scope than Berck *et al.* (2008). They use scanner data on frozen orange juice and refrigerated orange juice (which allows them to explore the role of durability in the determination of sales) across a large number of stores in 24 US cities. Reflecting the characteristics of the UK food retailing sector, our data covers all the major retailers in the UK and for a larger number of product groups than has hitherto been explored.

The paper is organised as follows. Section 2 describes the data set – the largest of its kind in the UK - that forms the basis of our empirical work. Section 3 presents a discussion of the methodological issues in identifying sale periods and Section 4 a summary of their empirical characteristics. In Section 5, prices are summarised and an initial assessment of the importance of sales presented. Section 6 offers a summary of our, as yet preliminary, findings.

2. Price Data

In the empirical analysis we utilise a unique, extensive and high frequency panel of supermarket food prices derived from electronic point of sale (EPOS) data obtained from A.C. Nielsen (UK), a leading market research company to whom all major UK supermarket chains submit data relating to in-store transactions. Our data derives from the records of the seven largest of these supermarkets, which as a group represented around three-quarters of all food sales in the UK during the sample period. The supermarkets include all mainstream and one prestige grocery retailer. [4]

[2] Although not reported in this version of the paper, the data set outlined below also allows us to test differences in pricing strategies between national brands and own-brands.

[3] Identifying pricing behaviour for identical brands/products across retailers also offers the scope for exploring whether price are coordinated and whether such actions are against the public interest. Important in this regard is the recent investigation by the UK Office of Fair Trading which addressed the issue that UK retailers were coordinating the price of milk. See below.

[4] The seven supermarkets included in the sample are Tesco, Sainsbury, ASDA, Safeway, Somerfield, Kwik Save and Waitrose. The remaining 25% of sales are accounted for by small national and regional supermarket chains and independent retailers. Discounters such as Lidl, Netto and Aldi did not submit data to Neilsen at the time of the sample, but together accounted for less than 3% of market share. Marks and Spencer do not sell branded goods and are excluded for this reason.

The price information contained in the dataset is based on the details recorded by laser barcode scanners as products pass through supermarket check-outs. As a result, prices are based on 100% of transactions of the sampled products rather than derived from consumer surveys. Overall, the sample consists of 231,069 weekly price observations on around 500 products in 15 categories of food.[5] They relate to a (137 observation) sample frame running from 8th September 2001 to 17th April 2004. Some 90% of products are available throughout this period, the minimum number of observations for any product being 103 weeks. [6]

Each price observation in the sample represents the simple average of the prices posted in each of the retailers' stores on the Saturday of each week. Price observations are thus retailer-based national (Great Britain) averages. For a retailer such as Tesco, (the largest in the UK) prices are averaged over those posted in several hundred of its stores nation-wide. While store managers may have some flexibility over pricing, particularly for perishable items, the large number of products stocked in most stores (which typically exceeds 25,000) mitigates against widespread differences between stores. Also, major UK grocery retail chains claim to that national pricing strategies are the norm for the barcoded food products (Competition Commission 2000).

Discounts relating to store 'loyalty' cards are not included since they apply to the consumer's total spend rather than the prices of specific products. Promotion-inclusive prices are particularly attractive to the analysis of retail price dynamics owing to the fact that 'sales' are widely acknowledged to play a highly influential role in price setting in multiproduct firms in general, and food retailing in particular (see Hosken and Reiffen, 2004 for the US). However, as the previous section has highlighted, analyses of prices at the micro-economic level are rare. The current analysis represents the first in UK food retailing and as such affords unprecedented insights in to the prices faced by the modern food shopper.

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 dataset contains retailer-specific prices of such 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 prices series, the

[5] The 15 categories are orange juice, instant coffee, breakfast cereals, teabags, yoghurt, wrapped bread, tinned tuna, tinned tomatoes, tinned soup, corned beef, fish fingers, frozen peas, frozen chips, Jam and frozen pizza.

[6] Time series are contiguous (in that there are no missing observations once the time series has begun) in 100% of cases, although some (10%) start later than 8th September. All time series finish in the week ending 17th April 2004.

distribution of which is summarised in Table 1. [7] Data (% of dataset) are most prevalent in the bread (34%), soup (18%), coffee (8%) and orange juice (6%) categories, each of which contain in excess 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 dataset is by no means a representative sample of consumer spending on food (fresh fruit and vegetables are not part of the dataset since they do not carry unique barcode indicators) but the range of categories is relatively broad, spanning beverages and foods across a range of formats (fresh, chilled ambient and frozen).

As Table 1 also shows, seven categories contain products in both branded (sold with the manufacturer's name) and own label (sold under the retailer's name) forms. Own label products with the same product profile (e.g. an 800 gram standard medium sliced white loaf) are treated as one product and have the same product code in the database. Retailer-specific prices of these products (i.e. UPCs) represent the Tesco own label 800 gram standard medium sliced white loaf, or the Sainsbury own label 800 gram standard medium sliced white loaf, for example. Hence, own label versions of the same product are treated analogously to the branded products stocked by multiple retailers in the Nielsen dataset. In the UK, where sales of own label products account for a significant minority of the total consumer spend, this dimension of the dataset offers potentially insights in to any differences between the pricing of manufacturer- and retailer- branded products. Own label products account for nearly one-fifth of the products listed in the dataset.

[7] Each price observation is uniquely identified by its UPC and the (week ending) date but because the data set is an unbalanced panel (in that not all products are sold in all supermarket chains in all weeks) summary statistics vary slightly depending on the standardisation that is used. For example, orange juice accounts for 5.33% of the product codes, 6.34% of the UPCs (product code × retailers stocking the product) and 6.40% of the observations (product code × retailers × weeks). Unless specifically stated, UPC (i.e. the product code-retailer combination) will be taken to represent the principal unit of analysis when describing the dataset.

Table 1: Distribution of Unique Product Codes by Category

Category	Branded Products	Own Label	All Products	%
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 presents the prices of eight well-known branded products selected from the dataset on the basis that they are sold in most if not all of the retail chains. Hence, for each product there are seven UPCs representing the national average prices in each of the retailers at weekly intervals. Without wishing to generalise they display a number of interesting features, in particular is the way that sales punctuate the time series, albeit with as frequency and intensity that varies by products and retailer. When not on sale prices tend to coalesce around particular levels, although this regular (i.e. non-sale) price is constant over the sample, particularly so for the products in tinned tomatoes and wrapped bread categories. It is also apparent that despite representing the prices of identically bar-coded products, there are persistent, and substantial differences in the prices charged by retail chains.

Coverage of the dataset by retailer is summarised in Table 2. All seven supermarkets are well-represented in the sample, Tesco having the largest number of observations at 17%, Waitrose the fewest at 11%. One of the most interesting aspects of the dataset is that prices are available by retail chain, allowing time series analysis of identically bar-coded products across retailers. Issues relating to the synchronisation and staggering of prices (Lach and Tsiddon 1996; Berck *et al. op. cit.)* are most relevant here.

Figure 1: Weekly Prices (pence) of a Selection of Products sold by UK Retail Chains

Table 2: Distribution of Unique Product Codes by Retailer

Chain	UPCs	% of total
Tesco	292	17.14
Sainsbury	275	16.14
Asda	228	13.38
Safeway	263	15.43
Somerfield	242	14.20
Kwik save	221	12.97
Waitrose	183	10.74
Total	**1,704**	**100.00**

Table 3 shows how the 507 products (i.e. separate product codes) are distributed across retailers, broken down by label. [8] Overall, 63% (320/507) of the products were sold in at least 2 retailers, with 17% sold in all seven. Many of the products are sold in just one retailer; indeed, this is the most common outcome over both the branded and own label groups. In terms of (manufacturer) branded products 71% (267/375) were sold in at least 2 retailers with 21% sold in all seven. For own (retailer) label products comparable statistics are 43% and 11%.

Table 3: The Distribution of Products (by label) Stocked by the Supermarkets

	Number of Supermarkets							
	1	2	3	4	5	6	7	**Total**
Branded	**108**	**45**	**36**	**37**	**28**	**43**	**78**	**375**
	28.80	12.00	9.60	9.87	7.47	11.47	20.80	100.00
	59.02	67.16	92.31	75.51	84.85	97.73	84.78	73.96
Own Label	**75**	**22**	**3**	**12**	**5**	**1**	**14**	**132**
	56.82	16.67	2.27	9.09	3.79	0.76	10.61	100.00
	40.98	32.84	7.69	24.49	15.15	2.27	15.22	26.04
Total	**187**	**69**	**38**	**51**	**32**	**42**	**88**	**507**
	36.88	13.61	7.50	10.06	6.31	8.28	17.36	100.00
	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Note: : Cell entries in bold represent the number of series in each classification, with this number expressed as a percentage of the row (label) and column (sold in) totals respectively in the cell entries beneath.

[8] Note that the entries in the table refer to products (e.g. 100g jar of Nescafe 'Gold Blend' instant coffee) rather than UPC (100g jar of Nescafe 'Gold Blend' instant coffee in Tesco) and thus proportions in the table need not correspond to those in Table 1. See previous footnote for clarification.

3. Creating Sales Data

In order to investigate the impact of promotional discounting (what we refer to as sales) on these price data an indicator variable is created to identify sale periods. Prices are generally not declared as 'sale' or 'regular' by manufacturers or retailers in the datasets constructed by government departments or commercial organisations, with the result that the identification of sale periods is based upon the prices themselves. Broadly speaking, sales are simply periods of temporary price reduction, although the precise definition employed in empirical work varies. For example, prices supplied by the Bureau of Labour Statistics (BLS) for the construction of the U.S. Consumer Price Index is identified as 'on sale' if the product is sold below its 'regular selling price', as assessed by the Bureau's field agents in the monthly survey. Nakamura and Steinsson (2005) use this sales flag to identify sale periods. Hosken and Reiffen (2004) also use BLS data but define a sale with reference to the actual behaviour of prices, rather than the field agents' assessment. Specifically, if a price falls by more than some fixed percentage (they consider 10% and 20%) between two adjacent months (m-1 and m) but is then reversed in the following month (i.e. between m and m+1), the product is treated as being on sale in month m. In other words, a product is recorded as being 'on sale' in month m if the prices in months m-1 and m+1 are markedly (i.e. 10% or 20%) higher. Unlike Nakamura and Steinsson's (2005) measure, this definition implicitly restricts sales to last one month, although this is probably not an unreasonable assumption in the majority of cases. Nevertheless, because price setting typically occurs on a weekly basis, some of the finer detail of price adjustment is inevitably obscured by the use of monthly data.

In those studies analysing weekly price data, definitions of sale prices are modified to reflect the data's higher frequency. For example, Campbell and Eden (2005) define a sale to have occurred if a price decline of 10% or more between weeks w-1 and w is then completely reversed within two weeks (i.e. in w +1 or w+2). Using this definition, a sale last for a minimum of one and a maximum of three weeks. Berck *et al.* (2008) consider price falls of 25% 35% and 50%, but use the store's modal price ($\tilde{p}$) over a two year period as the basis for comparison rather than the price in the week preceding the decline. Using this definition, sales are recorded for all weeks w in which the price falls below $\tilde{p}$. While this approach leaves the length of sales unconstrained, it relies upon the mode being representative of the non-sale (or regular) price. Where the regular prices changes over time (reflecting general inflation or specific changes in production costs, for example) use of the mode to identify sales is arguably less than ideal (see below).[9]

[9] It is noteworthy that in their study of refrigerated and frozen orange juice, Berck *et al.* (2008) find evidence of a clear mode in fewer than half the refrigerated products and in two-thirds of frozen products they analyse.

The foregoing discussion serves to highlight some of the practical difficulties in identifying sales - 'periods of temporary low prices' -, from price data alone, namely the duration of 'temporary', the magnitude of 'low' and the reference price used in the assessment of each. Mindful of these considerations we define a sale as a period during which price falls by at least *x*% of the observation immediately preceding the decline in prices, and then which is reversed within 12 weeks. In this definition notice that:

(i) sales of long duration (i.e. less than 3 months) are allowed for. While sales of 2-4 weeks are typically believed to be the norm and sales of longer than 6 weeks rare in UK food retailing (Competition Commission, 2000 p.116) this measure does allow for the (albeit infrequently observed) longer sale durations;

(ii) it is the cumulative price drop (i.e. the peak-to-trough difference) that it used rather than any week-on-week change in price that is used to define the magnitude of price change. This allows for price changes that are staggered over more than a single week at the start of or end of a sale period. This may be important in our data if a national price promotion is implemented over adjacent weeks;

(iii) it is actual prices that act as the reference price. This is likely to be useful in cases where there is no single non-sale price (e.g. 99p) that applies over the entire sample. Hence, the *regular price*, refers to a state of nature (paralleling the status of the term *sale price*) rather than a fixed value such as the mode;

(iv) in recognition of the fact that results inevitably depend on the price decline that is chosen, we consider three thresholds, namely x = 10%, 25%, and 35%;

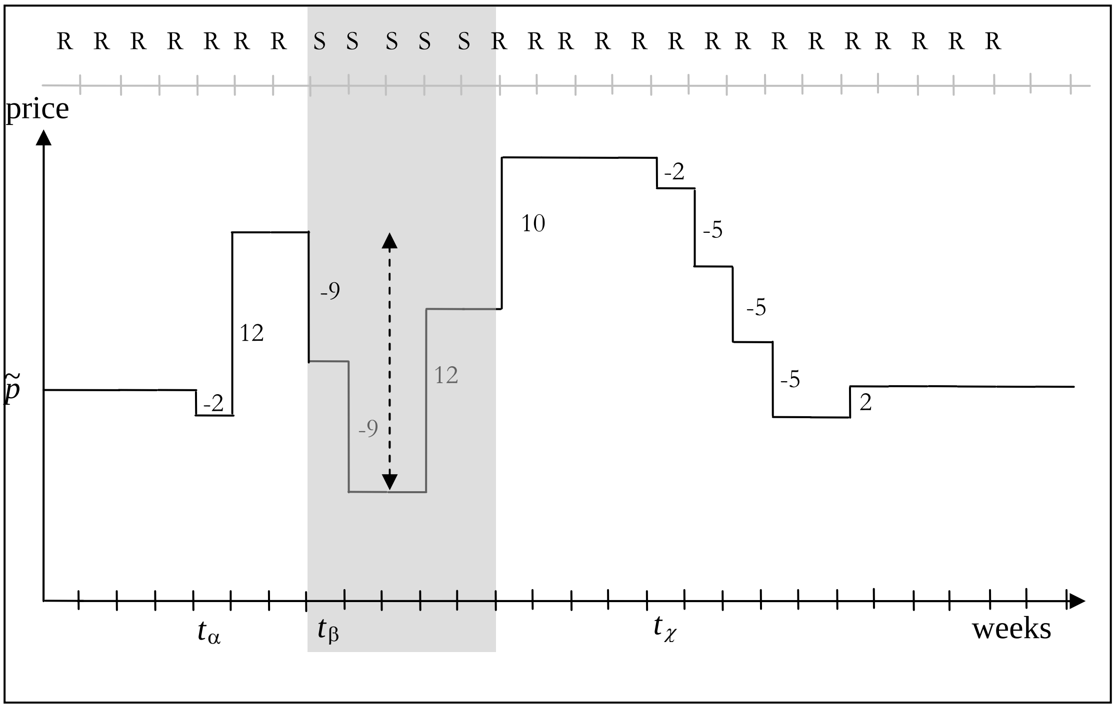
(v) the sale period ends when prices return to a level that is *at least* as high as the price preceding the sale. With this condition, prices do not need to return to their pre-sale level. This is useful since it allows us to investigate whether changes in the regular price are more likely to occur following sale periods that at other times;

(vi) all prices between the initial decline and the subsequent reversal are counted as sale prices. Where the start (end) of a sale occurs in adjacent weeks in different regions, the fall (rise) in the national average price will be staggered and tend to overstate the duration of the sale slightly.

In order to illustrate the effect of these conditions in the creation of a sale indicator consider the stylised weekly time series of prices depicted in Figure 2. While the figure is by no means representative of the price series in the dataset, it exhibits some of the more problematic features that characterise some series, most notably changes in the regular price and staggered sale prices. At the top of the figure are labels (R and S denoting regular and sales prices respectively) generated by application of our sales algorithm with a 10% threshold to the stylised data. Numbers adjacent to the price levels denote the (non-zero) week-on-week percentage price changes.

In the figure there are three episodes of lower prices, commencing at weeks t_α, t_β and t_χ, although only one (at time t_β and shaded grey for clarity) is recorded as a 'sale' by our definition. This owes to the fact this is the only one of the three periods which satisfies the conditions that (a) the peak-to-trough price drop exceeds 10% and (b) the price returns to (at least) the pre-sale level within 12 weeks. Note that neither of the (9%) falls that make up the 18% cumulative decline would trigger a sale using a 10% week-on-week criterion. The two other periods of price decline shown (at t_α and t_χ) do not qualify as sales by our definition because in the first case the price decline (of 2%) is too small; and in the second case, the decline in price is not reversed within 12 weeks. Finally, note that $\tilde{p}$ represents the mode of the series, underlining the limitations of its use in sales identification when there are changes in the regular price.

Figure 2: Sales Identification in a Stylised Weekly Time Series of Prices



While the flexibility of the sales definition we have adopted in this paper seeks to overcome some of the key issues in data of this sort, it should not be overlooked that any set of criteria designed to distinguish sales from changes in the regular price using prices alone are to some extent arbitrary. As a guiding principle, it seems reasonable that the larger the transitory price decline is, the more likely it is that the observations represent a sale. For this reason we in initially consider (peak-to-trough) price drops of 10%, 25% and 35% in sales identification, a description of which is presented in the following section.

4. Sales Data

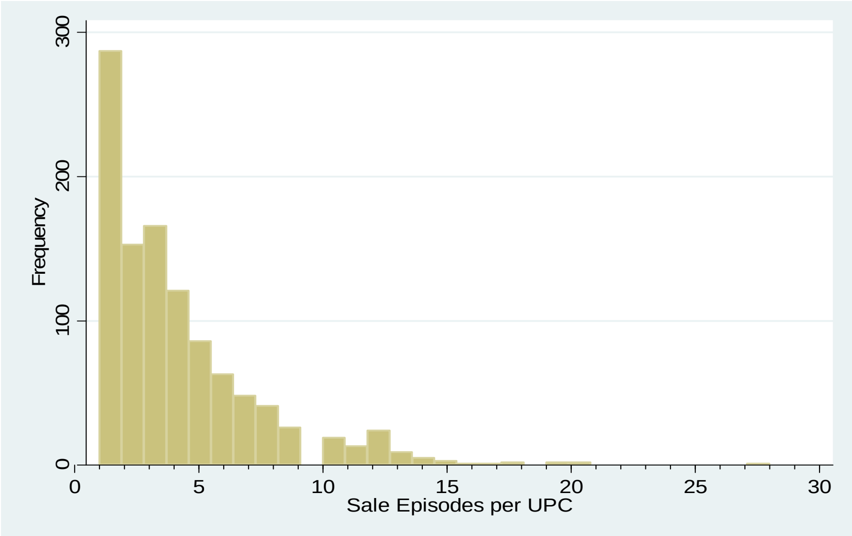
Table 4 reports summary statistics of the sales defined according to 10, 25 and 35 percent thresholds. It shows that nearly 8% of prices are classed as 'on sale' using the 10% threshold, a figure that drops to 3.5% and 1.4% using the larger discounts. Thus while sales are clearly the exception to the normal rule of pricing, only very deep sales are rare. Table 4 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 time series 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 (using the 10% measure). The figures in Table 4 also suggest that sales tend to be around four weeks long, irrespective of their depth.

Table 4: Summary Statistics of the Sales Data

	Sale Threshold		
	10%	**25%**	**35%**
Frequency (%)	7.79	3.45	1.41
Products on sale (%)	62.97	36.80	20.07
Average Duration (weeks)	4.52	4.39	4.17
Sale episodes	4,309	2024	822
Sales per UPC (all products)	2.50	1.19	0.48
Sales per UPC (sale products)	4.01	3.22	2.40

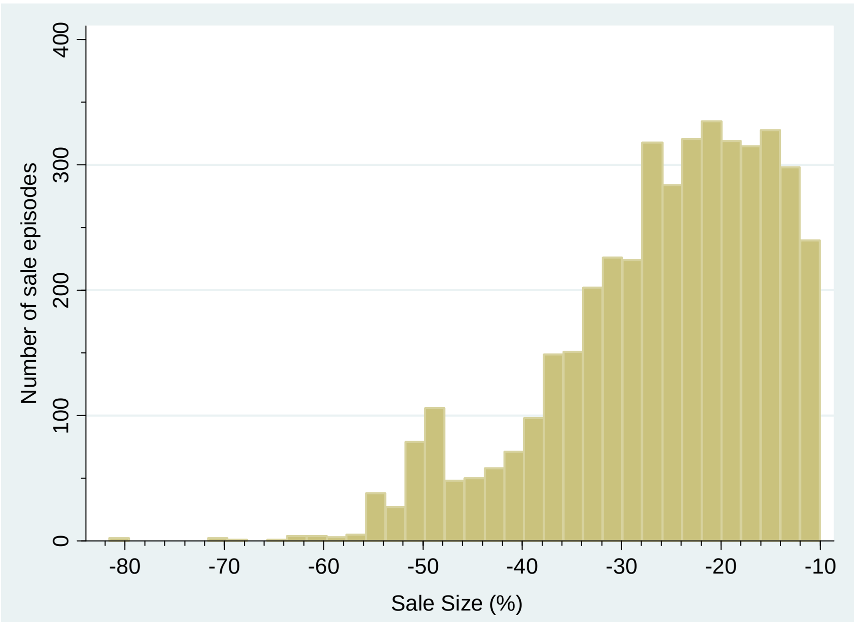
In terms of the number of sale episodes (rather than the number of observations) the dataset contains 4,309 10% sales, of which around half (2,024) are sales of >25% and one-fifth being sales of >35%. These figures imply that on average a UPC will experience a 10% sale around 2.5 times during the sample (a little under once per year), a figure that rises to four (just over 1.5 per year) if we consider only those UPC that have ever been discounted. As is clear from the distribution of sale episodes per UPC displayed in Figure 2, most products that have been on sale, have only ever been discounted once during the sample frame, although there are a small number of frequently discounted products.

Figure 2: The Frequency of Sales



For a more complete picture the distribution of sales by size consider Figure 3. While sales in excess of 80% have occurred in a handful of cases, discounts in excess of 50% are rare, accounting for less than 5% of all sales. The majority of sales represents discounts of between 10 and 30%, the median discount being 24%.

Figure 3: Size Distribution of Sales



The statistical description of sales presented above suggests a number of stylised facts, that may be conveniently summarised as:

- *Sales are unusual*, occurring less than 10% of the time;
- *Most UPCs experience a sale*, although around one-third do not;

- *A small number of products are discounted frequently*
- *Sales last around four weeks*;
- *Sales typically represent discount of 25%.*

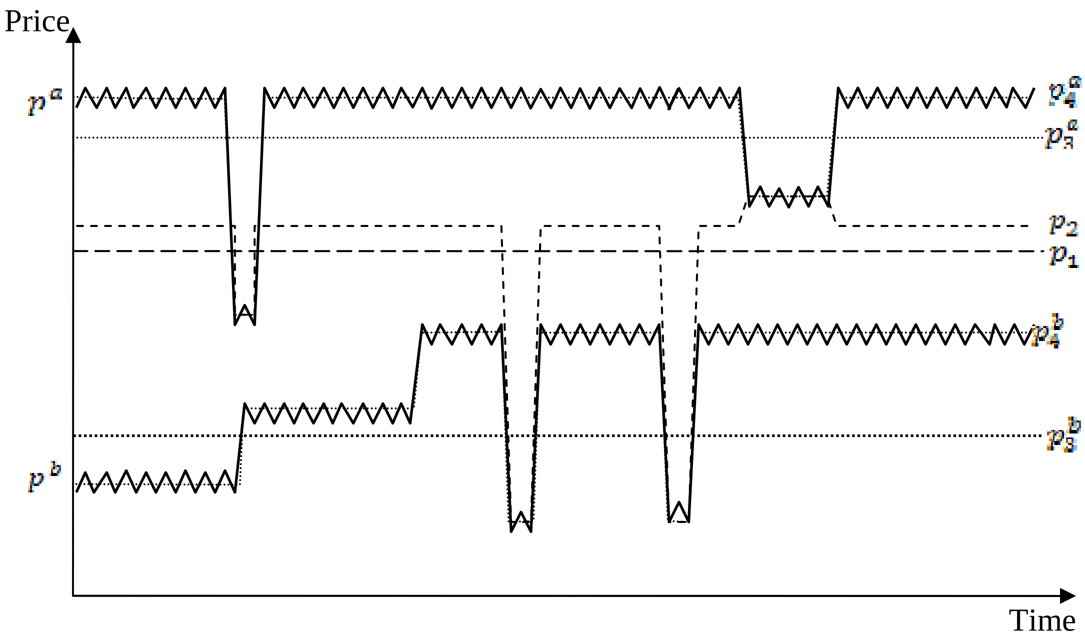
5. The Importance of Sales in Price Variation

Method

The previous section has highlighted the important role played by promotional sales in product pricing, and so now we turn to quantify the contribution of sales in the overall variation in prices. To do this we estimate price regressions containing increasingly rich sets of dummy variables. As with any such analysis, there are two components to the issue, the deviation from the reference point through time, and the choice of the reference point itself. While the average price of each barcoded product is the most natural reference point, it is apparent from the data that the prices of like-products, even those with the same barcode, are seldom the same across retailers. Indeed, with the average dispersion being some 25%, price parity is very much the exception rather than the rule. As a result we estimate four regressions: in Model 1, we fit a product-based set of dummy variables to the price data (i.e. one dummy for each product code). The coefficient attached to each dummy represents the average weekly price of a barcoded product (such as a 200g jar of Nescafe standard blend instant coffee) over the sample across all retailers. Where differences in the price of products are relatively small, this barcode average represents a sensible baseline against which the importance of sales can be assessed. This is done in Model 2 which augments the product-based dummies with a set of sales dummies, in which each and every sale episode is identified by a separate dummy variable. Comparison of the fit of Models 1 and 2 indicates the importance of sales relative to the mean price of the product code across all retailers. In Model 3, the product-based set of dummies is replaced by a set of UPC-based dummies, so that the coefficient on each dummy represents the average weekly price of each bar-coded product in each supermarket. In Model 4, these UPC-based dummies are augmented by the same set of sales dummies as used in Model 2.

To illustrate the approach consider Figure 6 which depicts the prices of an identically bar-coded product in just two supermarkets (given by the sold lines p^a and p^b). For the purposes of illustration it has been assumed that the product sells at different prices in each supermarket, as might be expected if supermarket *a* were a luxury retailer and supermarket *b* a discounter. While both supermarkets periodically promote the product, sales are not necessarily of the same duration and intensity. Also note that while supermarket *a* maintains a constant regular price, supermarket *b* changes the regular price during the sample. Using this hypothetical data, estimation of Model 1 would deliver a mean price for the product, of say p_1, and the extent to which p^a and p^b deviate from it would indicate the typical variation in the price of the product. Augmenting the regression with sales dummies produces fitted values given by p_2, against which the variation of actual prices (p^a and p^b) may be evaluated. Comparison of results from Model 1 and 2 offers an indication of the contribution of sales to overall variation in price. As the figure illustrates however, this assessment is confounded by the large difference that exists between p^a and p^b, and may even give rise to positive coefficients on the sales dummies, as shown in the figure. Allowing for retailer-specific mean prices involves estimating a price regression using a set of UPC-based dummies, which we have labelled Model 3 and gives rise to the UPC mean prices p_3^a and p_3^b. Assessing price variation against these reference prices is likely to be appropriate when large differences in prices for like-products exist (as seems to be the case in our data). In fact, Model 3 performs a dual purpose: not only does it indicate the contribution of 'retailer' effects in overall price variation when results are compared to those from Model 1 but it also provides a more representative baseline when assessing the importance of sales. This is done in Model 4 which augments the UPC-based dummies with the same set of sales dummies used in Model 2 producing the fitted values p_4^a and p_4^b. As the also figure indicates, the importance of sales in price variation diminishes the more changeable the regular price actually is.

Figure 6: Modelling the Contribution of Retailer and Sales in Price Variation



Any of the usual criteria (such as $\bar{R}^2$ or AIC) may be used to compare the fit of each of the models however in the empirical analysis we express prices in natural logarithms so that the regression standard errors (rse) represent the average size of the residuals as a proportion of product price, and it is this that we use to evaluate the effect of retailer and sales on the variation in prices.

Results

The empirical analysis is conducted on a sample containing 507 products and 1,704 UPCs. Using the 10% (peak-to-trough') criterion to identify a sale we also have 4,309 sales dummies, each of which represents an individual 'sales episode' (of 1 to 12 weeks in duration). The parameters of Models 1 to 4 are estimated by ordinary least squares using 231,069 price observations and regression standard errors (rse) reported in Table 6. Results for the 'all products' regressions head the table with results by category reported underneath. Moving across the table the entries represent the regression standard errors from Models 1, 2, 3 and 4. To facilitate comparison between models, figures in parentheses represent the percentage reduction in standard error. The percentages in the columns headed Model 2 and Model 4 represent the contribution of sales to the overall variation in prices relative to the product (Model 1) and UPC (Model 3) sample averages respectively. The percentage under the heading Model 3 represents the reduction in standard error relative to Model 1, and as such assesses the importance of retailer in overall price variation rather than sales. To gauge the importance of changes in the regular price, the final column reports the percentage of the regression standard error of Model 1 unaccounted for by either sales or retailer (i.e. Model 4).

Focussing on the 'all categories' results at the head of the table it is apparent that when using barcode-specific means (Model 1) as the reference point, the regression standard error of 0.132, implies that the typical variation of a product's price around its sample mean is about 13% across all 507 products. To assess the importance of sales in this variation, we estimate Model 2 which has an rse of 0.114, which represents a fall of some 13%. In other words, at the barcode level, sales account for around 13% of the variation in prices. On this measure, sales are most important for products in yoghurt (29%), frozen pizza (26%) and teabags (24%); and least important for products in bread (6%), jam (6%) and fish fingers (2%).

Given the price disparity of like-products by retailer it may be helpful to consider the variation around UPC - rather than product code - averages. This is done in Model 3, which fits the data to a set of 1,704 UPC-based dummies. Model 3 establishes a new baseline – the price of bar-coded products by supermarket – against which the importance of sales can be assessed. This yields Model 4 which has an rse of 0.080 implying that over all the products in the sample, sales account for 23% of price variation at the UPC level. The breakdown by category produces broadly similar results to those reported above, although the proportion of price variation attributable to sales is higher since the differences between retailer and product averages have been removed from the overall price variation.

To put these figures in to context we may compare the effect of sales on price variation with that attributable to retailer as a yardstick. This merely involves a comparison of Models 2 and 3 relative to Model 1. Focussing on the All Categories figures it is apparent that at 20%, retailer-type effects account for a higher proportion of price variation than sales (at 13%) at the barcode-specific level.

Table 6: The Importance of Retailer and Sales on the Variation in Food Product Prices (Regression Standard Errors of Models 1 to 4).

Category	Model 1	Model 2[a]	Model 3[a]	Model 4[b]	Unexplained variation[d]
	Product dummies	Product & Sales dummies	UPC dummies	UPC & Sales dummies	
All Categories	**0.131504**	**0.11407349 (13%)**	**0.105043 (20%)**	**0.080498 (23%)**	**61%**
Orange Juice	0.159544	0.148196 (7%)	0.096230 (40%)	0.078449 (18%)	49%
Instant Coffee	0.134481	0.108324 (19%)	0.109188 (19%)	0.078246 (28%)	58%
Tinned Tuna	0.160784	0.1270385 (21%)	0.144194 (10%)	0.098433 (32%)	61%
Tinned Tomatoes	0.244132	0.2305431 (6%)	0.225925 (7%)	0.209942 (7%)	86%
Tinned Soup	0.127419	0.0963458 (24%)	0.116859 (8%)	0.077048 (34%)	60%
Frozen Chips	0.129171	0.1109475 (14%)	0.092703 (28%)	0.060721 (34%)	47%
Corned Beef	0.168753	0.1553559 (8%)	0.101734 (40%)	0.0715 (30%)	42%
Frozen Peas	0.084937	0.0767492 (10%)	0.059936 (29%)	0.040443 (33%)	48%
Fish Fingers	0.079044	0.0775685 (2%)	0.051563 (35%)	0.047841 (7%)	61%
Breakfast Cereal	0.101026	0.0778872 (23%)	0.092611 (8%)	0.068105 (26%)	67%
Teabags	0.141824	0.1084825 (24%)	0.118821 (16%)	0.075658 (36%)	53%
Yoghurt	0.114293	0.081271 (29%)	0.104655 (8%)	0.065109 (38%)	57%
Wrapped Bread	0.113602	0.1073524 (6%)	0.079992 (30%)	0.070232 (12%)	62%
Jam	0.119301	0.1122818 (6%)	0.080611 (32%)	0.066366 (18%)	56%
Frozen Pizza	0.201402	0.1486645 (26%)	0.181079 (10%)	0.107721 (41%)	53%

[a] Percentages represents the reduction in regression standard error compared to Model 1.

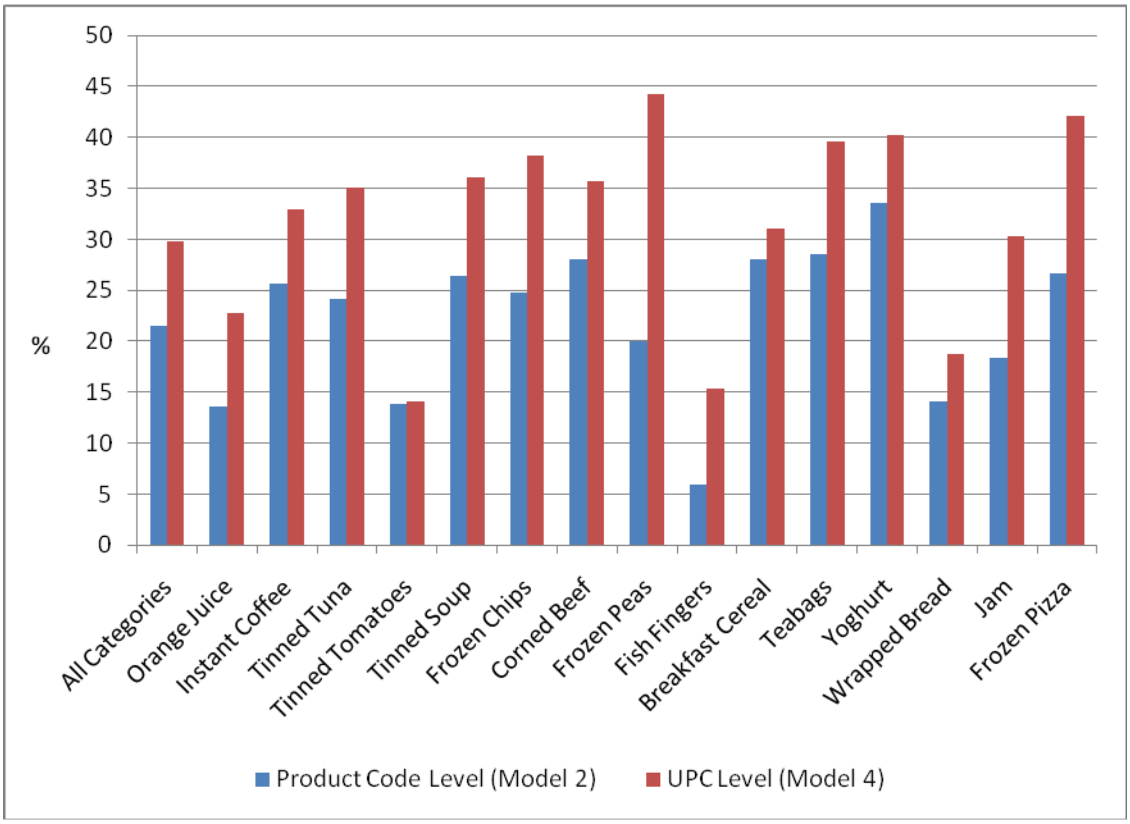
[b] Percentages represents the reduction in regression standard error compared to Model 3.

[c] Percentages represents that part of the regression standard error in Model 1 unaccounted for Model 4.

So, while sales are important in price variation they are generally less influential than the effect of retailer. Moreover, both sales and retailer effects are dominated by the influence of factors not accounted for in the models, principally, changes in the regular price. Comparing regression standard errors of Models 1 and 4 it is apparent that over all categories 61% of the variation in prices in Model 1 is not accounted for by either sales or retailer, and for tinned tomatoes this rises to 86% owing to the high rate of inflation exhibited by this category over the sample period.

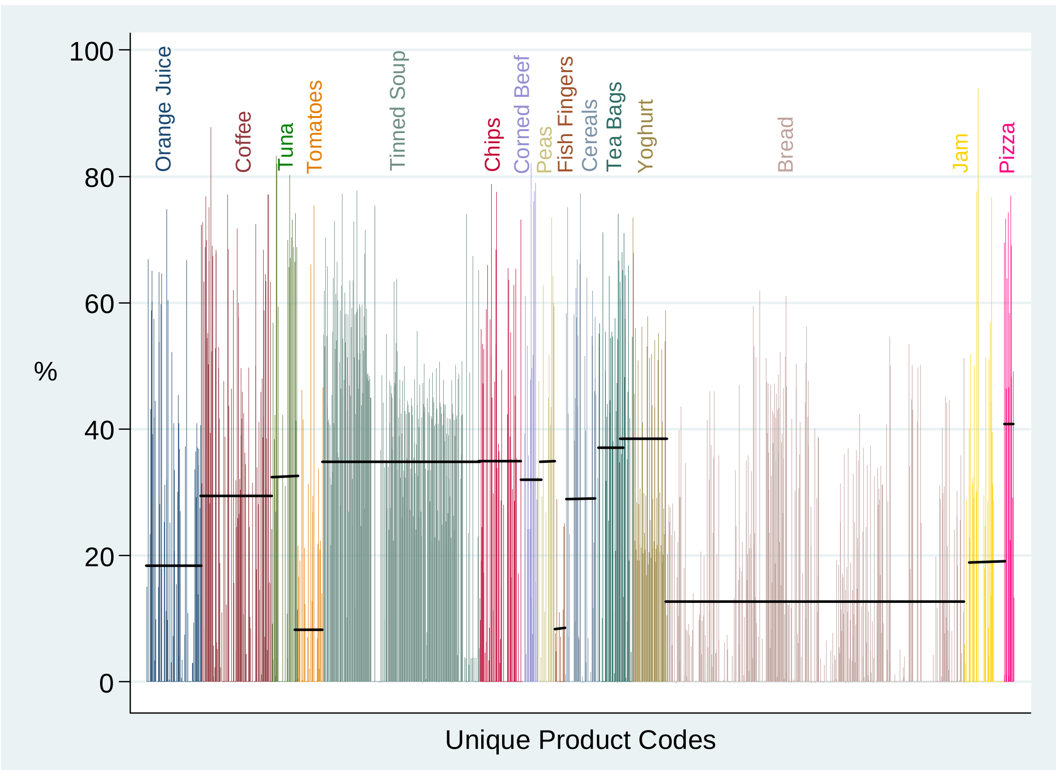
While the above results give an appropriate assessment of the importance of sales in the variation of food prices, it may be of interest to repeat the analysis on the subset of products that have actually experienced a sale, particularly so since around one-third of UPCs did not. Figure 7 reports the percentage contribution of sales at the product and UPC levels, paralleling the percentages under Models 2 and 4 in Table 6. Referring to the 'All Categories' figures, the results show that when those UPCs which had never been 'on sale' are excluded, the contribution of sales to price variation rises from 13% to 21% at the barcode (i.e. product) level and 23% to 30% when evaluated at the UPC level. Thus even when assessed against this somewhat contrived yardstick, sales have a relatively modest influence on the variation of food prices.

Figure 7: The Importance of Sales in Food Price Variation at Product and UPC Levels



As a final step to the analysis we drill down through these category averages to consider the contribution of sales in price variation for each and every UPC. Results are reported in Figure 8, where for the purposes of illustration we focus on the percentage of price variation accounted for by sales relative to UPC – rather than product – means. The height of each spike in the figure is the percentage reported under Model 4 in Table 6 for each UPC, with horizontal bars denoting category averages. Gaps represent UPCs that have not experienced a sale, so the figure illustrates not only the intensity of sales but the categories in which sales are most common. Recalling from Table 6 that the 'All Categories' average is 23% (30% if non-promoted products are excluded) it is unsurprising that there are only a handful of UPCs for which sales account for more than 80% of their price variation. The highest figure - of 94% - corresponds to a UPC in the Jam category, although sales accounts for a relatively low proportion of price variation in this category as a whole. As is also evident from the graph, sales are both relatively deep and widely used in many categories such as tinned soup, frozen chips, teabags and frozen pizza in contrast to wrapped bread, tinned tomatoes and fish fingers where they are typically lower and more sparse. While we do not attempt to account for these differences here, further analysis along these lines is likely to be fruitful, in particular with respect to retailer pricing strategies ('EDLP' vs 'Hi-Lo'), product branding (national brands vs own labels) and the number of retail chains stocking the products.

Figure 8: Sales in Food Price Variation by UPC



6. Conclusions

This paper presents a preliminary analysis of promotional sales in an extensive high frequency micro-economic database of supermarket prices in the UK. Given that the timing of sales is not recorded it must be inferred from the prices themselves and the method that we have adopted for the identification of sales is described. Using a 10% peak-to-trough measure we find that 92% of prices are regular, the remainder represent sales. So although sales are unusual, the majority of products experience a sale, the norm being around one sale per year. While there are marked differences in the number of times products are on sale (one third of the 1,700 UPCs considered were never on sale in the sample period) promotional activity generally lasts 4 weeks and represent discounts that average 25%.

The paper also offers an assessment of the importance of sales in overall price variability. The underlying message conveyed by our simple measures is that the influence of sales on price variation is at best, rather modest, accounting for 13% of price variation at the barcode level. Even when price variation is evaluated relative each retailer's mean for each barcode (i.e. the UPC), sales are shown to account for less than a quarter of price variation. In fact, the impact of sales on price variation is considerably less than the influence of the retail chain from which the product is purchased. Phrased slightly differently, where you shop is more important than when you shop. Indeed, one surprising feature in the data is the extent to which the prices of identically barcoded products differ across the retail chains. Even if sales are excluded, the average difference in prices of like-products is about 25%, suggesting that despite having identical barcodes products are far from homogenous.

The results also contribute to the debate regarding price stickiness. According to the basket of products analysed here, prices tend to vary around 13% of their mean level by barcode and 10% of their UPC mean. While this variability is not excessive, there is little evidence characterising the prices of food products as sticky by our measures. Moreover, however the results clearly show that lionshare of price variation emanates from movements in the regular price, brought about by general inflation and idiosyncratic shocks, not promotional activity. Perhaps one of the most interesting corollaries of this research is what it has to say about the 'success' of modern marketing strategy in exaggerating perceptions regarding the true extent of price promotion.

References

Berck, P. J. Brown J.M. Perloff and S.B. Villas-Boas (2008) 'Sales: Tests of Theories on Causality and Timing' *International Journal of Industrial Organisation*, 26(6): 1257-1273.

Bils, M. and P.J. Klenow (2004) 'Some evidence on the importance of sticky prices' *Journal of Political Economy*, 112: 947-985.

Calvo, G.A. (1983) 'Staggered prices in a utility-maximising framework' *Journal of Monetary Economics*, 12: 383-398.

Caplin, A. and D. Spulber (1987) Menu costs and the neutrality of money' *Quarterly Journal of Economics*, 102: 703-725.

Clarke, R., Davies, S., Dobson, P. and Waterson, M. (2002), *Buyer Power and Competition in European Food Retailing*, Edward Elgar, Cheltenham.

Dhyne, E., L.J. Álvarez, H.L.Bihan, G. Veronese, D. Dias, J. Hoffmann, N. Jonker, P. Lunnemann, F. Rumler, and J. Vilmunen (2005) 'Price setting in the Euro Area and the United States: some facts from individual consumer price data' *Journal of Economic Perspectives*, 20: 171-192.

Fougére, D., H.L. Bihan and P. Sevestre (2005) 'Heterogeneity in consumer price stickiness: a micro-econometric investigation' Working paper No. 536, European Central Bank.

Hosken, D. and D. Reiffen (2004) 'Patterns of retail price variation' *RAND Journal of Economics*, 35: 128-146.

Lach, S. and D. Tsiddon (1996) 'Staggering and Synchronisation in Price Setting: Evidence from Multiproduct Firms' *American Economics Review*, 86(5): 1175-1196.

Lal, R. and J.M. Villas Boas(1998) 'Price promotions and trade deals with multiple retailers' *Marketing Science*, 44: 935-949.

McCorriston, S. (2002) 'Why should industrial organisation matter to agricultural economists?' *European Review of Agricultural Economics*, 29: 349-371.

Nakamura, E. and J. Steinsson (2006) 'Five facts about prices: a re-evaluation of menu cost models' Mimeo, Harvard University.

Pesendorfer, M. (2002) 'Retail sales: a study of pricing behaviour of supermarkets' *Journal of Business*, 75: 33-66.

Sobel, J. (1984) 'The timing of sales' *Review of Economic Studies*, 51: 353-368

Varian, H. (1984) 'A model of sales' *American Economic Review*, 70: 651-659

Villas Boas, J.M. (1995) 'Models of competitive price promotions: some empirical evidence from the coffee and saltine crackers markets' *Journal of Economics and Management Strategy*, 4: 85-107.

Warren, E. and R. Barsky (1995) 'The timing and magnitude of retail store markdowns: evidence from weekends and holidays' *Quarterly Journal of Economics*, 110: 321-352.