Øyvind Thomassen, Howard Smith, Stephan Seiler and Pasquale Schiraldi

Multi-category competition and market power: a model of supermarket pricing

Article (Accepted version)
(Refereed)

Original citation: Thomassen, Øyvind, Smith, Howard, Seiler, Stephan and Schiraldi, Pasquale (2017) *Multi-category competition and market power: a model of supermarket pricing.* American Economic Review . ISSN 0002-8282

Multi-Category Competition and Market Power: A Model of Supermarket Pricing*

Øyvind Thomassen† Howard Smith††
Stephan Seiler‡ Pasquale Schiraldi‡‡

March 2017

Abstract

In many competitive settings consumers buy multiple product categories, and some prefer to use a single firm, generating complementary cross-category price effects. To study pricing in supermarkets, an organizational form where these effects are internalized, we develop a multi-category multi-seller demand model and estimate it using UK consumer data. This class of model is used widely in theoretical analysis of retail pricing. We quantify cross-category pricing effects and find that internalizing them substantially reduces market power. We find that consumers inclined to one-stop (rather than multi-stop) shopping have a greater pro-competitive impact because they generate relatively large cross-category effects.

JEL Numbers: L11: L13: L81

1 Introduction

In many competitive settings consumers buy multiple categories and find it convenient to obtain them all from a single store, location, or firm. This shopping behavior can generate complementary cross-category pricing effects, as an increase in the price of one category may lead a consumer to transfer away all his category purchases. The magnitude of cross-category pricing effects depends on shopping behavior: a consumer that prefers to purchase all categories at a single store may

*Authors are in reverse alphabetical order. All authors contributed equally. We thank the editor and two anonymous referees for very helpful comments, which have greatly improved the paper. We are also grateful for comments from Simon Anderson, Mark Armstrong, Martin Browning, Peter Davis, Jean-Pierre Dubé, Thierry Magnac, Ariel Pakes, Kate Smith, John Thanassoulis, Christopher Wilson, and seminar participants at Royal Economic Society Annual Conference, Toulouse, Paris, LSE, University of Zurich, the 8th workshop on the Economics of Advertising and Marketing (Nuffield College, Oxford), and the NBER Summer Institute (2016). Smith acknowledges financial support from the Milk Development Council, UK, and the UK's Department of the Environment, Food & Rural Affairs. Thomassen acknowledges financial support from the Institute of Economic Research of Seoul National University and Overseas Training Expenses for Humanities & Social Science from Seoul National University (SNU).
†Seoul National University, ††Oxford University, IFS, and CEPR, ‡Stanford University, ‡‡London School of Economics

generate larger cross-category effects than a shopper willing to use multiple stores, since the latter can easily switch stores only for the category affected by a price change and not for other categories.

Whether sellers internalize such cross-category effects depends on the organization of supply.[1] In supermarket organization there is a maximal level of internalization, as a single seller sets prices for all categories sold at the same store. In malls, streets or public market places, on the other hand, separate categories have independent vendors—e.g. butchers for meat, bakers for bread, etc.[2] There are some cases with incomplete levels of internalization, such as stores that lease a section of their floor space, and delegate pricing, to an independent category seller.[3]

It has long been recognized that the internalization of complementary pricing effects can substantially mitigate market power. In the monopoly case in Cournot (1838) a single seller of two strictly complementary categories sets an overall Lerner index that is half as high as would arise with two independent sellers. In oligopoly settings—where categories sold by any firm are pricing complements because of the costs to shoppers of buying from multiple firms—internalization can greatly intensify price competition (see Nalebuff (2000)). This is closely related to the finding from the compatibility literature that two multi-product firms may set more competitive prices if their products are incompatible (i.e. have infinite "shopping costs")—so that consumers must buy only from one firm—than when they are not (see Matutes and Regibeau (1988), Economides (1989)).

The market power of supermarkets is an issue of widespread interest. The industry's revenues are a large share of GDP and its behavior affects many interest groups from consumers to suppliers. The analysis of pricing in the supermarket industry has typically been conducted at two alternative levels. The first is the level of the individual supermarket category, e.g. breakfast cereals, alcoholic drinks, etc., where there are concerns that prices are set inefficiently, either too high because of market power (see Hausman et al. (1994), Nevo (2001), Villas Boas (2007), Bonnet and Dubois (2010)), or too low because of a predatory intent or a negative consumption externality (see Griffith et al. (2010)). The second is the level of the retailer as a whole, setting prices across a range of categories (see for example Chevalier et al. (2003) and Smith (2004)), which is the focus of antitrust investigations into supermarket competition (see Competition Commission [CC] (2000, 2008)) and retail merger cases such as the proposed merger of Whole Foods and Wild Oats (considered by the Federal Trade Commission [FTC]).[4] At this level there has been much interest in the growth of large retail firms such as Walmart and Carrefour (see Basker (2007) for a survey). Sometimes public policy has been introduced to protect traditional forms of retail organization such as streets and market places, which do not internalize cross-effects, by curtailing the growth of supermarkets: e.g. in France a law (*Loi Raffarin,* 1996) imposed restrictions on new supermarkets for this purpose.

[1]A well-known example outside of retailing is the selling of component parts for an aeroplane. The proposed GE-Honeywell merger would have resulted in a single seller of two categories (aircraft engines and avionics) and the consequences of internalization of complementary cross-category effects was a central issue in the European Union's approach to the merger. See Nalebuff (2009).

[2]We use the term category or product category to refer to a group of similar product lines that are close substitutes, as in these examples.

[3]For example, retailers such as Sears and Walmart sometimes rent out space within their stores to independent sellers, in return for a rental payment (see Wall Street Journal Sept. 22, 2010). These arrangements are sometimes referred to as "stores within a store" or "in-store concessions".

[4]FTC v. Whole Foods Markets, Inc., 533 F.3d 869 (D.C. Cir. July 29, 2008).

For pricing analysis at each of these two levels it is important to understand the extent to which the internalization of cross-category effects mitigates market power.

A related issue, in the supermarket industry and more generally, is whether consumers that prefer one-stop shopping (i.e. to use a single store) constrain market power more than those that choose multi-stop shopping (who use multiple stores). This has been an important question in prominent antitrust investigations. One possibility is that the former—known in some contexts as "core" or "single-homing" shoppers—have the greater pro-competitive impact, because they generate a relatively large cross-category effect when they change store. The opposite can also be argued, however: multi-stop shoppers may have the greater pro-competitive effect as they find it easier to substitute any individual category between stores. In the UK's CC inquiries into the supermarket industry, some firms claimed that "since supermarkets could not price discriminate [in favor of multi-stop shoppers], these other outlets collectively placed a competitive constraint on the grocery retailer's offer" and that multi-stop shoppers "effectively [...] determined supermarket prices across the board." See CC (2000, paragraph 2.31). According to one of the main firms in the CC investigation "it was the marginal shopper—with the greatest tendency to migrate—who determined prices" and this firm claimed that it "had a high proportion of secondary shoppers and could not be indifferent to them in terms of its price setting." (See CC (2000, paragraph 4.68)). In the US in the proposed Whole Foods/Wild Oats merger the parties to the proposed merger argued that many of their customers "cross-shop" in a wide range of other firms, buying different categories from different stores, and that these multi-stop shoppers constrain prices more than one-stop (or core) shoppers. In these investigations the authorities had to decide whether to focus on promoting competition between retailers that are substitutes for one-stop shoppers, or between retailers combined by multi-stop shoppers. In both cases there was a debate as to which consumer type constrained supermarket prices more, with implications for whether a narrow or wide definition of the market was appropriate for competition analysis.

In this paper we have two main goals. First, we develop a multi-store multi-category model of consumer demand that belongs to a class of models used widely in the theoretical literature to analyze retail pricing, and estimate it using data on shopping choices at consumer-store-category level. Recent demand models used to study retail market power have not considered cross-category externalities, despite the prominence of this issue in the theoretical literature. Second, we use the model to study two policy-relevant issues in retail pricing (as mentioned above): (i) the implications of the internalization of cross-category externalities for market power and (ii) the relative impact on market power of consumers inclined to one-stop and multi-stop shopping. We define categories to correspond to product groups sold by traditional independent sellers of grocery products in streets and public market places (butchers for meat, bakers for bread, etc.) in order to analyze cross-effects that are internalized in supermarket organization but not in a well known alternative organization of supply.

In the model each consumer decides whether to use a single store or multiple stores for their purchases in a given shopping period. For each category a consumer makes a discrete choice of store(s), and a continuous choice of how much to buy. There is differentiation between stores at two levels. The first is at individual category level: the consumer views stores as being different for any category. We allow this differentiation to be partly vertical, reflecting differences in the average quality of stores for any category, and partly horizontal, reflecting variation in individual consumer preferences. The second level of differentiation is at the overall shopping level: each consumer views the fixed costs of shopping at each store differently because of spatial variation in consumer location. For any consumer the benefit of multi-store shopping—going to the best store for each category—must be weighed against the fixed costs of using multiple stores. Consumers inclined to one-stop shopping either have relatively high shopping costs or view a single store as being the best for all categories.

There are two main econometric challenges in estimating the taste parameters that enter category-specific demands. First, a significant number of zero expenditures are observed at category level, so that there are binding nonnegativity constraints in the consumer's continuous category demand problem. Second, given that a consumer's unobserved store-category tastes influence both his choice of store and his category demands, the consumer's unobserved tastes are not independent of the observed characteristics of the stores the consumer selects. To overcome these problems we estimate the consumer's utility parameters in a single step which jointly models both the consumer's nonnegativity constraints and his combined discrete-continuous choice of store and category demand.

The estimated parameters imply complementary pricing effects between categories sold by the same retailer. We calculate the Lerner index of market power implied by these elasticities in Nash equilibrium, using the retailers' first-order pricing conditions. We find that ignoring cross-category effects and analyzing each category in isolation can result in market power being overestimated substantially: accounting for complementary cross-category effects reduces the estimated Lerner index by more than half for most categories and firms. To quantify the externality between product categories (internalized by a supermarket) we compute the implicit marginal (Pigouvian) subsidy per unit of output that must be offered to an independent category seller to ensure it does not increase prices relative to the observed levels (set by supermarkets). We find that the externality is about 47% of the price of the category (on average across firms and categories).[5] This externality is analogous to the "pricing pressure" concept that measures the effects of a merger, introduced in Farrell and Shapiro (2010).[6] The absolute value of our estimates are larger than standard externality levels used to flag an adverse merger, i.e. our estimates indicate that supermarket organization mitigates market power significantly.

To assess whether consumer types inclined (because of their taste type) to one-stop shopping (at observed prices) have a greater competitive impact than those inclined to two-stop shopping,

[5] The presence of large external effects between product categories at a retail location is consistent with the theoretical literature on multi-category sellers, as discussed in Nalebuff (2000), and supported empirically by a study of rental payments in shopping malls in Gould et al. (2005), which found that mall owners offered large rent subsidies to stores that generate a positive externality (by drawing consumers to the mall) for other stores in different product areas.

[6] Supermarket organization (or any form of retailing in which cross-category effects are internalized) can be interpreted as a merger of independent category sellers in a shopping location (see Beggs (1992)). This leads to *downward* pricing pressure, because the categories have complementary cross-price effects, the reverse of the standard upward pricing pressure that follows from a merger of substitutes.

we compare the effect of a marginal price change (for one firm and one category at a time) on the profits from each consumer group. We find that the profit from one-stop shopping types falls and profit from two-stop shopper types increases. This implies that the former have the greater pro-competitive impact. We find that this is a consequence of the greater cross-category effects of one-stop shopping types. Since shopping costs are an important determinant of one-stop shopping we also compare consumers by shopping costs and find similar results: we find that those with high shopping cost have a greater pro-competitive impact than those with low shopping costs. This is consistent with the approach ultimately taken by the CC and FTC in the cases mentioned above, where the focus of the authorities was on maintaining a competitive market for shoppers inclined to one-stop shopping.

The theoretical literature makes extensive use of a multi-store multi-category modeling framework to study retail pricing. Some papers in this literature impose one-stop shopping (Stahl (1982), Beggs (1992), Smith and Hay (2005)) while others model the multi-stop shopping decision (Klemperer (1992), Armstrong and Vickers (2010), Chen and Rey (2012)).[7] The empirical literature on retail market power—in contrast to the theoretical literature, as noted in Smith and Thomassen (2012)—has typically not incorporated cross-category externalities. We adapt the multi-store multi-category theoretical framework for empirical analysis. We develop a model that is multiple-discrete-continuous, in that the consumer can choose one or more (discrete) stores and makes a (continuous) non-negative choice of quantity for every category. We build on the existing literature on multiple-discrete choice (see Hendel (1999), Dubé (2005), Gentzkow (2007)), and discrete-continuous choice (see Dubin and McFadden (1984), Haneman (1984), Smith (2004)).[8] Our multi-category multi-store model brings together the empirical literature that measures market power for a single supermarket category (e.g. Nevo (2001) and Villas Boas (2007)), with the literature on spatial competition between retail outlets in which the choice of category is not modelled (e.g. Smith (2004), Davis (2005) and Houde (2012)). The paper also relates to an established literature in quantitative marketing studying store choice (e.g. Bell et al (1998), Fox et al. (2004)) and another studying multi-category demand (e.g. Chintagunta and Song (2007), Mehta (2006)) by studying these different aspects of demand in a unified multi-store multi-category utility framework.

The rest of the paper is organized as follows. In Section 2 we discuss relevant features of the market and the data. We discuss the model in Section 3 and estimation in Section 4. We report estimates in Section 5 and in Section 6 we analyze supermarket pricing.

[7]Some papers (e.g. Lal and Matutes (1994), Lal and Rao (1997)) use this theoretical framework to study aspects of supermarket pricing (e.g. "Hi-Lo" pricing, advertising of specific product prices, etc.) that we ignore because they are more relevant at a product level rather than the broader category level of analysis we adopt.

[8]These discrete-continuous papers consider a single discrete and a single continuous choice in which zero demand is not allowed. We generalize to allow for multiple continuous choices and we allow for zeros so that the paper is also related to the literature on demand estimation subject to nonnegativity constraints, notably Wales and Woodland (1982) and Kim, Allenby and Rossi (2002).

2 The Market and the Data

Supermarkets became widespread in the US and UK in the mid 20th century. Until then broad grocery categories had been sold by independent sellers in streets, public market places, or through direct delivery to households. The categories used in this paper are defined to correspond approximately to the products sold by these traditional vendors. They are shown in Panel B of Table 2—Bakery, Drink, Fruit & Vegetables, among others—along with indicative products in each category.[9] Thus products in the Bakery category are sold by a traditional baker, Drink in a liquor store, Fruit & Vegetables by a greengrocer, and so on. This definition allows us to analyze cross-effects that are internalized by supermarkets but not in a familiar alternative organization of supply.

We adopt a week as the shopping period in which the consumer plans his shopping. A weekly shopping frequency was found in survey evidence in CC (2000, paragraph 4.77 and Appendix 4.3) in which 982 respondents were asked the question: "How often do you carry out your main grocery shopping?" A large majority (70%) reported a weekly frequency, with 14% less frequently and 16% more frequently. We aggregate store choices and expenditures to the weekly level and assume that decisions on how much to spend in each store are made for the whole week.

To analyze shopping behavior we use data from the TNS Superpanel (now run by Kantar), which records the grocery shopping of a representative panel of households in Great Britain. Our sample is for the three-year period October 2002 - September 2005. The data are recorded by households, who scan the bar code of the items they purchase and record quantities bought and stores used. The grocery items include all products in the categories listed in Table 2 including those sold in irregular weights such as fruit, vegetables and meat. Prices of items bought are obtained from the expenditure and quantity information that the household records, and cashier receipts are used to confirm these prices. 26,191 households participated in the consumer panel in the period of our data with an average of 67.6 weeks recorded per household. Demographic and location information for each household is recorded annually. We treat the household as a single decision-making agent and we use the term consumer to refer to this agent.

In the rest of this section we discuss the construction of the data used for estimation and provide descriptive statistics. The model is too computationally burdensome to estimate on the full sample of consumer-weeks and we therefore select a subsample for estimation. Furthermore we want to maintain multiple observations per consumer in order to use the panel structure of the data. We therefore construct a sample comprising a panel of 2000 consumers and three weeks per consumer. We choose the three weeks for each consumer in such a way that they are spaced at quarterly (13 week) intervals in order to avoid interdependencies between weeks for a given consumer and we pick different sets of weeks (randomly) for each consumer, which allows us to make use of time-series price variation across the full sample period. Further details of sample construction are given in Appendix B where we also show that the estimation sample is representative by comparing demographics between our sample, the full TNS sample and census data.

[9]Supermarkets appear to think about their product offering at a category level when determining price and quality positions: they often define management jobs by category, and thus organize product selection and and pricing decisions this level. (For more discussion of these points, see CC(2008), Appendix 8.1, paragraphs 10-13).

To obtain consumer choice sets we match each consumer to stores based on the distance from the consumer's home. We use a store dataset from the Institute for Grocery Distribution (IGD), which includes the postcode of all supermarket outlets in Great Britain. To compute the distance between consumers and stores we use postcode information in the consumer and store data.[10] We assume that the consumer's choice set is made up of the nearest 30 stores to the consumer's home.[11] The fraction of expenditure on store visits outside the nearest 30 is 1.2%. For each store that is chosen by a consumer the TNS survey indicates the firm (e.g. ASDA, Tesco, etc.) and for stores operated by the main firms it usually records the postcode. The postcode is known for 70% of store choices. When it is not known we assume the consumer goes to a store (in the choice set) operated by the firm they choose.[12]

We compute price indices at firm level rather than store level. This aligns with the policy of firms in the period of the data, which is to set national prices that do not vary by store location.[13] The existence of this pricing policy is helpful as we can use prices observed in any transaction in a given week to compute each firm's (national) weekly price index. We use the full sample of transactions in the TNS data to compute the price indices. To compute the price indices we aggregate over two hierarchical levels, following standard practice in price index construction (see for example Chapter 2 in Office for National Statistics (2014)). At the lower level we compute a price index for a series of narrowly defined product groups, listed in Appendix A (e.g. shampoo is a product group in the Household category), using quantity from the transactions data to weight the individual products. At the upper level we compute a price index for each category (e.g. Household) using sales revenue from the transactions data to weight the lower-level price indices. At both levels the weights are fixed over time to ensure that intertemporal changes in the price index reflect changes in prices rather than composition effects in the weights; at the upper level weights are fixed across firms so that differences between firms in the price index are driven by prices rather than firm-specific weights, which avoids selection bias from the possibility that the consumers selecting a particular firm have tastes that differ from the population. The weights are computed separately for eight demographic types, depending on household size and occupational class, to allow different types of consumers to have different price indices depending on their tastes. The resulting prices are at a firm-category-week-demographic type level. We normalize the prices so that the price of ASDA in week 1 for each category and demographic type is 1. To obtain

[10]Geographic coordinates for every postcode in Great Britain are available from the Postcode Directory, produced by the UK's Office for National Statistics. For each store in the IGD data we therefore have an exact location. The location of each consumer is known at a slightly coarser level (to preserve anonymity), namely the postal sector. We locate each consumer at the average coordinates of the residential postcodes in their postal sector (listed in the Postcode Directory).

[11]The store data include all stores operated by supermarket chains. Where a chain operates more than one very small store in any choice set—defined as having a sales area of less than 10,000 square feet we use only the nearest of these stores to the consumer; this avoids choice sets from filling up quickly with very small stores.

[12]In many cases there is just one candidate store; in cases with more than one we pick one at random, using empirical probability weights that depend on distance and store size. We use the store's predicted probability (conditional on choice of firm) from a reduced form multinomial logit model of store choice, estimated using the full sample of consumers, for consumers whose store choices are known.

[13]"Most retailers set their prices uniformly, or mostly uniformly, across their store network [...]. Various other facets of the retail offer, such as promotions, may also be applied uniformly, or mostly uniformly, across a retailer's store network" (CC (2008), para. 4.98 p. 498-501).

Table 1: Descriptive Statistics: Demographic and Choice Set Characteristics in Estimation Sample

A: Demographic Variables (2000 consumers)					Mean		St. Dev.	
Household size					2.77		1.33	
Household Weekly Income (£) per Head					226.20		119.04	
B: Choice Set Variables (180,000 consumer-week-stores: 6000 consumer-weeks×30 stores)								
B1: Store Characteristics		Stores	Min Dist (km)		Size (1000sqft)		Price	
Firm	Type of Firm	(share)	Mean	St. Dev.	Mean	St. Dev.	Mean	
ASDA	Big Four	0.03	6.54	7.88	45.49	14.04	1.01	
Morrison	Big Four	0.06	6.41	6.80	33.31	8.78	1.06	
Sainsbury	Big Four	0.13	4.72	5.70	31.81	16.57	1.21	
Tesco	Big Four	0.17	3.61	4.56	31.44	22.43	1.10	
M&S	Premium	0.03	6.84	7.75	8.76	1.27	1.81	
Waitrose	Premium	0.04	6.58	6.36	18.71	8.27	1.39	
Aldi	Discounter	0.03	6.76	7.62	8.17	1.67	0.87	
Lidl	Discounter	0.02	10.31	10.17	9.83	3.01	0.78	
Netto	Discounter	0.02	8.76	9.14	6.70	1.76	0.76	
Iceland	Frozen	0.03	4.60	6.18	4.97	1.26	1.12	
Others	Small Chains	0.39	1.33	1.61	13.14	9.16	1.19	
B2: Prices by Category	Bakery	Dairy	Drink	Dry	Fr,Veg	H'hold	Meat	Milk
Mean	1.15	1.21	1.10	1.15	1.30	1.08	1.13	1.13
St. Dev.	0.19	0.17	0.19	0.19	0.23	0.20	0.19	0.07
St. Dev. (within firm)	0.03	0.04	0.03	0.03	0.05	0.03	0.04	0.04

Notes: The unit of observation for the statistics in Panel B is the consumer-week-store. There are 180,000 consumer-week-store observations in the choice sets: 6000 consumer-week choice sets and 30 stores per choice set. Panel B1 presents means across 180,000 consumer-week-stores of: a firm indicator, store size and price (mean across categories within the store), as well as the mean across 6000 consumer-weeks of the minimum (across 30 stores) distance to a store operated by the given firm (omitting consumer-weeks where the firm is not in the choice set). In Panel B2 the price statistics are for the 180,000 consumer-week-store price observations for each category.

category quantities used in estimation we aggregate expenditure to the store-category-week level for each consumer and divide by price.[14] Further details of price index construction are in Appendix C.

Table 1 presents descriptive demographic and choice set statistics for the estimation sample. Panel A reports demographic characteristics for the 2000 consumers.[15] Panels B1 and B2 present statistics on the stores in consumer choice sets. The unit of observation for this part of the table is the consumer-week-store: 2000 consumers, 3 weeks per consumer, and 30 stores per choice set,

[14]See Appendix E for a discussion of conditions that allow aggregation from product to category level.

[15]We use the household income variable to allow price sensitivity to depend on demographics. The TNS data includes a rich list of discrete demographic variables but not income. The UK's Expenditure and Food Survey (EFS) includes a variable for gross current household income (variable p352). We estimate household income by regressing this income variable (for years 2003-2005) on other demographic variables in the ESF that map to those in the TNS survey, namely indicator variables for the number of cars $(0, 1, 2, \geq 3)$, adults $(1, 2, \geq 3)$ children $(0, 1, 2, \geq 3)$, household size $(1, 2, ..., \geq 6)$, geographic region in Great Britain (10 regions), social class (6 classes as described in Appendix C), tenure of residence (dummies for whether the home is privately owned, privately rented, or public housing, structure of residence (detached house, semi-detached/terrace, and apartment), year, sex of the Household Reference Person (HRP), and age of the HRP ($\leq$24, 25-34,35-44,45-54,55-64,$\geq$ 65) We dropped the top and bottom 1% household incomes to avoid outliers. The R^2 is 0.51 and the number of observations in the regression is $17,335$.

yielding 180,000 observations. Each consumer-week has a unique choice set because of geographic and time differences. Panel B1 displays store characteristics by firm. We classify the firms into a number of groups. ASDA, Morrisons, Tesco and Sainsbury are traditional supermarkets and we refer to them as the Big Four. They operate large stores that stock a wide range of products in all categories. M&S and Waitrose have an emphasis on high quality fresh food, and we refer to these as Premium firms. Aldi, Lidl, Netto sell a limited range of grocery products at low prices and are referred to as Discounters. One firm (Iceland) emphasizes frozen food but (like all the firms) it sells all eight categories. The remaining firms (combined in the table as Others) are smaller chains that each have a low market share (namely Co-op, Somerfield, and a group of very minor chains). The table reports the distance from the consumer to the nearest store of each firm: consumers tend to have shorter distances to firms that have many stores. Panel B2 presents the price information at category level. Since category prices are normalized to 1 for ASDA in period 1, a mean category price above 1 (e.g. 1.15 for bakery) indicates that for most weeks and firms the price is higher than in ASDA in period 1. The within-firm standard deviation is due to price variation over time after controlling for firm.

Table 2 reports descriptive statistics on shopping outcomes for the estimation sample. The unit of observation for the table is the consumer-week: 2000 consumers and 3 weeks per consumer. Panel A1 shows that 36% of the consumers use one store per week for all three weeks, while 17% use multiple stores in all three weeks. The remaining 47% of consumers switch in different weeks between using one and multiple stores. When consumers use more than one store in a given week we note two features of the data, presented in Panels A2 and A3 respectively. First, shopping outside the top two stores by spending is minimal (Panel A2). In our model we therefore allow consumers to visit up to two stores per week, and the observed store choices used in estimation are the top two stores by spending. Second, within any individual category, multi-stop shoppers concentrate expenditure in just one store (the identity of which differs by category). Across all consumers (whether one- or multi-stop) the share of category spending in the category's second store is 4% (Panel A3). Our model therefore makes the simplifying assumption that consumers use only one store per category. Accordingly, observed category demands used in estimation are the purchases made in the consumer's main store for the category from the consumer's top two stores.

Panel B shows that a substantial proportion of consumers have zero expenditure for a given category in a given week. This may be because consumers do not wish to purchase all categories every week or because they buy from non-supermarket sellers (e.g. doorstep deliveries are sometimes used for milk in Great Britain). In line with this our model allows for zero expenditures in individual categories.

Panel C illustrates how multi-stop shoppers allocate their spending between firms of different kinds. We consider the three main types of firm introduced in Table 1—Big Four, Discounters, and Premium. We consider the shopping choices of multi-stop shoppers that use, as their top two stores, a Big Four store in combination with (i) a Discounter store or (ii) a Premium store. To represent the Big Four we use ASDA and Tesco respectively. The table shows the proportion

Table 2: Descriptive Statistics: Shopping Outcomes in Estimation Sample

Observations: 6000 consumer-weeks (2000 consumers × 3 weeks)					
A1: Use of Stores (shares sum to one)			Share of consumers		
One store per week for all three weeks			0.36		
One store per week for two of three weeks			0.26		
One store per week for one of three weeks			0.20		
One store per week for none of three weeks			0.17		
A2: Consumer-weeks with >1 store visits			Mean		St. Dev.
Expenditure share in 1st store by weekly spending (store A)			0.71		0.16
Expenditure share in 2nd store by weekly spending (store B)			0.23		0.13
A3: Share of category spending in 2nd store for category from stores (A,B)			Mean		St. Dev.
All consumer-weeks			0.04		0.11
Consumer-weeks with >1 store			0.10		0.15
B: Category Expenditure in 1st Store for category from stores (A,B)		Expenditure (£/wk)			Zeros
Category	*Illustrative Products*	Mean	St. Dev.		(share)
Bakery	Bread, Cakes, Desserts	3.47	3.33		0.09
Dairy	Cheese, Yogurt, Butter	3.33	3.27		0.17
Drink	Wine, Spirits, Lager, Cola	4.91	8.20		0.31
Dry Grocery	Breakfast Cereals, Confectionery, Coffee	5.68	5.24		0.10
Fruit & Vegetables	Fruit and Vegetables (including frozen)	7.11	6.09		0.06
Household	Pet Food, Detergents, Toilet Tissues	6.19	6.98		0.21
Meat	Ready Meals, Cooked Meats, Fresh Beef	10.34	9.00		0.07
Milk	Low Fat Fresh Milk, Organic Fresh Milk	1.17	1.29		0.30
All Categories		42.21	27.02		0.00
C: Category choices, Consumer-weeks with >1 store					
Top two stores (A,B)		Tesco/ASDA top store for category (1/0)			
		Bakery	Drink	H'hold	Meat
Tesco & Discounter		0.80	0.51	0.71	0.66
Tesco & Premium		0.65	0.73	0.86	0.55
ASDA & Discounter		0.68	0.44	0.59	0.72
ASDA & Premium		0.60	0.80	0.89	0.59

Notes: Taylor Nelson Sofres (TNS) Superpanel survey of consumers in Great Britain, October 2002 - September 2005. The statistics are calculated at consumer-week level (i.e. we aggregate expenditures to the week) for the 6000 consumer-weeks (2000 consumers and 3 weeks per consumer) used in the estimation sample. Stores A and B referred to in the table are the consumer's first and second stores by overall spending in any week. In Panel B the illustrative products in each category are from TNS's list of 269 most granular product classifications; the full list of such products by category is shown in Appendix A. In Panel C the (1/0) dummies take the value 1 if the consumer's top store (by expenditure) for the category is ASDA or Tesco where the figures displayed are averages for multi-stop shoppers whose top two stores (by spending) are as listed the first column.

of these shoppers that use these Big Four firms as their main store for the category. Thus for example 80% of Tesco-Discounter shoppers select Tesco for Bakery, 49% of the same group of shoppers select the Discounter for Drink, and so on. A pattern emerges in which the Discounter is strong (relative to the Big Four supermarkets) in categories where products tend to be non-perishable (e.g. Drink and Household goods), but less strong in perishable categories (e.g. Bakery and Meat). The Premium firms have the opposite pattern: relatively strong for perishables and relatively weak for non-perishables.

3 Utility and Demand

3.1 A Simple Multi-store Multi-category Model

We start with a simple version of the model to build intuition for the price incentives in the full model. Suppose there are two stores A and B and each stocks three categories 1, 2 and 3.[16] A consumer has unit demand for every category, and selects a store for each one. Let the store-category choice (A, B, A) indicate store A for 1, store B for 2 and store A for 3. There are eight possibilities: $(A, A, A), (A, A, B), \ldots, (B, B, B)$. Let p_{jk} denote the unit price at store $j = A, B$ for category $k = 1, 2, 3$. The consumer's sensitivity to price is α.

Suppose utility is additively separable in categories and that a unit of category k at store j gives a gross utility μ_{jk}. A shopping cost $\Gamma \geq 0$ is incurred if the consumer visits both stores (two-stop shopping) and avoided if he visits one (one-stop shopping). Thus, for example, choices (A, A, A) and (A, B, A) respectively give the following utilities net of price:

$$\begin{aligned} U(A, A, A) &= (\mu_{A1} - \alpha p_{A1}) + (\mu_{A2} - \alpha p_{A2}) + (\mu_{A3} - \alpha p_{A3}) \qquad \text{and} \\ U(A, B, A) &= (\mu_{A1} - \alpha p_{A1}) + (\mu_{B2} - \alpha p_{B2}) + (\mu_{A3} - \alpha p_{A3}) - \Gamma. \end{aligned}$$

If the consumer has no shopping cost ($\Gamma = 0$) he selects the store with the highest net utility ($\mu_{jk} - \alpha p_{jk}$) for each category, independent of the store chosen for the other categories. Alternatively, if his shopping costs are positive ($\Gamma > 0$) his category choices are interdependent, as he may give up the benefit of shopping around to avoid the shopping cost. This interdependence tends to generate complementary effects between categories at the same store: a price increase for one category at the store may induce the consumer to buy *all* categories from the other store.

Consumer tastes—characterized by store-category preferences μ_{jk} (for all j and k), price sensitivity α, and shopping cost Γ—are heterogeneous in the population. Consumers of different (μ, α, Γ)-type respond to a price increase in different ways, which vary in the extent of the cross-category effects they imply. Consider a price increase at store A for category 1. Any consumer that initially bought category 1 at store A, and is marginal in the sense that he stops buying it there after the price increase, can be classified into one of the following four exhaustive response

[16] It is common in the multi-store multi-category theory literature, discussed in the introduction, to assume $J = 2$ and $K = 2$ and to define an individual consumer's tastes as a point in a unit square. In this subsection we use $K = 3$ because one of the consumer responses below, namely (2b), is impossible with $K = 2$.

classes:

1. Initial one-stop shopper: (A, A, A)

    (a) Drop store A for all categories: change to (B, B, B).

    (b) Retain A but drop it for at least category 1: change to (B, A, A), (B, B, A) or (B, A, B).

2. Initial two-stop shopper: (A, A, B) or (A, B, B) or (A, B, A).

    (a) Drop store A for all categories: change to (B, B, B).

    (b) Retain store A but drop it for category 1: change to (B, A, B) or (B, B, A).

Of these four responses, (1a) has the maximal cross-category effect: consumers in this class transfer all three categories from the store. At the other extreme, response (2b) has no cross-category effect. The response class into which a consumer falls depends on his (μ, α, Γ)-type. Consumers with very high shopping costs Γ, for example, are likely to fall into class (1a) as they strongly prefer to use a single store. Consumers with strong (and varying) store preferences $\mu_{jk} - \alpha p_{jk}$ by category, on the other hand, are likely to fall into (1b) and (2b) as they are not willing to transfer all their category demands away from store A even though by doing so they would avoid shopping costs.

We can now relate the model to the two main questions we study in Section 6. First, when margins are positive, complementarity between categories has a pro-competitive effect if it is internalized—as is the case in supermarket organization—and the magnitude of this effect depends on the distribution of consumer preferences (e.g. it is greater the larger is the proportion of consumers in class (1a) relative to class (2b)).

Second, we can compare the pro-competitive effects of consumer types that choose one-stop and two-stop shopping respectively by comparing how much each group of shoppers punishes the supermarket for the price increase. We have just seen that, conditional on being marginal (i.e. responding to the price change), initial one-stop shoppers tend to have larger cross-category externalities than two-stop shoppers. This does not, however, imply that one-stop shoppers penalize the firm more than two-stop shoppers, because a relatively low proportion of one-stop shoppers may be marginal: a one-stop shopper, unlike a two-stop shopper, cannot switch an individual category (say category 1) between stores A and B without incurring shopping costs Γ, as he initially does not use both stores. Whether one-stop shoppers penalize the firm more than two-stop shoppers thus depends not just on the magnitude of the cross-category externalities per marginal shopper but also on the proportion of each shopper group that is marginal.

3.2 Full Demand Model

We now specify the full model, which generalizes the number of stores and categories and relaxes the unit demand and additive separability assumptions for categories. We also allow shopping costs to depend on distances to the chosen stores. At a general level one can think of the consumer's weekly problem as choosing how much to purchase in each category at any store. To make the model tractable, based on the data patterns presented in Section 2, we assume that consumers visit no more than two stores, use only one store per category, and only stores that are among the thirty nearest to the consumer (denoted by the set $\mathcal{J}$).[17] In Appendix D we show how, based on these restrictions, the utility specification presented below can be derived from a general model of quantity choice under a budget constraint.

We call the one or two stores visited by a consumer his shopping choice, which we denote by c and write as a set: if c has two stores j and j' then $c = \{j, j'\}$ and if it has one store j then $c = \{j\}$. Let $n(c)$ denote the number of stores in c. The set of available shopping choices $\mathcal{C}$ comprises all unordered pairs and singletons from the set of available stores $\mathcal{J}$. For each shopping choice the consumer has a shopping cost $\Gamma(c)$, which depends on the number $n(c)$ of stores in c and their locations. Shopping costs include the financial, time, and psychological costs involved in shopping at the stores in c.

There are K demand categories, indexed by k, at each store. For each category the consumer selects a store $j \in c$. The store-category choices are summarized in d, a vector that lists the store chosen from shopping choice c for each category. As an illustration suppose $K = 3$, as in subsection 3.1, and $c = \{A, B\}$. Then $d = (A, B, A)$ is an example of a store-category choice. We write $\mathcal{D}_c$ for the set of possible alternatives for d given shopping choice c. For each category the consumer makes a non-negative continuous quantity choice. The quantity choices are given by the $K \times 1$ vector q. Let p be the full vector of store-category prices p_{jk} and μ the full vector of store-category tastes μ_{jk}.

The consumer's utility from shopping choice c, store-category choice d, and quantity q, at prices p, is given by

$$U(c, d, q, p) = u(q, d) - \alpha p_d' q - \Gamma(c) + \varepsilon_c \tag{1}$$

$$= (\mu_d - \alpha p_d)' q - 0.5 q' \Lambda q - \Gamma(c) + \varepsilon_c \tag{2}$$

where $u(q, d) = \mu_d' q - 0.5 q' \Lambda q$. The $K \times 1$ vectors μ_d and p_d collect the tastes and prices (respectively) that are relevant for each category given store-category choice d. Λ is a symmetric $K \times K$ matrix of parameters, common across consumers.[18] The first two terms in (1) are variable utility (in terms of q): $u(q, d)$ is gross utility from the categories bought, and $p_d' q$ is the consumer's total payment for them. The price sensitivity scalar α corresponds to the marginal utility of expenditure on non-supermarket consumption (in which utility (1) is quasi-linear).

The full model has two sources of product differentiation. First, as in the simple model, there is differentiation between stores at category level, as the consumer views stores differently for any category (captured in the store-category taste vector μ). Second, there is differentiation across shopping choices c at the level of fixed utility, captured in $\Gamma(c)$, which now (unlike the

[17] These assumptions can in principle be relaxed in our framework. A relaxation of the second assumption would allow the consumer to select two stores (each with a nonnegative quantity) for each category. This can be accommodated by extending the quadratic utility specification to allow the number of continuous quantities to be $2K$ instead of K when $n(c) = 2$, with extra second-order parameters that govern inter-store intra-category substitution. This can also be accommodated in the econometric framework in Section 4. Given that category spending in the category's second store is low (see Section 2) we decided not to generalize in this way.

[18] Unlike many forms (e.g. AIDS), the quadratic is suitable for our purposes as it can naturally accommodate zero demands at category level. Quadratic utility demand is used in Wales and Woodland (1982).

simple model) includes spatial variation in store locations relative to the consumer. A consumer's (μ, α, Γ)-type fully characterizes his tastes up to ε_c, a iid type-1 extreme value term which is iid across store choices and which captures any residual unobserved utility that arises from the shopping choice. Tastes (μ, α, Γ) vary in the population of consumers; we specify how in Section 3.4.

The consumer maximizes $U(c, d, q, p)$ by selecting c, d, and q. The shopping choice c that gives the highest total utility net of shopping costs is obtained by solving

$$\max_{c \in \mathcal{C}} [w(c, p) - \Gamma(c) + \varepsilon_c] \tag{3}$$

where

$$w(c, p) = \max_{d \in \mathcal{D}_c} \max_{q \in \mathbb{R}^K_{\geq 0}} [(\mu_d - \alpha p_d)' q - 0.5 q' \Lambda q] \tag{4}$$

is the consumer's indirect variable utility function (i.e. the maximum variable utility from a choice of (d, q) given shopping choice c and prices p.

The outer maximization problem in (4) determines the vector d of store-category choices. Note that for any category the consumer always does best to select the store $j \in c$ with higher linear term in the quadratic equation (2) so that we can write the optimal store-category choices given shopping choice c and prices p as[19]

$$\begin{aligned} d(c, p) &= [d_1(c, p), \ldots, d_K(c, p)] \\ &= [\arg\max_{j \in c}(\mu_{j1} - \alpha p_{j1}), \ldots, \arg\max_{j \in c}(\mu_{jK} - \alpha p_{jK})]. \end{aligned} \tag{5}$$

The inner nested maximization problem in (4) implies a system of K category demands conditional on store-category choice vector d

$$q(d, p) = \arg\max_{q \in \mathbb{R}^K_{\geq 0}} [(\mu_d - \alpha p_d)' q - 0.5 q' \Lambda q] \tag{6}$$

where the consumer may choose a zero demand for any k. We write the k-the element of $q(d, p)$ as $q_k(d, p)$.

Thus indirect variable utility (4) is $w(c, p) = (\mu_{d(c,p)} - \alpha p_{d(c,p)})' q(c, p) - 0.5 q(c, p)' \Lambda q(c, p)$ where $q(c, p) = q(d(c, p), p)$. Expressions (3), (5), and (6) give our model's predictions of consumer behavior.

3.3 Cross-Category Effects at Store Level

In this subsection we return to the discussion in 3.1 of how consumers respond to a price change. To do this we aggregate across shopping choices (c) to present the consumer's behavior at store level (j), which is the level most relevant for thinking about a store's pricing incentives. (We note

[19] Optimal d is invariant in the level of q. This follows by the absence of store-specific effects in Λ.

that the discussion in this subsection is useful for understanding the pricing analysis of Section 6 but does not develop anything that enters our estimator.)

First consider continuous demand responses conditional on store-category choice d. If d is such that the consumer chooses store j for category k, then the solution to the problem in (6) implies

$$q_k(d,p) = \max\left[\frac{1}{\Lambda_{kk}}\left(\mu_{jk} - \alpha p_{jk} - 0.5\Sigma_{k'\neq k}\Lambda_{kk'}q_{k'}(d,p)\right), 0\right]. \tag{7}$$

Equation (7) illustrates how category demands respond to price (holding d constant). The diagonal second-order quadratic terms, i.e. Λ_{kk} for any k, scale demand and (since α is fixed across categories) allow own-price effects to vary across categories.[20] The off-diagonal second-order terms $\Lambda_{kk'}$ determine cross-category effects between k and k' (a positive value indicates substitutes, a negative value complements).

We now consider the consumer's demand for category k at the level of store j. We aggregate over the set $\mathcal{C}_j$ of shopping choices that include store j:

$$Q_{jk}(p) = \sum_{c\in\mathcal{C}_j}\left\{q_k(d(c,p),p)\times 1[j = d_k(c,p)]\times 1[c = \arg\max(w(c,p) - \Gamma(c) + \varepsilon_c)]\right\}. \tag{8}$$

This expression does not condition on the chosen c or d and hence allows both to change in response to a price change. The right hand side of (8) therefore allows us to identify three distinct types of consumer response that follow a marginal increase in store j's category k price p_{jk}: (i) an intensive margin change in the consumer's continuous conditional demand q_k holding store-category choice d constant; (ii) a change in the store-category choice d_k for category k, holding shopping choice c constant; and (iii) a change in shopping choice c. In the simple example in 3.1 two of these responses were present: (ii) and (iii). Response (i) is now added because we allow for continuous demands. As was the case in the simple example in 3.1 complementary cross-category pricing effects are generated at store j via demand response (iii): a shopper may switch store for all products because the increase in the price of product k. In the full model there may be further cross-category effects via response (i) which can be of either sign depending on the off-diagonal second-order quadratic utility parameters $\Lambda_{kk'}$ in equation (7).

3.4 Specification of Consumer Type Heterogeneity

In this subsection we specify how tastes (μ, α, Γ) vary across consumers i and weeks t. We now introduce consumer and time subscripts.

We begin with consumer i's taste at time t at store j for category k which is written in terms of observed and unobserved taste-shifters:

$$\mu_{itjk} = \xi_{fk} + \beta_{0k}\left(\beta_1 hz_i + \beta_2 sz_j + \beta_T T_t + \sigma_1\nu_i^{\mu} + \sigma_2\nu_{it}^{\mu} + \sigma_3\nu_{ik}^{\mu} + \sigma_4\nu_{ijk}^{\mu}\right). \tag{9}$$

The firm category effect ξ_{fk} is common to all consumers and may vary by firm (f) and category (k)

[20] The specification thus allows the demand elasticity (conditional on store-category choice d) to vary across categories, as the slope and intercept both have a distinct parameter for each category.

because different firms do not offer the same branded products, and because many products (e.g. private labels) are firm-specific (see e.g. Corstjens and Lal (2000)). Variation in firm-category strengths was suggested by the data presented in Section 2.[21]

The remainder of μ_{itjk}, i.e. $\beta_{0k}(\beta_1 hz_i + \ldots)$, allows the utility for consumer i and store j at time t to deviate from the firm-category mean ξ_{fk}. In the interest of parsimony the parameters (but not all of the random terms) in this component are common across categories, up to a scaling term β_{0k} which allows the size of the effect to vary across categories (to normalize we set $\beta_{0k} = 1$ for $k = K$). Note that the scaling term allows the variance of the random utility shocks to differ across categories.

The observable variables are: household size hz_i, which allows continuous demand to be greater for larger households, the log of the store's floor space sz_j which allows larger stores to offer greater quality (e.g. because of a better selection of products), and quarter and year dummies T_t, which allow for seasonal and year effects.[22] The remaining terms in (9) are four random taste components (each iid $N(0,1)$): a consumer effect ν_i^μ, a consumer-time effect ν_{it}^μ, a category-specific effect ν_{ik}^μ, and a store-category effect ν_{ijk}^μ. These are scaled by parameters $(\sigma_1, \ldots, \sigma_4)$. The last of these introduces (horizontal) product differentiation at store-category level, allowing each consumer to view stores differently for any category. Thus store-category differentiation is partly vertical, reflecting differences in the average quality of stores for any category, and partly horizontal, reflecting variation in individual consumer preferences. Finally note that (i) with the exception of ν_{it}^μ the unobserved heterogeneity through the ν-terms in utility are constant across time, capturing permanent unobserved effects and (ii) the first two terms ($\nu_i^\mu$and ν_{it}^μ) are common across categories and hence allow for correlation in the unobserved taste shocks across categories.

The price coefficient α_i is specified to allow heterogeneity in price sensitivity

$$\alpha_i = (\alpha_1 + \alpha_2 / (y_i/hz_i))\, \nu_i^\alpha \tag{10}$$

where y_i is household i's income, hz_i is household size. ν_i^α is a Rayleigh(1) random shock which introduces heterogeneity in a parsimonious way while ensuring positive price sensitivity $\alpha_i > 0$ for all i, as long as α_1 and α_2 are positive.

Shopping costs are given by

$$\Gamma_{it}(c) \;=\; \left(\gamma_{11} + \gamma_{12}\nu_{i1}^\Gamma\right) 1_{[n(c)=2]} + \left(\gamma_{21} + \gamma_{22}\nu_{i2}^\Gamma\right) dist_{ic} \tag{11}$$

where $n(c)$ is the number of stores in c and $dist_{ic} = 2\sum_{j\in c} dist_{ij}$ is the total distance of traveling to each store and back. ν_{i1}^Γ and ν_{i2}^Γ are each iid $N(0,1)$.

[21]The main firms are listed in Table 1 in Section 2. To economize on ξ parameters we aggregate two groups of smaller firms: the "Discounters" (Aldi, Lidl, Netto), which have a similar quality position across categories, and the Others, which are smaller chains (namely Co-op, Somerfield, and minor chains). This results in 9 firms (or firm groups) that have a distinct ξ for each k: ASDA, Morrison, Sainsbury, Tesco, M&S, Waitrose, Iceland, Discounters, and Others.

[22]The use of a firm dummy with a store size variable to pick up the quality on offer at a store is consistent with the following quotation from CC(2008): "Product range for many retailers is also, in large part, uniform across stores with variations for the most part being a function of store size [...]" (para. 6.33).

The model allows for correlation in unobserved utility between alternative shopping choices c with stores in common through the ν_{ijk} terms that enter (9), as well as through the unobserved components of the shopping cost term (11). The idiosyncratic term (ε_{itc}) in equation (2) is iid for each (i, t, c) combination and captures any residual unobserved utility that arises from the shopping choice.[23]

Consumers are observed in our data only if they have positive expenditure in at least one category (see Section 2). In theory it is possible for our model to predict that a consumer visits a shopping choice c but makes zero purchases, for example as a consequence of a high price sensitivity shock ν_i^α or a low overall demand shock (ν_i^μ or ν_{it}^μ). Our model is flexible enough, however, to ensure that this happens only with a negligible probability (0.014 at our parameter estimates), thereby closely matching what is observed in the data.

4 Estimation and empirical strategy

The full set of parameters to be estimated is $\theta = (\theta^w, \gamma)$, where $\theta^w = (\beta, \xi, \sigma, \alpha, \Lambda)$ groups the parameters in variable utility and γ those in shopping costs. In the same way, let $x_{itc}^w = (x_{itj}^w)_{j\in c}$ and $\nu_{itc}^w = (\nu_i^\mu, \nu_{it}^\mu, \nu_{ik}^\mu, \nu_{ijk}^\mu, \nu_i^\alpha)_{j\in c, k=1,\ldots,K}$ be the observables (including prices) and taste shocks in variable utility, and $(x_{itc}^\Gamma, \nu_i^\Gamma)$ be the observables and taste shocks (excluding ε_{itc}) entering fixed shopping costs. Finally, we define $x_{it} = (x_{itc}^w, x_{itc}^\Gamma)_{c\in\mathcal{C}_{it}}$, $\nu_{it} = [(\nu_{ict}^w)_{c\in\mathcal{C}_{it}}, \nu_i^\Gamma]$, and $\varepsilon_{it} = (\varepsilon_{itc})_{c\in\mathcal{C}_{it}}$.

We make use of two core demand expressions in estimation. These are based on $w(c, p)$, $d(c, p)$ and $q(d, p)$, as given in (4), (5) and (6), now with i and t subscripts added to indicate dependence on consumers and time as described in Section 3.4. The first is the quantity demanded of category k in store j (by i in week t) conditional on shopping choice c:

$$q_{cjk}(\theta^w, x_{itc}^w, \nu_{itc}^w) = q_{itk}(d_{it}(c, p_t), p_t) \times 1[j = d_{itk}(c, p_t)]. \tag{12}$$

Since this conditions on shopping choice, the arguments $(\theta^w, x_{itc}^w, \nu_{itc}^w)$ of this function relate to variable utility but not shopping costs. The second demand expression is the indicator for whether consumer i in week t chooses shopping choice c:

$$I_c(\theta, x_{it}, \nu_{it}, \varepsilon_{it}) = 1[w_{it}(c, p_t) - \Gamma_{it}(c) + \varepsilon_{itc} \geq w_{it}(c', p_t) - \Gamma_{it}(c') + \varepsilon_{itc'},\ \forall c' \in \mathcal{C}_{it}]. \tag{13}$$

The arguments $(\theta, x_{it}, \nu_{it}, \varepsilon_{it})$ of this function relate to both variable utility and shopping costs.

Two important issues arise when estimating the class of demand model proposed here. First, since consumers self-select into which store to visit, and a common set of parameters and unob-

[23] The scale of the parameters is determined by normalizing the parameter on the random shopping cost disturbance ε to unity so that it is a Type-1 Extreme Value draw. Note from (7) that conditional demands are homogeneous of degree zero in parameters (μ, α, Λ), i.e., writing demand as a function of parameters, $q(d, p; \kappa\mu^*, \kappa\alpha^*, \kappa\Lambda^*) = q(d, p; \mu^*, \alpha^*, \Lambda^*)$, where $(\mu^*, \alpha^*, \Lambda^*)$ represents some arbitrary value. This does not, however, allow another normalization because variable utility (4) is homogeneous of degree one in the same parameters, i.e. $w(c, p; \kappa\mu^*, \kappa\alpha^*, \kappa\Lambda^*) = \kappa w(c, p; \mu^*, \alpha^*, \Lambda^*)$, so that their scale κ determines the relative importance of variable utility as compared to shopping costs in the consumer shopping choice problem (3). Since we have not normalized (μ, α, Λ) we do not need a further parameter to multiply $w(c, p)$ in (3).

servables drive both store and quantity choice, there is a selection issue when estimating quantity choice (12). Secondly, corner solutions, where a consumer does not have positive demands for all categories, occur often in the data.

We deal with the first issue by simultaneously estimating all the parameters in a framework where store and quantity choices are jointly determined by the combination of (12) and (13). Smith (2004) uses a similar approach in a setting with a single continuous choice. Importantly, in our model, the variables that determine shopping choice are not the same as those that determine quantity choice: distance enters shopping costs but is excluded from variable utility. This exclusion restriction is helpful to identify the parameters in shopping costs separately from those in variable utility.[24]

We address the second issue, zero category demands, by explicitly modeling the nonnegativity constraints in the consumer's category demand problem (6). For each combination of binding and nonbinding constraints, we find the quantities that satisfy the relevant optimality conditions, and then choose the solution that yields the maximum variable utility, $w_{it}(c, p)$.[25] As can be seen from (13), to form a prediction of whether c is chosen, we need to find $w_{it}(c', p)$ for each $c' \in \mathcal{C}_{it}$. In comparison to standard discrete-choice models where the utility of each alternative is a closed-form function of the parameters, the calculation of $w_{it}(c', p)$ (as the solution to a quadratic programming problem with nonnegativity constraints) adds a new layer of computation that must be carried out for each draw of consumer tastes.[26]

Our approach to the two issues (selection and corner solutions) brings together two strands of demand modeling: (i) the discrete-continuous literature that jointly models a (discrete) store and (continuous) category choice (but does not allow zero category demands), e.g. Dubin and McFadden (1984) and Smith (2004), and (ii) the literature that models non-negativity constraints explicitly (but does not consider store choices), e.g. Wales and Woodland (1982) and Kim et al. (2002).

We estimate the model's parameters using the generalized method of moments (GMM) rather than maximum likelihood.[27] Both methods involve calculating the probability of the consumer making the observed shopping choice c, i.e. the expected value of (13). As discussed above, this is computationally demanding because of the presence of corner solutions, and would be prohibitive

[24]We also note that distance is a strong predictor of store choice. The literature on selection models (see Moffitt (1999)) emphasizes the importance of excluded variables that are strongly correlated with the selection outcome (store choice in our setting) and excluded from the outcome equation (quantity choice). Distance plays the role of such an exclusion restriction on our setting.

[25]For a simple example, if $K = 2$, there are 2^K such combinations: (i) $q_1 > 0, q_2 > 0$, (ii) $q_1 > 0, q_2 = 0$, (iii) $q_1 = 0, q_2 > 0$, (iv) $q_1 = q_2 = 0$. In each case (i)-(iii), a first-order condition implies a quantity for the nonzero categories, e.g. $\mu_1 - \alpha p_1 - \Lambda_{11} q_1 = 0$ for (ii). If these quantities are all positive, this is a candidate solution. The final step is to compare the utilities resulting from these candidate solutions as well as from case (iv). In the general setting where $K = 8$, there are $2^8 = 256$ combinations for each consumer-week/shopping choice/random draw if all K categories are interacted in utility.

[26]In a standard multinomial logit discrete-choice framework, the choice probability of c is $e^{v_c} / \sum_{c'} e^{v_{c'}}$ where utility has a closed form: $v_c = x_c\theta$. In our model, utility depends on a new level of optimization, with no closed-form solution: $v_c = \Gamma_c + \max_d \max_{q \geq 0}[(\mu_d - \alpha p_d)' q - 0.5 q' \Lambda q]$. Thomassen (2009, 2017) uses a simpler version of the same structure, where $v_c = \mu_c + \max_k [x_{ck}\theta]$ and x_{ck} is horsepower (and price) of engine variant k of car model c.

[27]The papers cited in the previous paragraph mostly use a likelihood approach. The likelihood for our model nests the likelihoods in those papers and is derived in Appendix G.

for more than a small number of draws of ν for each consumer. Accordingly, we use the method of simulated moments, which has the advantage of being consistent for a fixed number of draws of ν (see McFadden (1989), Pakes and Pollard (1989)).

A standard concern is that prices depend on unobserved factors that affect demand. To deal with this possibility, we include fixed effects ξ_{fk} to control for market-level unobserved quality for each firm-category combination, and time dummies to control for market-level year and quarter effects. We believe this ensures that market-level demand shocks are included in the model, and therefore assume that there is no dependence between price and the unobservables ν and ε.

To construct our estimator, we use (12) and (13) to form conditional expectations of choice outcomes that we observe in the data. (We condition on the explanatory variables x, but not on choice variables like shopping choice.) In this paragraph, to bring out the underlying structure shared by all our moment conditions, we let Y^* denote any choice outcome that we use as a dependent variable, e.g. quantity purchased of k in j within c. (For simplicity we drop subscripts.) For each dependent variable Y^*, we assume that its value is determined by x, ν and ε through a known relationship that is given by our structural model evaluated at the true parameter value θ_0:[28]

$$Y^* = Y(\theta_0, x, \nu, \varepsilon), \tag{14}$$

$$\nu, \varepsilon | x \sim \text{with known density } f(\nu, \varepsilon | x). \tag{15}$$

In light of the discussion in the previous paragraph, we assume independence between (ν, ε) and x so that $f(\nu, \varepsilon|x) = f(\nu, \varepsilon)$.[29] We can write the population conditional expectation in terms of the model's primitives as

$$Y(\theta, x) = \mathrm{E}[Y(\theta, x, \nu, \varepsilon)|x] = \int\int Y(\theta, x, \nu, \varepsilon) f(\nu, \varepsilon|x) d\varepsilon d\nu, \tag{16}$$

so that the following condition, which we use for estimation, holds:[30]

$$\mathrm{E}[Y^* - Y(\theta_0, x)|x] = 0. \tag{17}$$

As we explain in more detail in subsections 4.1-4.3, we use moment conditions, all of which are of the type (17), for a number of different dependent variables Y^*, to construct a GMM estimator.[31] For each dependent variable we use a star, as in Q^*_{itcjk}, to indicate both a variable's realization in our sample and the corresponding population variable. To denote the conditional expectation (16) given the explanatory variables, we use the variable's symbol (without a star) and with θ and x

[28] For instance, when Y^* is the quantity Q^*_{itcjk} purchased of category k in store j and in shopping choice c, $Q^*_{itcjk} = Q_{cjk}(\theta_0, x_{it}, \nu_{it}, \varepsilon_{it}) = q_{cjk}(\theta_0^w, x^w_{itc}, \nu^w_{itc}) I_c(\theta_0, x_{it}, \nu_{it}, \varepsilon_{it})$.

[29] We note that (14) and (15) are the type of assumptions needed to form a maximum likelihood estimator. Therefore our GMM estimator is based on exactly the same structural assumptions that would be used for MLE.

[30] This is an instance of the general fact that $\mathrm{E}[y - \mathrm{E}(y|x)|x] = 0$. To see why this is true here, first note that since $f(\nu, \varepsilon|x)$ is a density, $\int\int f(\nu, \varepsilon|x) d\varepsilon d\nu = 1$. Then, the left-hand side of (17) can be rewritten, using (14) and (16), as $Y(\theta_0, x) - \int\int Y(\theta_0, x) f(\nu, \varepsilon|x) d\varepsilon d\nu = Y(\theta_0, x) - Y(\theta_0, x) \int\int f(\nu, \varepsilon|x) d\varepsilon d\nu = 0$.

[31] These are the conditional moment restrictions (22)-(24) and the unconditional moment restrictions (32)-(36) and (41)-(42).

as arguments, e.g. $Q_{cjk}(\theta, x_{it})$. We also informally discuss the identification of the parameters in 4.1-4.3. Subsection 4.4 details how the moment conditions are combined in the estimator.

4.1 Within-period moment conditions

A distinctive feature of our data, compared to many discrete-choice applications, is that choice sets are different for each individual consumer because of location differences. This results in variation in store characteristics and household attributes that is useful for estimating the effects of these variables on choice outcomes. In this subsection we discuss moment conditions that rely primarily on this type of cross-sectional variation, while the next two subsections consider the role of repeated choices by the same consumer.

We start by defining the three dependent variables discussed in this subsection. For each consumer-week it, let Q^*_{itcjk} be the quantity purchased of category k in store j and shopping choice c. D^*_{itcjk} equals one if that quantity is positive and zero otherwise, and I^*_{itc} equals one if c is chosen and zero otherwise. (For instance, if i in t visits j and no other store, then $Q^*_{itcjk} = D^*_{itcjk} = I^*_{itc} = 0$ for all $c \neq \{j\}$.)

Before defining the conditional expectation functions (16) for the three dependent variables, we note that the conditional expectation of (13) given (x_{it}, ν_{it}) can be written as

$$P_c(\theta, x_{it}, \nu_{it}) = \int I_c(\theta, x_{it}, \nu_{it}, \varepsilon_{it}) f(\varepsilon_{it}) d\varepsilon_{it} = \frac{\exp[w_{it}(c, p_t) - \Gamma_{it}(c)]}{\sum_{c' \in \mathcal{C}_{it}} \exp[w_{it}(c', p_t) - \Gamma_{it}(c')]}, \tag{18}$$

by the assumption that ε_{itc} is iid type-1 extreme value for each c and independent of ν_{it}. We can now write the conditional expectation functions as:[32]

$$\begin{aligned} Q_{cjk}(\theta, x_{it}) &= \int q_{cjk}(\theta^w, x^w_{itc}, \nu^w_{itc}) P_c(\theta, x_{it}, \nu_{it}) f(\nu_{it}) d\nu_{it} && (19) \\ D_{cjk}(\theta, x_{it}) &= \int 1[q_{cjk}(\theta^w, x^w_{itc}, \nu^w_{itc}) > 0] P_c(\theta, x_{it}, \nu_{it}) f(\nu_{it}) d\nu_{it} && (20) \\ I_c(\theta, x_{it}) &= \int P_c(\theta, x_{it}, \nu_{it}) f(\nu_{it}) d\nu_{it}. && (21) \end{aligned}$$

The conditions (17) are now:

$$\begin{aligned} \mathrm{E}\Big[Q^*_{itcjk} - Q_{cjk}(\theta_0, x_{it}) | x_{it}\Big] &= 0, \text{ for } k = 1, \dots, K, && (22) \\ \mathrm{E}\Big[D^*_{itcjk} - D_{cjk}(\theta_0, x_{it}) | x_{it}\Big] &= 0, \text{ for } k = 1, \dots, K, && (23) \\ \mathrm{E}\Big[I^*_{itc} - I_c(\theta_0, x_{it}) | x_{it}\Big] &= 0. && (24) \end{aligned}$$

Expressions (22)-(24) imply orthogonality conditions (between the prediction errors and functions of the explanatory variables) that we use, by the analogy principle, to form the following empirical

[32]To see this for (19): $Q_{cjk}(\theta, x_{it}) = \int\int q_{cjk}(\theta^w, x^w_{itc}, \nu^w_{itc}) I_c(\theta, x_{it}, \nu_{it}, \varepsilon_{it}) f(\nu_{it}) f(\varepsilon_{it}) d\varepsilon_{it} d\nu_{it} = \int q_{cjk}(\theta^w, x^w_{itc}, \nu^w_{itc}) \left[\int I_c(\theta, x_{it}, \nu_{it}, \varepsilon_{it}) f(\varepsilon_{it}) d\varepsilon_{it}\right] f(\nu_{it}) d\nu_{it} = \int q_{cjk}(\theta^w, x^w_{itc}, \nu^w_{itc}) P_c(\theta, x_{it}, \nu_{it}) f(\nu_{it}) d\nu_{it}$.

moments (that our estimator seeks to bring as close to zero as possible in the sample):

$$g_i^{(1)}(\theta) = \begin{bmatrix} \sum_{t=1}^{T} \sum_{c \in \mathcal{C}_{it}} \sum_{j \in c} Z_{itcj1}^{Q}(Q_{itcj1}^{*} - Q_{cj1}(\theta, x_{it})) \\ \dots \\ \sum_{t=1}^{T} \sum_{c \in \mathcal{C}_{it}} \sum_{j \in c} Z_{itcjK}^{Q}(Q_{itcjK}^{*} - Q_{cjK}(\theta, x_{it})) \\ \sum_{t=1}^{T} \sum_{c \in \mathcal{C}_{it}} \sum_{j \in c} Z_{itcj1}^{D}(D_{itcj1}^{*} - D_{cj1}(\theta, x_{it})) \\ \dots \\ \sum_{t=1}^{T} \sum_{c \in \mathcal{C}_{it}} \sum_{j \in c} Z_{itcjK}^{D}(D_{itcjK}^{*} - D_{cjK}(\theta, x_{it})) \\ \sum_{t=1}^{T} \sum_{c \in \mathcal{C}_{it}} Z_{itc}^{I}(I_{itc}^{*} - I_{c}(\theta, x_{it})) \end{bmatrix}. \tag{25}$$

Here (and elsewhere, unless stated otherwise) the subscript t refers to the tth time period (week) in the sample for consumer i. (As noted in Section 2 we draw three weeks at quarterly intervals for each consumer (so that T=3) where the weeks differ across consumers.) The vectors Z_{itcjk}^{Q}, Z_{itcjk}^{D} and Z_{itc}^{I} are functions of the explanatory variables x_{it}. We now discuss their components (which are also listed in Appendix F) and explain informally how they help with the estimation of the parameters. Parameters can be grouped according to whether they are in μ, α, Λ, or γ. We discuss them in this order. We leave discussion of the spread parameters for unobserved heterogeneity until the next two subsections.

Consider first the parameters in μ. The vectors Z_{itcjk}^{Q} and Z_{itcjk}^{D} contain log of store size, household size, eight firm dummies and a constant. The moments involving these variables are particularly useful for estimating the parameters β_1, β_2 and the firm-category fixed effects ξ_{fk}. Since these moments are category specific, they also contribute to the identification of the category-specific scaling terms β_{0k} (normalized to one for $k = K$). Z_{itcjk}^{Q} also includes quarter and year time dummies, which help with the estimation of β_T.

Next consider the parameters in α and Λ, all of which relate to price effects. To estimate these parameters we include prices in all three moments: Z_{itcjk}^{Q} and Z_{itcjk}^{D} contain the price of category k at store j and the price for each of the other categories k' with which k has an interaction term $\Lambda_{kk'}$.[33] In Z_{itc}^{I} we include the average of the store-category prices in c, i.e. $\Sigma_{j \in c} \Sigma_k p_{jk}/(n(c)K)$, and the same average divided by the per-capita household income of consumer i. The same-category (diagonal) terms Λ_{kk} and the price parameters α_1, α_2 all affect the overall price response, and to estimate them we exploit the fact that we observe price effects both at the quantity and store-choice levels: Λ_{kk} is estimated by observing how much less consumers buy of k when its price changes, whereas α_1 and α_2 are estimated by observing how much less likely consumers are to visit a store when its prices change.[34]

Now consider the cross-category terms $\Lambda_{kk'}$. Here the price of k' is useful. Gentzkow (2007) discusses the challenge of separately identifying correlation in tastes and cross-category effects

[33] When $n(c) = 1$ we use the price of k' in store j and when $n(c) = 2$ we use the average price of k' in the two stores in c.

[34] We see from (7)—i.e. $q_k = (\mu_k - \alpha p_k)/\Lambda_{kk}$ (supposing $q_k > 0$ and $\Lambda_{kk'} = 0$)—that $1/\Lambda_{kk}$ scales category-k demand. Since α is common to all categories, Λ_{kk} determines a k-specific price effect for given α_1 and α_2. Still, α_1 and α_2 determine the scale of variable utility $(\mu - \alpha p)q - 0.5q'\Lambda q$, relative to shopping costs, Γ (see the discussion in footnote 23), and are therefore identified by the changes in shopping choice associated with variation in the price levels across alternatives c.

(complementarity or substitutability). In our setting cross-category effects in variable utility are driven by the cross-category terms $\Lambda_{kk'}$.[35] By contrast, correlation in the taste for categories is determined by the variance of the overall spending shocks $\sigma_1\nu_i^\mu + \sigma_2\nu_{it}^\mu$. Observing that consumers tend to demand either a lot of both k and k' or little of both is consistent with complementarity as well as correlation in tastes. A shock to the price of k' is a natural experiment that allows us to distinguish between these two possibilities: if correlation between demand for k and demand for k' is due entirely to correlation in unobservable tastes (large σ_1, σ_2), the price shock to k' should have no effect on the demand for k (since k' is in this case excluded from the utility of k). On the other hand, if there is complementarity ($\Lambda_{kk'} < 0$), a positive price shock to k' will reduce demand for both k' and k.

Finally, with a view to estimating the mean shopping cost parameters $(\gamma_{11}, \gamma_{21})$, Z_{itc}^I contains the observable variables in the shopping cost term (11): a dummy for whether c is a two-store shopping choice (this is also included in Z_{itcjk}^Q and Z_{itcjk}^D), and the total distance of traveling to each store and back. To further help with identification of shopping costs we also include distance squared and the interaction of distance and the two-store dummy.

4.2 Cross-period moment conditions

We use the panel aspect of the data to estimate the spread parameters on taste shocks that are constant across time. Ackerberg et al. (2007) discuss how first- and second-choice data (like in Berry et al. (2004)) or repeated purchases by the same consumer (like in our data), help to pin down the effects of unobserved and observed sources of taste heterogeneity, since both drive correlation between choices, and cross-sectional data alone is sufficient to estimate the latter.[36] Each of the taste shocks in our model influences one observed outcome in particular, such as total spending, store used or distance traveled. By matching the observed covariance of each such outcome across time periods, in the same way that Berry et al. (2004, p. 74-5) "matched ... the covariances between the first-choice product characteristics and the second-choice product characteristics", we can pin down the parameter for the relevant shock.

Concretely, after taking the conditional expectation given (x_{it}, ν_{it}), overall spending, category-specific quantities, usage incidence of a given store j for category k, one-stop shopping, and total

[35] The cost of visiting multiple stores induces cross-category correlation in demands within a store, but this effect is separately identified since distance and $n(c)$ are observed and excluded from category utility.

[36] "... the correlation between the x-intensity of the first choice and the second choice of a given individual is a function of both θ^o [parameter on unobserved heterogeneity] and the θ^u [parameter on observed heterogeneity] terms, and the θ^o terms should be able to be estimated from only the first choice data. A similar comment can be made for repeated choices, at least provided the utility function of the consuming unit does not change from choice to choice." (Ackerberg et al. (2007), p. 4193.)

distance traveled are:

$$r(\theta, x_{it}, \nu_{it}) = \Sigma_{c\in\mathcal{C}_{it}}\Sigma_{j\in c}\Sigma_{k=1}^{K} p_{tjk} q_{cjk}(\theta^w, x_{itc}^w, \nu_{itc}^w) P_c(\theta, x_{it}, \nu_{it}) \tag{26}$$

$$q_k(\theta, x_{it}, \nu_{it}) = \Sigma_{c\in\mathcal{C}_{it}}\Sigma_{j\in c} q_{cjk}(\theta^w, x_{itc}^w, \nu_{itc}^w) P_c(\theta, x_{it}, \nu_{it}) \tag{27}$$

$$d_{jk}(\theta, x_{it}, \nu_{it}) = \Sigma_{c\in\mathcal{C}_{itj}} 1[q_{cjk}(\theta^w, x_{itc}^w, \nu_{itc}^w) > 0] P_c(\theta, x_{it}, \nu_{it}) \tag{28}$$

$$os(\theta, x_{it}, \nu_{it}) = \Sigma_{c\in\mathcal{C}_{it}} 1[n(c) = 1] P_c(\theta, x_{it}, \nu_{it}) \tag{29}$$

$$dist(\theta, x_{it}, \nu_{it}) = \Sigma_{c\in\mathcal{C}_{it}} dist_{ic} P_c(\theta, x_{it}, \nu_{it}), \tag{30}$$

where $\mathcal{C}_{itj}$ in (28) is the set of shopping choices that include store j for consumer i in period t, and p_{tjk} is the price of category k at store j and time t. The conditional expectation, given $(x_{it}, x_{i(t-1)})$, of the product of the respective quantities in adjacent time periods is

$$R(\theta, x_{it}, x_{i(t-1)}) = \int r(\theta, x_{it}, \nu_{it}) r(\theta, x_{i(t-1)}, \nu_{i(t-1)})\, f(\nu_{it}, \nu_{i(t-1)}) d(\nu_{it}, \nu_{i(t-1)}), \tag{31}$$

and similarly for (27)-(30). For simplicity we now let f denote the joint density of $(\nu_{it}, \nu_{i(t-1)})$. Like above, we use a star, e.g. R_{it}^*, to indicate a given variable before taking expectations (with respect to any unobservable). By the logic set out above, the equivalent of (17) holds for R_{it}^* etc. This implies the following unconditional moment conditions:[37]

$$\mathrm{E}\Big[R_{it}^* - R(\theta_0, x_{it}, x_{i(t-1)})\Big] = 0 \tag{32}$$

$$\mathrm{E}\Big[Q_{itk}^* - Q_k(\theta_0, x_{it}, x_{i(t-1)})\Big] = 0 \tag{33}$$

$$\mathrm{E}\Big[D_{itjk}^* - D_{jk}(\theta_0, x_{it}, x_{i(t-1)})\Big] = 0 \tag{34}$$

$$\mathrm{E}\Big[OS_{it}^* - OS(\theta_0, x_{it}, x_{i(t-1)})\Big] = 0 \tag{35}$$

$$\mathrm{E}\Big[DIST_{it}^* - DIST(\theta_0, x_{it}, x_{i(t-1)})\Big] = 0 \tag{36}$$

Consider first condition (32), which concerns the product of overall spending in adjacent time periods. This condition helps to identify the parameter of the shock $\sigma_1 \nu_i^{\mu}$ that is consumer-specific but fixed across time periods t, categories k and stores j. A large value of σ_1 implies that some consumers tend to spend more in all time periods, whereas others tend to spend less in all time periods. Matching this covariance pins down the value of σ_1.

In the same way, we estimate the spread parameters on the category-specific (σ_3) and store-category-specific (σ_4) random shocks by matching the covariance between time periods of (non-store-specific) category quantities and of store-category purchase indicators, respectively, as required by (33) and (34). Concretely, the correlation in quantity purchased of k across periods reveals σ_3 and the extent to which consumers tend to use the same store for k across time periods reveals σ_4. The spread parameters on unobserved heterogeneity in shopping cost (γ_{12}) and distance traveled (γ_{22}) are identified by matching the observed and predicted covariances (between adjacent time periods) of one-stop shopping and distance traveled, respectively.

[37] To see this: $\mathrm{E}[R_{it}^* - R(\theta_0, x_{it}, x_{i(t-1)})] = \mathrm{E}\{\mathrm{E}[R_{it}^* - R(\theta_0, x_{it}, x_{i(t-1)}) | x_{it}, x_{i(t-1)}]\} = \mathrm{E}\{0\} = 0$.

The empirical moments are

$$g_i^{(2)}(\theta) = \begin{bmatrix} \sum_{t=2}^T (R_{it}^* - R(\theta, x_{it}, x_{i(t-1)})) \\ \sum_{t=2}^T \sum_{k=1}^K (Q_{itk}^* - Q_k(\theta, x_{it}, x_{i(t-1)})) \\ \sum_{t=2}^T \sum_{j \in \mathcal{J}_{it,(t-1)}} \sum_{k=1}^K (D_{itjk}^* - D_{jk}(\theta, x_{it}, x_{i(t-1)})) \\ \sum_{t=2}^T (OS_{it}^* - OS(\theta, x_{it}, x_{i(t-1)})) \\ \sum_{t=2}^T (DIST_{it}^* - DIST(\theta, x_{it}, x_{i(t-1)})) \end{bmatrix} \tag{37}$$

where $\mathcal{J}_{it,(t-1)} = \mathcal{J}_{it} \cap \mathcal{J}_{i(t-1)}$ is the set of stores that are in i's choice set in both period t and period $t-1$.

4.3 Cross-category moment conditions

We use cross-category moment conditions to distinguish between the time-invariant shock (with spread parameter σ_1) and the time-varying shock (with spread parameter σ_2). Both types of shock impact all categories. However, the time-varying shock generates correlation between spending on k and k' in the same period, but—contrary to the time-invariant shock—not between k in one period and k' in another period (see Gentzkow (2007) for a related argument). We therefore use separate moment conditions for within-period and cross-period covariances in spending across categories. Spending on category k, after taking the conditional expectation given (x_{it}, ν_{it}), is

$$r_k(\theta, x_{it}, \nu_{it}) = \Sigma_{c \in \mathcal{C}_{it}} \Sigma_{j \in c} p_{tjk} q_{cjk}(\theta^w, x_{itc}^w, \nu_{itc}^w) P_c(\theta, x_{it}, \nu_{it}). \tag{38}$$

We use this to define the following conditional expectations given $(x_{it}, x_{i(t-1)})$ for the average product between spending on k and k' within ("in") and across ("cr") adjacent time periods,

$$\begin{aligned} R_{kk'}^{in}(\theta, x_{it}, x_{i(t-1)}) &= \int \frac{1}{2}\Big[r_k(\theta, x_{i(t-1)}, \nu_{i(t-1)}) r_{k'}(\theta, x_{i(t-1)}, \nu_{i(t-1)}) \\ &+ r_k(\theta, x_{it}, \nu_{it}) r_{k'}(\theta, x_{it}, \nu_{it})\Big] f(\nu_{it}, \nu_{i(t-1)}) d(\nu_{it}, \nu_{i(t-1)}) \qquad (39) \\ R_{kk'}^{cr}(\theta, x_{it}, x_{i(t-1)}) &= \int \frac{1}{2}\Big[r_k(\theta, x_{i(t-1)}, \nu_{i(t-1)}) r_{k'}(\theta, x_{it}, \nu_{it}) \\ &+ r_k(\theta, x_{it}, \nu_{it}) r_{k'}(\theta, x_{i(t-1)}, \nu_{i(t-1)})\Big] f(\nu_{it}, \nu_{i(t-1)}) d(\nu_{it}, \nu_{i(t-1)}). \qquad (40) \end{aligned}$$

The unconditional versions of (17) are,

$$\mathrm{E}[R_{itkk'}^{in*} - R_{kk'}^{in}(\theta_0, x_{it}, x_{i(t-1)})] = 0, \quad \text{for } k < k' \tag{41}$$

$$\mathrm{E}[R_{itkk'}^{cr*} - R_{kk'}^{cr}(\theta_0, x_{it}, x_{i(t-1)})] = 0, \quad \text{for } k < k', \tag{42}$$

where, in the same way as above, the star indicates the respective quantities ((39) and (40)) before taking expectations (with respect to any unobservables). We use (41) and (42) to form the following empirical moments:

$$g_i^{(3)}(\theta) = \left[\begin{array}{c} \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{k'=k+1}^{K} (R_{itkk'}^{in*} - R_{kk'}^{in}(\theta, x_{it}, x_{i(t-1)})) \\ \sum_{t=2}^{T} \sum_{k=1}^{K} \sum_{k'=k+1}^{K} (R_{itkk'}^{cr*} - R_{kk'}^{cr}(\theta, x_{it}, x_{i(t-1)})) \end{array} \right]. \tag{43}$$

4.4 Estimation

To estimate the parameters we write $g(\theta) = N^{-1} \sum_{i=1}^{N} g_i(\theta)$, where $g_i(\theta)$ vertically stacks the three sets of moments $g_i^{(1)}(\theta)$, $g_i^{(2)}(\theta)$, and $g_i^{(3)}(\theta)$. The GMM estimator is

$$\hat{\theta} \quad = \quad \arg\min_{\theta} g(\theta)' W^{-1} g(\theta) \tag{44}$$

where the weighting matrix is the inverse of the covariance matrix $W = N^{-1} \sum_{i=1}^{N} g_i(\tilde{\theta}) g_i(\tilde{\theta})'$ and $\tilde{\theta}$ are first-stage estimates.[38]

5 Estimates and Model Fit

In this section we discuss parameter estimates and model fit. We present estimates for two specifications in Table 3. As a starting point Model 1 assumes independence between product categories in variable utility, so that the cross-category terms $\Lambda_{kk'}$ are set to zero. Model 2 relaxes this assumption.

Panel A shows the parameters that enter the store-category taste effects μ_{itjk} as specified in equation (9). The effect of household and store size have intuitive signs: β_1 and β_2 are both positive. The spread parameters ($\sigma_1, \ldots, \sigma_4$) are precisely estimated. Panel B reports the parameters in the matrix Λ of second-order terms in quadratic utility. The diagonal parameters Λ_{kk} are all positive so that own-price effects are negative. The off-diagonal parameters $\Lambda_{kk'}$ (in Model 2) are also positive, which implies that categories are intrinsic substitutes (i.e. substitutes in terms of the variable utility function), but the estimated parameters are small and insignificant for three of the five parameters. In the interest of parsimony we estimate only the interaction parameters that we believed a priori to be the most important. (In Section 6.6 we discuss results from a model with alternative interactions and we find that our main results are robust to this alternative). The parameters in the price sensitivity coefficient (10), reported in Panel C, are of the expected sign: α_1 and α_2 are positive so that consumers prefer lower prices and price sensitivity is decreasing in per capita household income. Finally, Panel D reports the parameters γ that enter the consumer's shopping costs $\Gamma(c)$. The mean and spread parameters for both shopping cost variables are precisely estimated.

[38] We first obtain preliminary estimates by using only the moments $g_i^{(1)}$, with the inverse of the covariance of the Z vectors as the weighting matrix. Then $(N^{-1} \sum_{i=1}^{N} g_i g_i')^{-1}$ evaluated at these preliminary estimates is the weighting matrix used to obtain the first-stage estimates $\tilde{\theta}$. We use a simple frequency simulator with one draw per observation (consumer) and a standard estimator for the asymptotic variance of $\hat{\theta}$ (see Wooldridge (2001), p. 527, eq. 14.14). To correct for simulation noise we multiply this variance by a factor of $1 + \frac{1}{r} = 2$, where $r = 1$ is the number of simulation draws per observation (see McFadden (1989), p. 1006).

Table 3: Estimated Parameters

		Model 1		Model 2	
		Estimate	Std. Error	Estimate	Std. Error
A: Store-category Taste Effects					
Bakery	β_{01}	2.374	0.074	2.085	0.037
Dairy	β_{02}	1.643	0.032	1.418	0.106
Drink	β_{03}	0.943	0.016	1.096	0.013
Dry Grocery	β_{04}	2.063	0.067	1.901	0.105
Fruit & vegetable	β_{05}	2.968	0.100	2.665	0.192
Household goods	β_{06}	1.252	0.022	1.115	0.028
Meat	β_{07}	2.733	0.128	2.309	0.135
ln(floor space)	β_1	0.425	0.005	0.456	0.006
Household size	β_2	0.383	0.008	0.477	0.024
Year & Quarter effects		Yes		Yes	
Scale of Taste Shocks (ν):					
Fixed across category/store	σ_1	0.197	0.024	0.208	0.012
Time-varying	σ_2	0.790	0.019	0.920	0.046
Category specific	σ_3	0.748	0.017	0.677	0.023
Store/category specific	σ_4	1.033	0.019	1.129	0.015
Firm-Category effects	ξ_{fk}	Yes		Yes	
B: Second-Order Quadratic Parameters	$\Lambda_{kk'}$				
Bakery	Λ_{11}	22.209	1.127	19.852	2.090
Dairy	Λ_{22}	12.308	0.383	11.239	1.404
Drink	Λ_{33}	3.261	0.173	3.802	0.192
Dry Grocery	Λ_{44}	11.332	0.582	10.536	0.916
Fruit & vegetable	Λ_{55}	16.113	0.739	15.952	2.138
Household goods	Λ_{66}	4.387	0.163	4.360	0.219
Meat	Λ_{77}	9.882	0.613	8.901	0.883
Milk	Λ_{88}	14.081	0.330	14.062	1.257
Drink - Dry Grocery	Λ_{57}	–	–	1.742	0.163
Milk - Dairy	Λ_{18}	–	–	1.368	0.555
Bakery - Fruit & Veg	Λ_{23}	–	–	0.269	0.890
Bakery-Meat	Λ_{23}	–	–	0.572	0.363
Fruit & Veg - Meat	Λ_{23}	–	–	0.076	0.926
C: Price Parameters					
Constant	α_1	1.936	0.047	1.839	0.037
1/[Weekly Income (£) per Head]	α_2	18.345	3.046	32.881	3.702
D: Shopping Costs					
Two Store Dummy	γ_{11}	9.665	1.385	7.528	0.917
Standard Deviation	γ_{12}	14.375	2.553	10.269	1.657
Distance	γ_{21}	0.430	0.026	0.440	0.027
Standard Deviation	γ_{22}	0.396	0.030	0.394	0.028

Notes: Parameters are estimated by GMM using 6000 consumer-week observations. Standard errors are corrected for simulation noise as detailed in Section 4. Year, quarter, and firm-category fixed effects are not reported.

Table 4: In-Sample and Out-of-Sample Fit

	In-Sample		Out-of-Sample	
	Quantity	Shoppers	Quantity	Shoppers
A: Correlation between predicted & observed demands				
$\rho(Q_{fk}, Q^*_{fk})$,$\rho(D_{fk}, D^*_{fk})$	0.994	0.994	0.986	0.991
B: Mean absolute prediction errors				
B1: Firm share of category demand (all firms and categories)				
$\lvert s_{fk} - s^*_{fk}\rvert$	0.008	0.008	0.011	0.009
B2: 1-stop shopper share of category demand (all categories)				
$\lvert s_{1ss,k} - s^*_{1ss,k}\rvert$	0.004	0.020	0.007	0.028

C: Firm-Level Demand Market shares of Firms

	In-Sample				Out-of-Sample			
	Revenues		Shoppers		Revenues		Shoppers	
	Pred	Obs	Pred	Obs	Pred	Obs	Pred	Obs
ASDA	0.209	0.214	0.180	0.170	0.204	0.230	0.177	0.176
Morrisons	0.128	0.113	0.115	0.096	0.134	0.111	0.117	0.092
Sainsbury	0.159	0.153	0.127	0.127	0.157	0.152	0.128	0.123
Tesco	0.327	0.329	0.266	0.256	0.315	0.317	0.256	0.255
M&S	0.005	0.011	0.009	0.026	0.007	0.012	0.010	0.029
Waitrose	0.021	0.016	0.019	0.020	0.017	0.018	0.017	0.021
Iceland	0.008	0.015	0.026	0.036	0.012	0.017	0.030	0.035
Discounter	0.025	0.029	0.052	0.055	0.031	0.028	0.056	0.053
Other	0.118	0.120	0.207	0.214	0.123	0.115	0.210	0.216

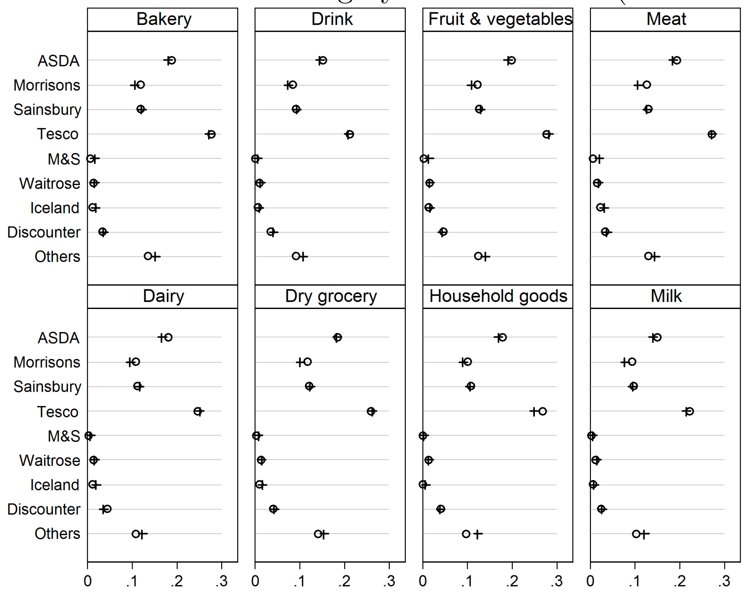
Notes: In-sample predictions use the estimation sample of 6000 consumer-weeks (and the taste draws) used in estimation. Out-of-sample statistics use a new sample of 6000 consumer-weeks and new taste draws. Panel A: Correlation coefficients are for number of firm-category shoppers D_{fk} and quantities Q_{fk}. Correlation statistics are for the 72 firm-category predictions. Panel B: Mean absolute prediction errors are given by $|s - s^*|$ where s is predicted and s^* is observed market shares as defined in text. B1 uses 72 firm-category market shares while B2 uses 8 category shares (the proportion of 1-stop shoppers in category demand). Panel C: All columns sum to 1.

Model 1 is a restricted version of Model 2, in which the five off-diagonal elements of Λ are all set to zero. We reject this restriction at a significance level of less than 1%. (The χ^2_5-distributed GMM distance statistic comparing Model 1 and Model 2 is 19.2). We use Model 2 as the baseline model for the results in Section 6. In the rest of this section we discuss the fit of this model.

The model generates choice outcomes at three levels: continuous and discrete demands at store-category level and a discrete choice of store(s) at the shopping choice level. In the rest of this section we check the fit of the model to ensure it is flexible enough to match these choice predictions accurately. For example firm-specific taste parameters appear only in variable utility (entering as firm-category effects ξ_{fk}) and they serve the joint purpose of fitting (continuous and discrete) category demands for each firm, and discrete shopping choices that include stores of each firm, so it is informative to check whether the model fits the data at these different levels. Along with the in-sample fit on the estimation sample of 6000 consumer-weeks, we consider the out-of-sample fit on a validation sample of 6000 consumer-weeks (with new random taste draws).

We begin at the category level. To check the fit of both demand intensity and demand incidence we consider a continuous and a discrete measure of category demand. The continuous prediction

Figure 1: Predicted and Observed Category Market Shares (in terms of Shoppers)



Notes: This figure shows observed (+) and predicted (o) market shares by category in terms of shoppers using the 6000 consumer-weeks in the estimation sample (and their taste draws).

Table 5: Observed and Predicted Market Shares by Firm Pair

		ASDA	Morr	Sains	Tesco	M&S	Wait	Icel	Disc	Other
ASDA	Pred	0.144	0.011	0.011	0.024	0.002	0.001	0.004	0.008	0.025
	Obs	0.131	0.012	0.012	0.022	0.008	0.001	0.007	0.011	0.029
Morrisons	Pred		0.087	0.008	0.017	0.001	0.001	0.002	0.005	0.016
	Obs		0.073	0.004	0.012	0.001	0.001	0.002	0.009	0.018
Sainsbury	Pred			0.098	0.023	0.001	0.002	0.001	0.004	0.014
	Obs			0.088	0.027	0.008	0.004	0.005	0.005	0.021
Tesco	Pred				0.209	0.002	0.005	0.006	0.010	0.045
	Obs				0.206	0.008	0.006	0.012	0.013	0.045
M&S	Pred					0.004	0.000	0.000	0.000	0.001
	Obs					0.004	0.002	0.001	0.001	0.004
Waitrose	Pred						0.011	0.000	0.000	0.003
	Obs						0.009	0.001	0.001	0.003
Iceland	Pred							0.013	0.001	0.004
	Obs							0.008	0.004	0.011
Discounters	Pred								0.029	0.008
	Obs								0.020	0.014
Other	Pred									0.149
	Obs									0.148

Notes: Predicted and observed market shares (in terms of shoppers) for each firm pair. The diagonal shows the proportion of consumers using the indicated firm only. Predictions and observations are for the 6000 consumer-weeks (and taste draws) used in estimation.

at the firm category level, given estimated parameters $\hat{\theta}$, is the total quantity of category k sold by the firm, i.e.

$$Q_{fk}(p) = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\Sigma_{j\in\mathcal{J}_f}\Sigma_{c\in\mathcal{C}_{itj}}Q_{cjk}(\hat{\theta}, x_{it}) \tag{45}$$

where $Q_{cjk}(\hat{\theta}, x_{it})$ is i's category demand at store j in shopping choice c defined in (19). The innermost sum is over the set $\mathcal{C}_{itj}$ of shopping choices c that include store j. $\mathcal{J}_f$ is the set of stores owned by firm f. The discrete demand measure, which we refer to as the number of shoppers, is the total number of consumers who buy a positive quantity of category k from firm f, i.e.

$$D_{fk}(p) = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\Sigma_{j\in\mathcal{J}_f}\Sigma_{c\in\mathcal{C}_{itj}}D_{cjk}(\hat{\theta}, x_{it}) \tag{46}$$

where $D_{cjk}(\hat{\theta}, x_{it})$ is the probability consumer i buys a positive quantity of k in store j in shopping choice c, defined in (20).

Panel A of Table 4 presents correlation coefficients which show that the predicted and observed firm-category demands from the model are highly correlated. This is also true for the out-of-sample predictions. A high correlation is found both for the discrete and the continuous demand measures.

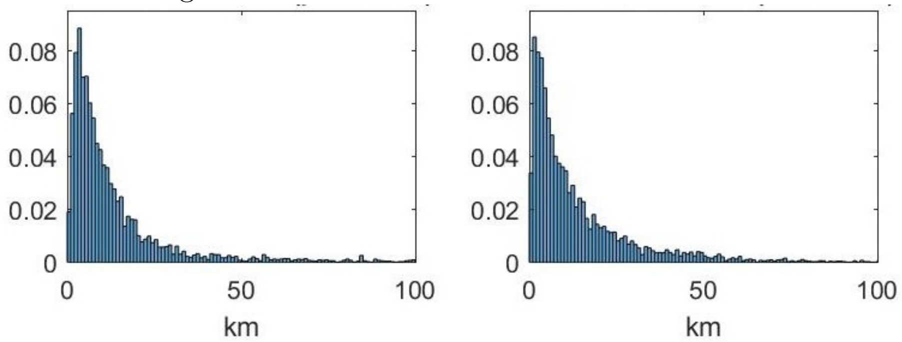
Panel B of Table 4 checks how close the observed and predicted demands are to each other. In B1 we consider how well the model predicts each firm f's share of category k, written s_{fk}, in terms of quantities $(Q_{fk}(p)/\Sigma_{f'}Q_{f'k}(p))$ and shoppers $(D_{fk}(p)/\Sigma_{f'}D_{f'k}(p))$. The absolute prediction error is the magnitude of the difference between the predicted and observed shares. There are 72 such prediction errors - one for each firm and category combination. We find that the average of these prediction errors is 0.008 for in-sample predictions (for both quantities and shoppers)—i.e. on average a firm's market share is predicted to be within about one percentage point of its observed value. A similarly small error is found for out-of-sample predictions. Figure 1 visualizes the predicted and observed market shares s_{fk} of each category for each firm in terms of number of shoppers (for the estimation sample).

The consumer's category demand is associated with a shopping choice c which may have either one or two stores. Substitution patterns between categories depend on this dimension of the model, as we saw in section 3.1. We therefore check the proportion of demand in each category from one-stop shopping choices, written s_k^{1ss}, in terms of quantities $(\Sigma_f Q_{fk}^{1ss}/\Sigma_f Q_{fk})$ and shoppers $(\Sigma_f D_{fk}^{1ss}/\Sigma_f D_{fk})$. There are 8 such prediction errors: one for each category. Panel B2 of Table 4 shows that the average of these prediction errors is 0.004 in terms of quantities and 0.020 in terms of shoppers for in-sample predictions. A similar level of fit is found for the out-of-sample predictions.

Panel C of Table 4 moves from the category level to the firm level and considers the aggregate (across categories) market shares for the main firms in the market. This allows us to check that the model predicts the market share accurately for each firm both in-sample and out-of-sample.

We now check how well the model predicts some other aspects of the shopping choices c. Table 5 looks at the in-sample fit for the market share of each possible combination of firms. Note that there is no direct parameter to capture this (like a "firm pair" dummy), so it is interesting to see whether the model fits this aspect of the data well. The diagonal gives the proportion of shoppers

Figure 2: Predicted and Observed Distances



Notes: The graphs display histograms of predicted (left) and observed (right) total shopping distances traveled (to and from stores) in kilometers for the 6000 consumer-weeks (and taste draws) used in estimation. The height of each bar is the relative number of observations (number of observations in bin / total number of observations). The sum of the bar heights is 1.

that shop only at one firm and the upper triangle gives the proportion that combine each pair of firms. Finally we provide a visual check of the spatial fit of the model. Figure 2 presents histograms of observed and predicted total shopping travel distances (to and from consumers' chosen stores) for the 6000 consumer-weeks in the estimation sample. The histograms indicate that relatively few travel more than 30km to and from their home. As well as being consistent with the observed data, these predictions are consistent with external survey evidence from CC (2000, 4.129), which found 91% of shoppers had a travel time of 20 minutes or less to their supermarket—a distance of about 15km (30km to and from) at standard driving speeds of 45km/hour.

6 Analysis of Supermarket Pricing

We now use the estimated model to analyze supermarket pricing. We first outline the pricing problem under two alternative organizational assumptions: supermarket pricing and independent category sellers. In 6.2 we report the own- and cross-category demand elasticities implied by the estimated parameters. In 6.3 we solve for the Nash equilibrium profit margins implied under the alternative organizational assumptions and compare them to external data on profit margins. We find that supermarket pricing fits the external data better. In the rest of the section we assume supermarket pricing and consider the two main policy-relevant questions of interest: we measure cross-category externalities and discuss their impact on market power, and we compare the impact of one-stop and two-stop shopper types on firms' pricing incentives.

Throughout this section we use the model's predictions of firm-category demands in a given week

$$Q_{fk}(p) = \Sigma_{i=1}^{N} \Sigma_{j \in \mathcal{J}_f} \Sigma_{c \in \mathcal{C}_{itj}} Q_{cjk}(\hat{\theta}, x_{it}) \tag{47}$$

for the N consumers in the estimation sample, where $Q_{cjk}(\hat{\theta}, x_{it})$ is given by (19), and $\mathcal{C}_{itj}$ is the set of shopping choice alternatives that include store j.[39] As we use only one period we suppress the t subscript in the rest of this section.

[39] For all consumers t is week 78 (the midpoint of the 156-week sample period). We use the same taste draws as in estimation. If week 78 is not in the estimation sample for consumer i we draw a new time-specific taste ν_{it}^{μ}.

6.1 Supermarket Pricing and Equilibrium Profit Margins

We compare two forms of organization: supermarkets, which set prices to maximize profit across all categories (internalizing cross-category effects), solving

$$\max_{p_{f1},\dots,p_{fK}} \Sigma_{k=1}^{K} Q_{fk}(p)(p_{fk} - mc_{fk}),$$

and independent category sellers, which set prices to maximize category profit, solving

$$\max_{p_{fk}} Q_{fk}(p)(p_{fk} - mc_{fk})$$

for $k = 1, \dots, K$. Here p_{fk} is the firm's (national) price and mc_{fk} the marginal cost. (The firms have a policy of national rather than storewise pricing, see footnote 13).

We assume Nash equilibrium prices, which implies the following set of first-order conditions for each f and k:

$$\underbrace{\frac{Q_{fk}}{\frac{\partial Q_{fk}}{\partial p_{fk}}} + p_{fk}}_{\text{marginal category revenue } (mr_{fk})} + \underbrace{\chi_f \Sigma_{k' \neq k} \left(\frac{\frac{\partial Q_{fk'}}{\partial p_{fk}}}{\frac{\partial Q_{fk}}{\partial p_{fk}}} (p_{fk'} - mc_{fk'}) \right)}_{\text{marginal externality } (me_{fk})} = mc_{fk} \tag{48}$$

where χ_f is 1 for supermarket pricing and 0 for independent category sellers. This condition states that the marginal benefit of inducing an extra unit of demand for category k (by means of a price change)—i.e. the marginal revenue mr_{fk} plus the marginal externality on other categories me_{fk}—is equal to marginal cost mc_{fk}. Note that the marginal externality imposed on any category $k' \neq k$ is the product of its markup $(p_{fk'} - mc_{fk'})$ and the cross-category diversion ratio

$$\frac{\partial Q_{fk'}}{\partial p_{fk}} \Big/ \frac{\partial Q_{fk}}{\partial p_{fk}} \tag{49}$$

which is the change in category k' demand at firm f for every unit of demand it loses on category k as a result of an increase in p_{fk}. Letting $\pi_{fk} = Q_{fk}(p)(p_{fk} - mc_{fk})$ and dividing (48) by price we obtain the following expression for the Lerner index

$$\frac{p_{fk} - mc_{fk}}{p_{fk}} = \frac{1}{\left| \frac{\partial Q_{fk}}{\partial p_{fk}} \frac{p_{fk}}{Q_{fk}} \right|} - \frac{\chi_f}{p_{fk}} \Sigma_{k' \neq k} \left(\frac{\frac{\partial \pi_{fk'}}{\partial p_{fk}}}{\frac{\partial Q_{fk}}{\partial p_{fk}}} \right). \tag{50}$$

This shows the relationship between market power and the cross-category externality: an independent category seller has a Lerner index that is equal to the inverse of its own-price elasticity, while a supermarket's Lerner index is lower than this by the extent of the marginal externality (as a proportion of p_{fk}).[40]

[40] In Appendix H.1 we derive (50) from product-level first-order conditions. In Appendix H.2 we show how group-specific prices (discussed in Section 2) result in the same expression as in (50) where p_{fk} is a weighted average of the group-specific prices. For simplicity we ignore group specific prices in the notation in this section.

6.2 Estimated Own- and Cross-Category Elasticities

The elasticities implied by the estimates are presented in Table 6 for six categories and three firms. The table consists of nine blocks of 6×6 sub-matrices. The three 6×6 within-firm elasticity matrices, along the principal block diagonal of the overall matrix, give own- and cross-elasticities for the categories of a given firm. Note that all the elasticities in these blocks are negative, so that any pair of categories at the same firm are pricing complements. This in turn implies that the diversion ratio (49) and the cross-category externality in (48) are positive. Some categories (e.g. meat) generate a much larger cross-elasticity than others (e.g. milk), which is likely to be a consequence of their relative size in the consumer's budget.

To decompose the cross-category complementarity effects into the consumer's discrete and continuous responses, Table 7 presents the 3×3 submatrix of overall cross-category within-firm elasticities (for two firms, ASDA and Tesco) alongside the corresponding conditional elasticities that hold (discrete) store-category choices d and shopping choices c constant and only allow continuous quantity choices to change. The cross-elasticities conditional on store choices are positive in sign. This shows that the cross-category complementarity derives from the consumer's shopping costs rather than from any intrinsic complementarity between the categories captured by $\Lambda_{kk'}$.

Returning to Table 6 note that the principal diagonal in each of the within-firm elasticity matrices gives own-price elasticities—i.e. same-firm same-category price elasticities. These are generally larger in magnitude than the cross-category same-firm price elasticities (which are on the off-diagonals). This difference is a consequence of two consumer responses that are allowed in the consumer model (shown in equation (8)): (i) a reduction (at the intensive margin) in the continuous demand for the category holding store choices fixed and (ii) a change of store for the category but not for other categories, which is possible for two-stop shoppers (response class (2b) in subsection 3.1).

Two further features of the own-price elasticities are noteworthy. First, they have less than unit magnitude in some cases. Elasticities of less than unit magnitude are inconsistent with positive marginal costs for a single-category seller (see (50) for the case of $\chi_f = 0$). Elasticities of this magnitude are, however, consistent with positive marginal costs when the firm internalizes a positive externality on other categories (see (50) for $\chi_f = 1$), which results in the firm setting prices at a lower level than otherwise. Second, the own-price elasticities vary across firms in a plausible way: they are higher for the discounter (Aldi) than for the Big Four firms, which may reflect (i) the relatively high share of two-stop shoppers among the discounter's customers (with their greater ease of substituting a category between stores) and (ii) the relatively high price-sensitivity of consumers attracted to discounter firms.

The off-diagonal 6×6 blocks give inter-firm elasticities. These are asymmetric in magnitude because of the differences in firm market shares: the effect of prices at Aldi (which has small market share) on demands at ASDA or Tesco (which have large market shares) is small (e.g. see the top-right 6×6 block) compared to the opposite price elasticities. Note that the pattern of elasticities within these off-diagonal blocks suggests there is a significant number of two-stop consumers that switch firms only for the category affected by the price change (the "middle"

Table 6: Cross Elasticities at Firm-Category Level for Selected Categories and Firms

	ASDA						Tesco						Aldi					
	Bakry	Drink	Fr,vg	H'hld	Meat	Milk	Bakry	Drink	Fr,vg	H'hld	Meat	Milk	Bakry	Drink	Fr,vg	H'hld	Meat	Milk
ASDA																		
Bakery	-0.619	-0.210	-0.275	-0.259	-0.372	-0.046	0.082	0.075	0.119	0.126	0.160	0.019	0.004	0.003	0.009	0.004	0.007	0.001
Drink	-0.135	-1.300	-0.251	-0.256	-0.363	-0.049	0.059	0.117	0.115	0.119	0.147	0.021	0.002	0.008	0.008	0.006	0.006	0.001
Fr,veg	-0.129	-0.183	-0.630	-0.239	-0.358	-0.040	0.058	0.078	0.156	0.113	0.149	0.018	0.003	0.004	0.014	0.006	0.008	0.001
H'hold	-0.132	-0.203	-0.260	-1.125	-0.364	-0.046	0.061	0.084	0.115	0.179	0.152	0.021	0.002	0.004	0.007	0.010	0.008	0.001
Meat	-0.119	-0.181	-0.245	-0.229	-0.830	-0.044	0.057	0.070	0.111	0.118	0.211	0.020	0.002	0.003	0.006	0.005	0.012	0.001
Milk	-0.121	-0.200	-0.224	-0.237	-0.363	-1.163	0.051	0.082	0.109	0.124	0.151	0.032	0.002	0.003	0.005	0.004	0.010	0.002
Tesco																		
Bakery	0.047	0.052	0.071	0.069	0.101	0.011	-0.604	-0.161	-0.258	-0.273	-0.326	-0.045	0.004	0.003	0.008	0.004	0.005	0.001
Drink	0.031	0.075	0.069	0.068	0.089	0.013	-0.115	-1.293	-0.226	-0.240	-0.277	-0.043	0.001	0.007	0.005	0.003	0.005	0.001
Fr,veg	0.034	0.052	0.097	0.065	0.101	0.012	-0.131	-0.160	-0.679	-0.247	-0.319	-0.043	0.002	0.004	0.010	0.005	0.004	0.001
H'hold	0.035	0.052	0.067	0.098	0.103	0.013	-0.133	-0.163	-0.238	-1.040	-0.311	-0.046	0.002	0.003	0.008	0.006	0.007	0.001
Meat	0.035	0.049	0.069	0.064	0.142	0.013	-0.123	-0.146	-0.237	-0.240	-0.809	-0.042	0.002	0.002	0.005	0.003	0.011	0.001
Milk	0.030	0.050	0.059	0.066	0.096	0.019	-0.124	-0.163	-0.230	-0.258	-0.306	-1.069	0.001	0.002	0.005	0.004	0.004	0.001
Aldi																		
Bakery	0.094	0.085	0.132	0.095	0.154	0.018	0.143	0.055	0.143	0.182	0.163	0.019	-0.907	-0.232	-0.537	-0.319	-0.379	-0.065
Drink	0.034	0.160	0.118	0.105	0.130	0.018	0.071	0.221	0.193	0.154	0.120	0.017	-0.139	-1.928	-0.499	-0.347	-0.451	-0.070
Fr,veg	0.053	0.075	0.180	0.081	0.105	0.011	0.087	0.069	0.204	0.163	0.132	0.020	-0.142	-0.220	-1.104	-0.264	-0.331	-0.059
H'hold	0.042	0.093	0.126	0.204	0.149	0.018	0.078	0.088	0.172	0.243	0.142	0.027	-0.150	-0.273	-0.469	-1.647	-0.416	-0.078
Meat	0.052	0.073	0.122	0.112	0.267	0.027	0.070	0.093	0.111	0.183	0.362	0.018	-0.125	-0.249	-0.413	-0.292	-1.501	-0.068
Milk	0.031	0.104	0.144	0.130	0.151	0.044	0.077	0.087	0.161	0.162	0.160	0.037	-0.155	-0.277	-0.535	-0.393	-0.491	-1.768

Notes: Each cell is elasticity of row demand with respect to column price.

Table 7: Within-Firm Cross-Category Elasticity Decomposition

	ASDA: Conditional			ASDA: Unconditional			Tesco: Conditional			Tesco: Unconditional		
	Bakery	Fr,veg	Meat	Bakery	Fr,veg	Meat	Bakery	Fr,veg	Meat	Bakery	Fr,veg	Meat
Bk	-0.373	0.005	0.020	-0.674	-0.282	-0.379	-0.367	0.006	0.019	-0.622	-0.268	-0.342
Fv	0.003	-0.232	0.002	-0.136	-0.646	-0.370	0.003	-0.274	0.002	-0.137	-0.700	-0.337
Mt	0.007	0.001	-0.261	-0.124	-0.173	-0.836	0.007	0.001	-0.271	-0.130	-0.250	-0.832

Notes: All elasticities are within-firm. Each cell is elasticity of row demand with respect to column price. Conditional elasticities hold discrete choices (d, c) constant and allow only continuous choices to change. Unconditional elasticities allow the consumer to change shopping choice c and the store $j \in c$ used for each category.

response in the decomposition in equation (8))—e.g. in the top-middle block a change in the price of Tesco Meat has a higher proportional effect on ASDA Meat (0.211) than ASDA Drink (0.147) because of two-stop shoppers that switch stores for Meat only.

6.3 Predicted and Observed Profit Margins

In this section we compare the margins (i.e. the Lerner index) implied by the model with bounds to profit margins calculated using external accounting data from competition inquiries CC (2000, 2008). This allows us to check whether the assumption of supermarket pricing is validated by external data. We do this comparison at an overall level (across all categories) as well as for a specific category (milk) for which there is relatively accurate margin data.

The margins reported by the CC are based on firms' accounting data and cannot be unambiguously mapped to our theoretical margin concept for two reasons. (The issues here are common when dealing with accounting data on costs and revenues. For instance Nevo (2001) discusses similar challenges when validating estimated margins with external data.) First, the CC only reports total revenues and total wholesale costs at the retailer level. Hence we do not observe the marginal wholesale price and need to make an assumption about the vertical contract that led to the reported payments to manufacturers.[41] Second, it is ambiguous what fraction of labor costs should be considered marginal. We therefore provide bounds based on alternative assumptions about vertical contracts and the fraction of labor costs that are marginal. Further details on how we derive the bounds on margins are in Appendix I.

To obtain the profit margins implied by the model we solve the system of first-order conditions (48) at estimated demand parameters under two alternative assumptions: supermarket pricing ($\chi_f = 1$ in (48)) and independent category pricing ($\chi_f = 0$). This gives a marginal cost mc_{fk} for each of the combinations of f and k. Based on these marginal costs a simple preliminary check on the two organizational assumptions is to ask whether either of them implies that marginal costs are

[41] We do not explicitly model the interaction between manufacturers and retailers (as for example in Sudhir (2001) Besanko et al. (2003)). The external bounds that we calculate adopt an agnostic position as to whether there is double marginalization or efficient pricing. There is relatively little direct evidence that discriminates between these two alternative models of vertical contracting. Villas-Boas (2007) and Bonnet and Dubois (2010) use a structural model of demand to compare the implications of the two modeling assumptions and reject double marginalization in favor of efficient pricing (where the retailer optimizes against true vertical marginal costs).

Table 8: Predicted Profit Margins and Observed Bounds to Margins

	Milk	All Categories
A: Bounds to margins from external data		
A1 [Lower bound]: Retail margin; labor included in marginal cost	0.20	0.16
A2 [Upper bound]: Full vertical margin; labor excl. from marginal cost	0.34	0.52
B: Median predicted margins (95% confidence intervals)		
B1: Supermarket Pricing, $\chi_f = 1$ in equation (50)	(0.22, 0.28)	(0.28, 0.31)
B2: Independent Category Sellers, $\chi_f = 0$ in equation (50)	(0.65, 0.80)	(0.71, 0.80)
C: Weighted Median predicted margins (95% confidence intervals)		
C1: Supermarket Pricing, $\chi_f = 1$ in equation (50)	(0.22, 0.33)	(0.37, 0.43)
C2: Independent Category Sellers, $\chi_f = 0$ in equation (50)	(0.68, 1.01)	(0.91, 1.09)

Notes: Panel A gives bounds to profit margins using external data in CC inquiries. See Appendix I for details. Panel B reports predicted median profit margins; medians are across categories and firms (All Categories) across firms (Milk). Weighted medians are weighted by firm-category revenues from the estimation sample to allow for heterogeneity in firm market shares.

negative. We find that all firm-category marginal costs are positive when we assume supermarket organization but 14.8% of them are negative when we assume independent category sellers.

To assess the two organizational assumptions more formally, we present confidence intervals for the profit margins implied when $\chi_f = 1$ and $\chi_f = 0$, respectively, and compare them with the external bounds to profit margins from the CC report. Panels B and C report 95% confidence intervals: the first column for the median margin in the milk category across firms, and the second column for the median margin across all categories and firms. Since each external margin is an industry-wide number, in Panel C we weight the medians by firm-category revenue to reflect variation in firm and category sizes. The weighted median is higher because firms with larger market shares (such as the Big Four) tend to have greater market power. To calculate confidence intervals, we take 2000 draws of the parameter vector from the (estimated asymptotic) distribution of our estimator where for each draw we compute margins using equation (50). The confidence intervals are then given by the 2.5th and 97.5th percentiles of the resulting distribution of medians.[42] If all the margins permitted by the external bounds fall outside the 95% confidence interval for one of the pricing assumptions, then this pricing assumption can be rejected at the 5% significance level (for all margins permitted by the external bounds).

Consider first the milk category. Here external bounds to margins are 0.20 and 0.34. Under the assumption of supermarket pricing the confidence intervals for margins—unweighted (0.22, 0.28), weighted (0.23, 0.33)—fall within the external bounds. Thus we cannot reject (at the 5% level) the null hypothesis of supermarket pricing (for all margins within the permitted bounds). Under the assumption of independent category pricing, on the other hand, the confidence intervals—(0.65, 0.80) unweighted, (0.69, 1.01) weighted—fall outside the margins permitted by external bounds. Thus we can reject (at the 5% level) the null hypothesis of independent category pricing.

[42]The same 2000 draws are used to generate the CIs in Tables 9 and 10. As a robustness check on the number of draws we calculated the CIs in Table 8 using 5000 draws and found the CIs were almost unchanged: for example the confidence intervals for weighted margins were $(0.23, 0.33)$ for milk and $(0.37, 0.43)$ for all categories under supermarket pricing, and $(0.71, 1.01)$ for milk and $(0.92, 1.09)$ for all categories under independent category sellers.

Table 9: Profit Margins and Cross-Category Externalities

	Median	Mean	(95% CI)	Mean							
	All Categories			Bakery	Dairy	Drink	Dry	Fr,Veg	Hhold	Meat	Milk
A: Profit Margins											
All Firms	0.30	0.31	(0.30, 0.33)	0.31	0.26	0.25	0.33	0.44	0.24	0.42	0.25
Big Four	0.37	0.38	(0.36, 0.41)	0.37	0.20	0.31	0.43	0.56	0.29	0.59	0.26
Discounter	0.28	0.29	(0.27, 0.31)	0.14	0.28	0.22	0.32	0.51	0.23	0.39	0.19
B: Inverse Own-Category Elasticity (Absolute Value)											
All Firms	0.75	0.78	(0.73, 0.84)	1.02	0.75	0.57	0.81	0.93	0.67	0.79	0.75
Big Four	0.98	1.04	(0.97, 1.14)	1.46	0.75	0.72	1.15	1.34	0.86	1.11	0.89
Discounter	0.80	0.83	(0.76, 0.90)	1.10	1.00	0.61	0.78	0.97	0.72	0.84	0.58
C: Marginal Externality me_{fk}/p_{fk}											
All Firms	0.42	0.47	(0.43, 0.52)	0.71	0.49	0.32	0.48	0.49	0.43	0.36	0.49
Big Four	0.61	0.66	(0.60, 0.74)	1.09	0.55	0.41	0.72	0.78	0.57	0.52	0.63
Discounter	0.46	0.54	(0.48, 0.60)	0.96	0.72	0.39	0.46	0.46	0.49	0.45	0.39

Notes: Profit margins and externalities implied by the model in Nash equilibrium (with supermarket pricing). Figures in column All Categories average across categories and figures in rows All Firms, Big Four, and Discounter average across firms of the stated type. By equation (50) figures in Panel A equal those in B minus those in C. We focus on means (as opposed to medians) in this table in order to preserve this adding-up property.

Now consider all categories. Here the external bounds to margins are 0.16 and 0.52. Under the assumption of supermarket pricing the confidence intervals—(0.28, 0.31) unweighted, (0.37, 0.43) weighted—fall within the external bounds so we cannot reject the null hypothesis of supermarket pricing. Under the assumption of independent category sellers the confidence intervals—(0.71, 0.80) unweighted, (0.91, 1.08) weighted—fall outside the margins permitted by the external data and we can reject the hypothesis of independent category pricing (at the 5% level). Thus, even though the external data have quite wide bounds, we can reject the null hypothesis of independent category sellers. We cannot reject the null hypothesis of supermarket pricing.[43] In the rest of this section we use marginal costs under the assumption of supermarket pricing.

6.4 Market Power and Cross-Category Externalities

With elasticities and marginal costs in hand we now analyze the extent to which cross-category externalities abate market power. Panel A of Table 9 reports profit margins for each category for alternative firm types. Recall from equation (50) that the profit margin may be decomposed as the category's inverse elasticity minus its marginal externality on other categories (as a fraction of p_{fk}). We report these two components in Table 9: the inverse elasticity in Panel B and the marginal cross-category externality in Panel C. To preserve the adding-up property of equation (50), we report means rather than medians (medians are reported for All Categories for comparison). Panel A is thus the corresponding figure in Panel B minus its counterpart in Panel C.

The overall mean (across all categories and firms) is 0.31 as reported in the top cell of the

[43] Alternatively, if supermarket pricing is taken as given, the confidence intervals for $\chi_f = 1$ can be used to test the null hypothesis that the model is correct, which, based on the same arguments as above, cannot be rejected.

second column. Profit margins are highest for the Big Four firms, which is not surprising given their large market shares. The inverse elasticities in Panel B are the profit margins that we would have obtained if we had assumed that observed category prices are generated by independent category sellers rather than supermarkets (i.e. if we had set $\chi_f = 0$ in (50) to back out marginal costs).[44] For the mean across all categories, profit margins under independent category pricing are more than double those with supermarket pricing (given in the first row of panel A). This shows that cross-category effects play an important role in correctly assessing market power in the supermarket industry.

The marginal externality reported in the table is a measure of the extent to which competition is intensified by supermarket organization. It can be interpreted as the (Pigouvian) marginal subsidy that must be offered to an independent seller to induce him to set prices that maximize the profits of the supermarket as a whole. The marginal externality is analogous to the concept of "upward pricing pressure" (see Farrell and Shapiro (2010)) used in antitrust policy to measure the anti-competitive effects from a merger of two substitute products. Supermarket organization is analogous to the merger of category sellers selling complementary goods; the marginal externality measures the downward pricing pressure implied.

As Panel C reports, the marginal externality as a fraction of price, i.e. me_{fk}/p_{fk}, is 0.47 on average across firms and categories. The positive sign of the externality indicates that in supermarket mode firms set prices closer to the competitive level than would be the case under independent category sellers. Its magnitude indicates that this pro-competitive effect is economically significant. As a benchmark we note that it is greater than the magnitudes conventionally used to identify problematic merger cases (see CC (2011, Chapter 4) for a discussion).

The table shows the variation in externalities by category. The externality for category k on any other category k' is given by the product of (i) the profit margin of category k' and (ii) the "diversion ratio" between k and k'—i.e. the demand lost to the store on category k' per unit of demand lost on category k. A category has a large externality if these two factors are relatively high. Note that marginal externalities can be high even for categories (such as Bakery) that are a small share of consumer budgets (as shown in Table 2). This suggests that cross-category effects can be important even when measuring market power for a category that is a small fraction of the retailer's sales.

Now compare externalities by firm. The Big Four firms have larger externalities than average. This is in part a consequence of their higher profit margins (as reported in Panel A) and in part a consequence of their shoppers having larger cross-category diversion ratios (as seen in the next subsection).

Table 10: Profit Effect of a Firm-Category Price Increase: Analysis by Shopper Types

Partition by:	(μ, α, Γ)						(Γ)
	All Firms		Big Four		Discounters		All
	Median	95% CI	Median	95% CI	Median	95% CI	Median
A: Derivative of Profit wrt Price $\frac{\partial \pi_f^g}{\partial p_{fk}}$							
(i) All shoppers $g = all$	0.00	(0.00, 0.00)	0.00	(0.00, 0.00)	0.00	(0.00, 0.00)	0.00
(ii) 1-stop types $g = 1$	-0.59	(-0.88, -0.47)	-3.69	(-5.40, -2.41)	-1.02	(-1.37, -0.71)	-0.19
[proportion negative]	[0.89]	(0.82, 0.94)	[0.84]	(0.75, 1.00)	[1.00]	(0.92, 1.00)	[0.67]
(iii) 2-stop types $g = 2$	0.59	(0.47, 0.88)	3.69	(2.41, 5.40)	1.02	(0.71, 1.37)	0.19
[proportion positive]	[0.89]	(0.82, 0.94)	[0.84]	(0.75, 1.00)	[1.00]	(0.92, 1.00)	[0.67]
B: Category Own-Price Shopper Elasticity $\frac{\partial D_{fk}^g}{\partial p_{fk}} \frac{p_{fk}}{D_{fk}^g}$							
(i) All shoppers $g = all$	-0.50	(-0.57, -0.45)	-0.42	(-0.45, -0.35)	-0.57	(-0.63, -0.51)	-0.50
(ii) 1-stop types $g = 1$	-0.39	(-0.45, -0.37)	-0.31	(-0.37, -0.29)	-0.51	(-0.56, -0.43)	-0.37
(iii) 2-stop types $g = 2$	-0.64	(-0.68, -0.56)	-0.51	(-0.54, -0.44)	-0.66	(-0.71, -0.59)	-0.59
C: Cross-category Shopper Diversion ratio: $\Sigma_{k' \neq k} \frac{\partial D_{fk'}^g}{\partial p_{fk}} \Big/ \frac{\partial D_{fk}^g}{\partial p_{fk}}$							
(i) All shoppers $g = all$	3.76	(3.51, 3.89)	4.55	(4.39, 4.67)	4.02	(3.78, 4.20)	3.76
(ii) 1-stop types $g = 1$	5.11	(4.79, 5.21)	5.64	(5.39, 5.74)	5.29	(4.99, 5.43)	5.43
(iii) 2-stop types $g = 2$	2.31	(2.23, 2.44)	3.06	(2.99, 3.17)	2.53	(2.39, 2.72)	2.79

Notes: All figures are medians across firm-category combinations for firms within stated type.

6.5 Competitive Implications of Alternative Shopper Taste Types

We now partition consumers into two groups based on their tendency to visit one or two stores and compare their impact on market power. Each consumer in the model is characterized by a given (μ, α, Γ)-type which fully describes his tastes for the shopping choices in his choice set, up to the idiosyncratic term ε (which is iid across shopping choices). A consumer's choice between one- and two-stop shopping depends on his (μ, α, Γ)-type: for instance, high shopping costs (Γ) and/or a high utility $(\mu - \alpha p)$ at the same store for all categories are conducive to one-stop shopping. After integrating out ε we calculate the probability, for each (μ, α, Γ)-type, that the consumer makes a one-stop shopping choice.[45] If this probability is greater than or equal to 0.5 at the observed prices we place the (μ, α, Γ)-type in the one-stop group. Otherwise he is in the two-stop group. We refer to these two groups as one-stop shopper types and two-stop shopper types respectively. Our goal is to assess which group abates market power to the greater extent. At the end of this subsection we consider an alternative partition based on Γ only.

[44] The 95% confidence intervals presented in this table differ from those in Table 8 because they are confidence intervals of the means rather than medians. By comparing the confidence interval for "A: All Firms" with the second column of Panel A of Table 8, we see that supermarket pricing is not rejected when using means, while the interval for "B: All Firms" shows that category pricing is rejected for all external margins when using means. That is, the conclusions in subsection 6.3 remain unchanged when using means instead of medians.

[45] The probability is $\sum_{c:n(c)=1} P_c(\mu, \alpha, \Gamma)$ where $P_c(\mu, \alpha, \Gamma)$ is given by (18) and $w_{it}(c, p_t)$ depends on (μ, α) as discussed in subsection 3.2.

To compare the impact of the two consumer groups on market power we consider how they respond to a marginal price change for one firm-category combination at a time. These effects are shown in Table 10. Note that the grouping into one-stop and two-stop shopper types is based on consumers' behavior at observed prices and is therefore exogenous to these price changes.

We first calculate profit effects. Let $\pi_f = \Sigma_k \pi_{fk}$ denote firm f's profits (from all categories). Row A(i) of Table 10 shows that the partial derivative of firm f's profit with respect to its price for category k is zero for each (f, k)-combination. This is true by construction since marginal cost has been calculated under the assumption that firms set each price to maximize profit. We now decompose the profit effect by shopper group g. Let $g = 1$ for one-stop shopper types (defined above), and $g = 2$ for two-stop shopper types. Let the profit derived from shopper group g be written π_f^g. Then the change in the firm's profits from the firm-category price increase can be decomposed

$$\frac{\partial \pi_f}{\partial p_{fk}} = \frac{\partial \pi_f^1}{\partial p_{fk}} + \frac{\partial \pi_f^2}{\partial p_{fk}} = 0$$

so that the effects for the two groups must be equal in magnitude and (if non-zero) opposite in sign. Rows A(ii) - A(iii) present the median group-specific profit effects along with 95% confidence intervals for these effects. The median effect is negative for one-stop shopper types and positive for two-stop shopper types. This pattern is seen for almost 90% of category store combinations. (The effects are larger for the Big Four because of their larger market shares). The confidence intervals imply that one-sided tests at the 2.5% significance level reject the null hypotheses that median profit effects are positive for the one-stop types and that they are negative for the two-stop types, respectively. The same is true if tests are performed separately for Big Four and Discounters. Together the results indicate that one-stop shopper types constrain supermarket prices (and hence abate market power) more than two-stop shopper types.

The next two panels in the table explore the factors underlying this finding. The first-order condition (50) implies that the impact of a firm-category (f, k) price change on profit hinges on two main factors: (i) the own-price elasticity of category k demand and (ii) the marginal cross-category externality for firm f. Indicators for these two factors are presented for each consumer type in Panels B and C of Table 10. We measure demand responses in these panels using the number of shoppers D_{fk} for each category (defined analogously to (47) using (20)) which gives a simple count measure of category demand that is easy to interpret and can be added across categories. We decompose D_{fk} by consumer group as follows $D_{fk} = D_{fk}^1 + D_{fk}^2$ where D_{fk}^g denotes the number of shoppers of group g.

Panel B of Table 10 considers own-price effects in terms of shoppers. Two-stop shopper types have higher own-price demand elasticities. This is consistent with the intuition that shoppers that tend to visit two stores can swap stores for a given category at relatively little cost.

Panel C turns to cross-category demand effects. To measure these we use the cross-category diversion ratio in terms of shoppers, i.e.

$$\sum\nolimits_{k' \neq k} \frac{\partial D_{fk'}^g}{\partial p_{fk}} \Bigg/ \frac{\partial D_{fk}^g}{\partial p_{fk}}, \tag{51}$$

Table 11: Robustness

(All Firms, All Categories)	Baseline	Ind p_{jk}	Alt $dist_{ij}$	Alt $\Lambda_{kk'}$	Ind $\Lambda_{kk'}$
A: Margins and Externalities					
Margin $(p_{fk} - mc_{fk})/p_{fk}$	0.30	0.32	0.30	0.30	0.29
Externality me_{fk}/p_{fk}	0.42	0.43	0.43	0.44	0.42
B: Effect of Firm-Category Increase					
1-stop shopper types	-0.59	-0.51	-0.66	-0.62	-0.58
[proportion negative]	[0.89]	[0.89]	[0.89]	[0.88]	[0.83]
2-stop shopper types	0.59	0.51	0.66	0.62	0.58
[proportion positive]	[0.89]	[0.89]	[0.89]	[0.88]	[0.83]

Notes: Entries are medians for All Firms and All Categories and therefore can be compared to the corresponding medians in Table 9 for Panel A and Table 10 for Panel B. The Baseline model is the model used elsewhere in Section 6 (Model 2 in Section 5). Parameters for the other models are in Appendix J.

where the numerator and denominator of this ratio are both negative. In the case of the numerator this is because the categories for a given firm f are complements. The ratio (51) is the total number of shoppers lost by firm f for categories other than k for every shopper (in group g) that stops buying category k. We find a strong difference between shopper types: the median diversion ratio for one-stop shopper types (5.11) is more than double that of two-stop shopper types (2.31).

Taken together the own- and cross-category effects of a price change, shown in Panels B and C, explain why one-stop shopper types constrain the market power of the firms more than the average shopper: such shoppers (i) generate much larger than average cross-category demand effects per marginal consumer (Panel C), and (ii) this effect is strong enough to outweigh the differences between one- and two-stop shopper types in terms of own-category demand elasticity (Panel B).

In the last column we consider an alternative partition of consumer types by shopping costs (Γ) only. Types with above-median shopping costs (for this classification defined as the cost of visiting a second store and traveling 10km) are now classified in group $g = 1$ and the remaining consumers in group $g = 2$. The results are similar despite this partition being less strongly associated with one-stop / two-stop shopping behavior than that based on the full set of tastes (μ, α, Γ).

6.6 Robustness

The results in Section 6 used Model 2 from Section 5. In this section we refer to this as the baseline model. We now consider four alternative specifications of the model. We discuss these specifications before showing that the main findings of the paper are robust.

The model labeled Ind p_{jk} changes the variable used for prices. It has price indices using weights that are computed using individual budget shares rather than budget shares for specific demographic groups (as done in the baseline model). The construction of these individual-specific price indices is described in Appendix C.2.

The model labeled Alt $dist_{ij}$ uses an alternative variable for distance. Instead of the sum of the distances from the consumer to each store and back we use the minimum distance the consumer would travel if he went in a triangle between home, store j, and store j'. This only makes a difference to the distance variable when the consumer is a two-stop shopper. Thus distance when $n(c) = 2$ is now given by $dist_{ic} = (dist_{ij} + dist_{jj'} + dist_{ij'})$ instead of $2(dist_{ij} + dist_{ij'})$. This allows for the possibility that the consumer takes advantage of geographic synergies when visiting the two stores.

The model labeled Alt $\Lambda_{kk'}$ estimates a different set of variable utility interaction parameters than in the baseline model. (These are bakery and dry, dairy and meat, and drink and milk). Cross-price variables in the Z vectors, discussed in 4.1, are modified to reflect the changes in interactions.

Finally the model labeled Ind $\Lambda_{kk'}$ allows for individual (observed and unobserved) consumer heterogeneity in the quadratic second-order terms (for $k' \neq k$). As household size is one of the most important forms of heterogeneity we use the specification

$$\Lambda_{ikk'} = \Lambda_{kk'}(1 + \lambda_1 1[hz_i = 2] + \lambda_2 1[hz_i > 2] + \sigma_5 \nu^{\Lambda}_{ikk'})$$

where $(\lambda_1, \lambda_2, \sigma_5)$ are additional parameters and the terms $1[hz_i = 2]$ and $1[hz_i > 2]$ are indicator variables for household size. $\nu^{\Lambda}_{ikk'}$ is an iid random draw from a standard normal distribution. In addition to the variables in Z^{Q}_{itcjk} and Z^{D}_{itcjk} in the baseline specification we include, for each category, the following two variables: indicators for whether household size is 2, and for whether it is greater than 2, each interacted with the mean price of all categories other than k in the stores in the relevant shopping choice.

We check for robustness to these alternatives by comparing their implications for the main results discussed in Sections 6.4 and 6.5 of the paper. This is done in Table 11. Panel A of the table presents the median profit margins and externalities, across all firms and categories, which we use in Table 9 to assess the pro-competitive effects of supermarket organization. Panel B compares the marginal profit effects of one-stop and two-stop shopper types as we did in Table 10. The table shows these alternative specifications give results that are very similar to those we found for our baseline model. The parameters for the alternative specifications are given in Appendix J.

7 Conclusions

In many important competitive settings, such as retailing, customers buy multiple categories and many prefer to do so from the same location or firm. We develop a multi-store multi-category model for estimation of consumer demand, relevant for the analysis of pricing in such settings. We estimate the model using data from the supermarket industry in the UK. We use the estimated model to analyze two policy-relevant questions: (i) the implications of the internalization of cross-category externalities for the market power of supermarkets and (ii) the relative impact on market power of shopper types inclined to one-stop and two-stop shopping.

The cross-category elasticities we estimate imply that supermarket organization substantially mitigates market power. This has implications for the analysis of retail pricing at two levels. First, at a single-category level of analysis, it indicates a role for considering cross-category effects when using demand elasticities to analyze prices for a given category of interest. In our application we found that accounting for cross-category effects implies a Lerner index typically less than half as large as the Lerner index that would be implied with independent category sellers, so that ignoring cross-category effects can result in market power being overestimated significantly.

Second, at a broader level, the results are relevant for analysis of the organization of the retail industry. Supermarket competition has received much attention—in part because of the large size of firms such as Walmart, Carrefour and Tesco—and policies are sometimes introduced with the aim of protecting or promoting alternative ways of organizing the industry: e.g. planning laws in the UK were tightened in the 1990s to protect town centre retailing, while in France a law (*Loi Raffarin,* 1996) imposed floor space limits on supermarkets with the objective of protecting small traditional retailers. Our empirical results highlight the pro-competitive nature of supermarket pricing relative to alternative ways of organizing retail supply in which pricing is decentralized to independent category sellers.

Comparing one-stop and two-stop shopping types we find that when supermarkets increase the price of a category marginally they lose profits earned on one-stop shopper types and gain profits from two-stop shopper types, which implies that the former constrain supermarket prices more than the latter. This finding suggests it can be appropriate for antitrust authorities to focus on competition for a firm's one-stop (or core) shoppers even where there are many multi-stop shoppers in the firm's customer mix. This is consistent with the position adopted by the FTC in the recent Whole Foods/Wild Oats antitrust case where the question was whether to allow the merger of firms that compete for the same group of one-stop (or core) shoppers, when the firms also sell to two-stop (or cross) shoppers. More generally the finding indicates that the presence of consumers inclined toward two-stop shopping (e.g. those with low shopping costs) does not necessarily enhance competitive pricing incentives.

References

[1] Ackerberg D, Benkard L, Berry S, and A. Pakes (2007) "Econometric Tools for Analyzing Market Outcomes" in Heckman J. and E. Leamer (Eds.) *The Handbook of Econometrics* Vol. 6A. Amsterdam: North-Holland

[2] Armstrong, M. and J. Vickers (2010) "Competitive Non-linear Pricing and Bundling" *Review of Economic Studies*, 77(1), 30-60

[3] Basker, E. (2007) "The Causes and Consequences of Wal-Mart's Growth" *Journal of Economic Perspectives*, 21(3), 177-198

[4] Beggs, A. (1992) "Mergers and Malls" *Journal of Industrial Economics,* 42(4), 419-28

[5] Bell, D., Ho T. and C. Tang (1998) "Determining Where to Shop: Fixed and Variable Costs of Shopping" *Journal of Marketing Research*, 35(3), 352-369

[6] Berry, S., J. Levinsohn, and A. Pakes (2004) “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market” *Journal of Political Economy*, 112(1), 68-105

[7] Besanko, D., Dubé, J-P., and S. Gupta (2003) “Competitive Price Discrimination Strategies in a Vertical Channel Using Aggregate Retail Data,” *Management Science,* 49, 1121-1138

[8] Bonnet, C. and P. Dubois (2010) “Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance” *RAND Journal of Economics,* 41(1), 139–164

[9] Chen Z. & P. Rey (2012). "Loss Leading as an Exploitative Practice," *American Economic Review* 102(7): 3462-3482.

[10] Chevalier, J., A. Kashyap, and P. Rossi (2003) “Why Don’t Prices Rise During Periods of Peak Demand? Evidence from Scanner Data,” *American Economic Review*, 93(1), 15-37.

[11] Chintagunta, P.K. and I. Song (2007), “A Discrete/Continuous Model for Multi-Category Purchase Behavior of Households,” *Journal of Marketing Research*, 44, 595-612

[12] Corstjens, M. and R. Lal (2008), “Building Store Loyalty through Store Brands,” *Journal of Marketing Research,* 37, 281-291

[13] Competition Commission (2000) *Supermarkets: A report on the supply of groceries from multiple stores in the United Kingdom*, The Stationery Office, London

[14] Competition Commission (2008) *The Supply of Groceries in the UK*, The Stationery Office, London

[15] Competition Commission (2011) *Commentary on Retail Mergers*, The Stationery Office, London

[16] Cournot, A. [1838] (1897) *Researches into the Mathematical Principles of the Theory of Wealth* (trans. N. Bacon), Macmillan, London.

[17] Davis, P. (2006) “Spatial Competition in Retail Markets: Movie Theaters” *The RAND Journal of Economics* 37(4), 964-982

[18] Dubé, J-P. (2005) “Product Differentiation and Mergers in the Carbonated Soft Drink Industry,” *Journal of Economics and Management Strategy*, 14(4), 879–904

[19] Dubin J. and D. McFadden (1984) “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption” *Econometrica,* 52(2), 345-362

[20] Economides, N (1989) “Desirability of Compatibility in the Absence of Network Externalities” *American Economic Review,* 79(5), 1165–1181

[21] Farrell, J. and C. Shapiro (2010) "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition" *The B.E. Journal of Theoretical Economics*, 10(1), Article 9

[22] Fox, E., Montgomery A., and L. Lodish (2004) "Consumer Shopping and Spending across Retail Formats," *Journal of Business,* 77, S25-S60

[23] Gentzkow, M. (2007). "Valuing New Goods in a Model with Complementarity: Online Newspapers" *American Economic Review,* 97(3), 713-744

[24] Gould, E., P. Pashigian, and J. Prendergast (2005) "Contracts, Externalities, and Incentives in Shopping Malls" *Review of Economics and Statistics,* 87(3), 411-422

[25] Griffith R., Nesheim L., and M. O'Connell (2010) "Sin taxes in differentiated product oligopoly: an application to the butter and margarine market," Cemmap working paper 37/10

[26] Haneman, W. (1984) "Discrete/Continuous Models of Consumer Demand" *Econometrica,* 52(3), 541-562

[27] Hausman, J., Leonard, G., and D. Zona (1994) "Competititve Analysis with Differentiated Products" *Annales d'Economie et de Statistique*, 34, 159-180

[28] Hendel, I. (1999) "Estimating Multiple Discrete Choice Models: An Application to Computerization Returns" *Review of Economic Studies*, 66, 423-446

[29] Houde, J.-F. (2012) "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline" *American Economic Review*, 102(5), 2147-2182

[30] Kim J., G. Allenby, and P. Rossi (2002) "Modeling consumer demand for variety" *Marketing Science,* 21(3), 229-250

[31] Klemperer, P. (1992) "Equilibrium Product Lines: Competing Head-to-Head May be Less Competitive" *American Economic Review*, 82(4), 740–755

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

[33] Lal, R. and R. Rao (1997) "Supermarket Competition: The Case of Every Day Low Pricing" *Marketing Science*, 16(1), 60-80

[34] McFadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration" *Econometrica*, 57(5), 995–1026

[35] Matutes, C. and P. Regibeau (1988) "Mix and Match: Product Compatibiltiy without Network Externalities" *RAND Journal of Economics,* 19(2), 221-234

[36] Mehta, D. (2006) "Investigating Consumers' Purchase Incidence and Brand Choice Decisions Across Multiple Product Categories: A Theoretical and Empirical Analysis" *Marketing Science*, 26(2), 196-217

[37] Moffitt, R. (1999), "New developments in econometric methods for labor market analysis", in Ashenfelter, O. and Card, D. eds., *Handbook of Labor Economics*, Vol. 3, Part A, Amsterdam: North-Holland

[38] Nalebuff, B. (2000) "Competing against bundles" in Hammond, P., and G. Myles, (Eds.), *Incentives, Organization, and Public Economics*, Oxford University Press, Oxford

[39] Nalebuff, B. (2009) "Bundling: GE-Honeywell (2001)," in Kwoka, J. and L. White, (Eds.) *The Antitrust Revolution*, 5th edition, 416–440, Oxford University Press, New York

[40] Nevo, A. (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry" *Econometrica*, 69(2), 307-342

[41] Office for National Statistics (2014) "Consumer Price Indices Technical Manual", available at http://www.ons.gov.uk

[42] Pakes, A. and D. Pollard (1989) "Simulation and the Asymptotics of Optimization Estimators" *Econometrica,* 57(5), 1027-57

[43] Smith, H. (2004) "Supermarket Choice and Supermarket Competition" *Review of Economic Studies,* 71(1), 235-263

[44] Smith, H. and D. Hay (2005) "Streets, Malls and Supermarkets" *Journal of Economics and Management Strategy*, 14(1), 29-59

[45] Smith, H. and O. Thomassen (2012) "Multi-category Demand and Supermarket Pricing" *International Journal of Industrial Organization*, 30(3) 309-314

[46] Stahl, K. (1982) "Location and Spatial Pricing Theory with Nonconvex Transportation Cost Schedules" *Bell Journal of Economics,* 13(2), 575-582

[47] Stern, S. (1997) "Simulation-Based Estimation" *Journal of Economic Literature*, 35(4), 2006-2039

[48] Sudhir, K. (2001) "Structural Analysis of Manufacturer Pricing in the Presence of a Strategic Retailer," *Marketing Science,* 20, 244-264

[49] Thomassen, Ø (2009) *Estimating the Demand for Automobile Engine Variants*, DPhil Thesis, Oxford University

[50] Thomassen, Ø (2017) "An Empirical Model of Automobile Engine Variant Pricing", Working paper

[51] Villas-Boas, SB (2007) "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data" *Review of Economic Studies,* 74(2), 625-652

[52] Vroegrijk, M., E. Gijsbrechts. and K. Campo (2013) "Close Encounter with the Hard Discounter: A Multiple-Store Shopping Perspective on the Impact of Local Hard-Discounter Entry" *Journal of Marketing Research,* 50(5), 606-626

[53] Wales T. and A. Woodland (1983) "Estimation of consumer demand systems with binding nonnegativity constraints" *Journal of Econometrics,* 21(3), 263-285

A Online Appendix: Category Definitions

TNS assigns to each transaction the variable "Retailer Share Track (RST) Market Code" that correspond to 269 narrowly defined product groups. We define our eight categories as follows where the names of product groups (including abbreviations) are those of TNS.

1. *Bakery*: Ambient Pizza Bases, Ambient Cakes and Pastries, Ambient Christmas Pudding, Ambient Sponge Puddings, Canned Rice Puddings, Childrens Biscuits, Chilled Breads, Chilled Cakes, Chilled Desserts, Chilled Pizza and Bases, Crackers & Crispbreads, Everyday Biscuits, Fresh/Chilled Pastry, Frozen Bread, Frozen Savoury Bakery, Healthier Biscuits, Morning Goods, Savoury Biscuits, Seasonal Biscuits, Tinned Sponge Puddings, Toaster Pastries, Total Bread.

2. *Dairy*: Butter, Defined Milk and Cream Products, Fresh Cream, Fromage Frais, Instant Milk, Margarine, Total Cheese, Total Ice Cream, Yoghurt, Yoghurt Drinks And Juices.

3. *Drink*: Ambient One Shot Drinks, Ambient Fruit or Yoghurt Juice and Drnk, Beer and Lager, Bottled Colas, Bottled Lemonade, Bottled Other Flavours, Bottled Shandies, Canned Colas, Canned Lemonade, Canned Other Flavours, Canned Shandies, Chilled One Shot Drinks, Cider, Fabs, Food Drinks, Fortified Wines, Ginger Ale, Lemon and Lime Juices, Mineral Water, Soda Water, Sparkling Wine, Spirits, Tonic Water, Wine.

4. *Dry*: Ambient Condiments, Ambient Slimming Products, Ambient Vegetarian Products, Artificial Sweetners, Breakfast Cereals, Chocolate Biscuit Bars, Chocolate Confectionery, Chocolate Spread, Confectionary. & Other Exclusions, Cooking Oils, Crisps, Dry Meat Substitutes, Dry Pasta, Dry Pulses and Cereal, Ethnic Ingredients, Everyday Treats, Flour, Frozen Confectionery, Gum Confectionery, Herbal Tea, Herbs and Spices, Home Baking, Honey, Instant Coffee, Lards and Compounds, Liquid and Ground Coffee and Beans, Mincemeat (Sweet), Mustard, Packet Stuffing, Peanut Butter, Pickles Chutneys & Relish, Powder Desserts & Custard, Preserves, RTS. Custard, Ready To Use Icing, RTS Desserts Long Life, Salt, Savoury Snacks, Sour and Speciality Pickles, Special Treats, Suet, Sugar, Sugar Confectionery, Sweet and Savoury Mixes, Syrup & Treacle, Table Sauces, Table and Quick Set Jellies, Tea, Vinegar.

5. *Fruit and Vegetables*: Ambient Olives, Ambient Rice and Savoury Noodles, Ambient Salad Accompaniment, Baked Bean, Bitter Lemon, Canned Fish, Canned Hot Meats, Canned Salads, Canned Vegetables, Chilled Fruit Juice and Drink, Chilled Olives, Chilled Prepared Fruit and Veg, Chilled Prepared Salad, Chilled Rice, Chilled Salad Accompaniment, Chilled Vegetarian, Cous Cous, Frozen Potato Products, Frozen Vegetables, Frozen Vegetarian Prods, Fruit, Instant Mashed Potato, Nuts, Prepared Peas & Beans, Tinned Fruit, Tomato Products, Total Fruit Squash, Vegetable.

6. *Household*: Air Fresheners, Anti-Diarrhoeals, Antiseptics & Liq. Disinfectant, Bath Additives, Batteries, Bin Liners, Bleaches & Lavatory Cleaners, Body Sprays, Carpet Clean-

ers/Stain Removers, Cat Litter, Cat and Dog Treats, Cleaning Accessories, Cold Sore Treatment, Cold Treatments, Conditioners and Creme Rinses, Contact Lens Cleaners, Cotton Wool, Cough Liquids, Cough Lozenges, Decongestants, Dental Floss or Sticks, Dentifrice, Denture Cleaners/Fixature, Deodorants, Depilatories, Dog Food, Electric Light Bulbs, Eye Care, Fabric Conditioners, Facial Tissues, First Aid Dressings, Foot Preparations, Furniture Polish, Hair Colourants, Hairsprays, Hand Wash Products, Hayfever Remedies, Home Perms, Household Cleaners, Household Food Wraps, Household Insecticides, Incontinence Products, Indigestion Remedies, Kitchen Towels, Laxatives, Liquid Soap, Machine Wash Products, Mens Hairsprays, Mens Mass Fragrances, Mens Skincare, Moist Wipes, Mouthwashes, Oral Analgesics, Oral Lesion/teething, Pot Pourri and Scented Candles and Oils, Razor Blades, Sanpro, Shampoo, Shaving Soaps, Shoe Care Products, Skincare, Sleeping Aids, Sun Preparations, Talcum Powder, Toilet Soap inc. Mens, Toilet Tissues, Topical Analgesics, Topical Antiseptics, Total Cat Food inc. Bulk, Total Dry Dog Food, Total Male and Female Styling, Total Toothbrushes, Upset Stomach Remedies, Vitamin and Mineral supplements, Wash Additives, Washing Up Products.

7. *Meat*: Ambient Cooking Sauces, Ambient Dips, Ambient Pastes and Spreads, Ambient Sandwich Fillers, Ambient Soup, Canned Pasta Products, Chilled Black and White Pudding, Chilled Burgers and Grills, Chilled Cooking Sauces, Chilled Dips, Chilled Gravy and Stock, Chilled Pate and Paste and Spread, Chilled Prepared Fish, Chilled Processed Poultry, Chilled Ready Meals, Chilled Sausage Meat, Chilled Frankfurter/Continental Sausages, Chilled Sandwich Fillers, Cold Canned Meats, Complete Dry/Ambient Meals, Cooked Meats, Cooked Poultry, Fresh Bacon Joint, Fresh Bacon Rashers, Fresh Bacon Steaks, Fresh Beef, Fresh Flavoured Meats, Fresh Lamb, Fresh Other Meat & Offal, Fresh Pasta, Fresh Pork, Fresh Poultry, Fresh Sausages, Fresh Soup, Frozen Bacon, Frozen Beef, Frozen Cooked Poultry, Frozen Fish, Frozen Flavoured Meats, Frozen Lamb, Frozen Meat Products, Frozen Other Meat & Offal, Frozen Pizzas, Frozen Pork, Frozen Poultry, Frozen Processed Poultry, Frozen Ready Meals, Frozen Sausage Meat, Frozen Sausages, Hens Eggs, Instant Hot Snacks, Loose Fresh Meat & Pastry, Meat Extract, Other Chilled Convenience, Other Frozen Foods, P/P Fresh Meat and Veg and Pastry, Packet Soup, Shellfish, Wet or Smoked Fish.

8. *Milk*: Total Milk.

B Estimation Sample: Construction & Representativeness

We implement the sample selection by drawing a week at random for each consumer to represent his third (and final) week in the estimation sample (this must be drawn from outside his first two quarters in the sample). To obtain the second and first weeks in the estimation sample we use the weeks that are one quarter-year and two quarter-years before the third week. When these exact weeks are not available we substitute the most recent available week (that is at least one quarter or two quarters before the final week). We drop consumers for whom three weeks cannot be obtained using this method because they do not participate long enough to be in the data for three successive quarters, which results in a loss of 23% of the initial sample of 26,191. We then draw 2000 of the remaining consumers at random to form an estimation sample of 6000 consumer-weeks.

Sample selection problems could arise either because the TNS sample is not representative for the UK population or because the subsample we select is not representative for the full TNS sample. Regarding the latter issue, note that we select consumers almost randomly subject to the constraint that they are in the sample long enough so that we observe each consumer in 3 different quarters. Regarding the former issue, TNS claims to survey a representative sample of consumers and has a commercial interest in making the sample representative.[46] Nevertheless, we analyze explicitly whether the sample is representative by comparing demographics across our sample, the full TNS sample and census data. In Table 12 Full Sample refers to the consumers in the raw sample, Estimation Sample refers to the 2000 consumers selected for estimation, and Validation Sample refers to the 2000 consumers used in the out-of-sample analysis in Section 5. A comparison of sample moments shows that they are similar. The column Great Britain refers to data from 2001 census and allows comparison between the TNS sample means and those of the population.

Table 12: Comparison of Sample Moments

	Full Sample		Estimation Sample		Validation Sample		Great Britain
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean
Household-level statistics:							
Number of Adults	2.0	0.8	2.0	0.8	2.0	0.8	NA
Number of Children	0.8	1.1	0.8	1.1	0.7	1.0	NA
Household size	2.8	1.4	2.8	1.3	2.7	1.3	2.4
Characteristics of Household Reference Person:							
Home owner (0/1)	0.74	0.41	0.78	0.41	0.76	0.43	0.69
Age	45.5	16.0	46.9	15.8	47.0	16.0	49.2
Retired (0/1)	0.17	0.37	0.17	0.38	0.18	0.38	0.17
Employed (0/1)	0.68	0.47	0.69	0.46	0.67	0.47	0.68
Unemployed (0/1)	0.03	0.16	0.03	0.15	0.03	0.16	0.03
Observations:	26,191×67.6		2000×3		2000×3		—

Notes: The Household Reference Person is a senior member of the household identified using criteria used for the 2001 census in Great Britain. All figures in the column marked Great Britain are for Great Britain from the 2001 Census with the following exceptions: (i) the figure for home ownership is from GB Housing Statistics, rather than the Census, and (ii) the figures for Retired, Employed, and Unemployed status in the last column are for England & Wales only as Scotland does not report this breakdown for the Household Reference Person (when Scotland is eliminated from the Full Sample, the Estimation Sample and the Validation Sample, it does not change the moments reported in the table for these variables).

[46] The commercial value of the data in the form of market analysis for firms requires a representative sample. Kantar describes the panel as a "purchase panel consisting of 30,000 demographically representative households in GB" (http://www.kantarworldpanel.com/en/Consumer-Panels-/alcohol", retrieved 10/7/2016).

C Online Appendix: Price Index Construction

C.1 Baseline Price Indices

The prices used in the model are computed at category-week-store-demographic group level for categories $k = 1, .., 8$ using the full sample of transactions in the TNS data. (See below for a description of the demographic groups).

In data there are two levels of aggregation below category k. First, in each category k (e.g. “Household Goods”), there is a set of narrowly-defined product groups g (e.g. “Shampoo”) listed in Appendix A. We drop some minor product groups that are not sold by all firms, which leaves 183 (out of 268) product groups that account for 96% of consumer expenditure. We define this set of product groups $\mathcal{G}_k$ for each k.

Second, within each product group $g \in \mathcal{G}_k$ there is a set of products h, each of which is a unique product and pack size (e.g. “Herbal Essences Fresh Balance Shampoo 200ml” is a product in the “Shampoo” group). Products h are numerous and there is a tail of products with low volume. For each firm f we select products h that appear in the data at least once in each year (2002 - 2005) and in more than six quarterly periods. This yields a set of products, $\mathcal{H}_{fg}$, for each firm f and product group g. For each store j product h and week t we compute price p_{jht} as the median price of product h for week t for stores operated by store j's firm $f(j)$. As noted in Section 2 the predominant pricing practice is national pricing, in which firms do not vary prices depending on the location of their stores. In cases where there are no observed prices for a particular week we impute the price using the median price for the quarter-year in which week t falls. We obtain 13 firm-level prices for each t and h: one for each of the following: ASDA, Morrison, Sainsbury, Tesco, M&S, Waitrose, Aldi, Lidl, Netto, Iceland, Co-op, and Somerfield, and smaller chains.

The aggregation to category k level thus proceeds in two stages: (i) from product h to product group g and (ii) from product group g to category k. In each of these stages we weight the prices to reflect their importance using information from the transactions data.

To allow for taste variation at an intra-category level we compute weights separately for the eight demographic types $m = 1, .., 8$ which are combinations of social class and household size categories. The TNS household characteristics data has six social class levels $(1, ..., 6)$ based on occupational group. These social class indicators are used widely in United Kingdom as a measure of socioeconomic status. A lower number on this scale has a higher average household income. We combine social class level 1 and 2, and likewise 5 and 6, as there are relatively few households in these groups, which yields four social class categories. For each of these we divide households into two size groups—small (one or two people) and large (more than two people)—which yields the eight demographic types.

In the first stage of aggregation the product group g price in store j for week t and demographic group m is given by $p^m_{jgt} = \sum_{h \in \mathcal{H}_{gf(j)}} w^m_{hf(j)} p_{jht}$ where $w^m_{hf(j)}$ are volume weights. We use volume weights at this stage since there is a common volume unit for products within each g (e.g. volumes in “Shampoo” are in ml). If each product were sold in each firm then we could proceed using volume weights $w^m_{hf} = Q^m_h / Q^m_g$ where Q^m_h is the total volume of product h sold to demographic group m

over the three year period and Q_g^m is the total volume sold in product group g to demographic group m over the three year period. However, each product h is not sold by all firms so we instead compute $\tilde{w}_{hf}^m = Q_h^m / Q_{g|h}^m$ where $Q_{g|h}^m$ is the volume sold in product group g to demographic group m by firms selling product h and let $w_{hf}^m = \tilde{w}_{hf}^m \Big/ \sum_{h \in \mathcal{H}_{gf}} \tilde{w}_{hf}^m$ in order to ensure that the weights add up to one for any firm (i.e. $\sum_{h \in \mathcal{H}_{gf}} w_{hf}^m = 1$ for any f) . This weights products using information that is not specific to firm f for products that are sold by more than one firm and uses firm f specific information otherwise.

In the second stage of aggregation we obtain the category price p_{jkt}^m using is a revenue-weighted average of product group price ratios p_{jgt}^m / p_{bg}^m (where p_{bg}^m is an arbitrary base price):

$$p_{jkt}^m = \sum_{g \in \mathcal{G}_k} \omega_g^m \left(\frac{p_{jgt}^m}{p_{bg}^m} \right). \tag{52}$$

The weights ω_g^m are the total expenditure share (over the three year period) of each product group g for demographic type m (where $\sum_{g \in \mathcal{G}_k} \omega_g^m = 1$ for each m). The weights are constant across stores and over time. Following common practice in price index construction (see for example Chapter 2 in ONS(2014)[47]) we (i) use sales rather than volume weights at this upper level of aggregation because the different product groups are often in different units, and (ii) use price ratios in (52) to ensure that p_{jkt}^m is independent of the units chosen within any product group. We set the arbitrary base price p_{bg}^m in the price ratio to be the price in the first week ($t = 1$) in ASDA stores.

C.2 Individual Price Indices

The individual price indices used in subsection 6.6 differ in the second stage of aggregation by using an individual-specific weighting term—instead of a demographic group weighting term—to aggregate from product group (g) to category (k) level. (We do the individual weighting at the product group g level but not the individual product h level because many individual products such as "private labels" are firm-specific and an individual consumer typically only visits a subset of the firms in the data). The category price p_{jkt}^i for individual consumer i is a budget share-weighted average of price ratios p_{jgt}^m / p_{bg}^m at product group level (where p_{jgt}^m and p_{bg}^m are as defined above for the baseline price indices):

$$p_{jkt}^m = \sum_{g \in \mathcal{G}_k} \omega_g^i \left(\frac{p_{jgt}^m}{p_{bg}^m} \right) \tag{53}$$

where weights ω_g^i are now the total expenditure share (over the three year period) of each product group g by consumer i and satisfy $\sum_{g \in \mathcal{G}_k} \omega_g^i = 1$ for each i. The weights are constant across stores and over time.

[47] Office for National Statistics (2014) "Consumer Price Indices Technical Manual", available at http://www.ons.gov.uk

D Online Appendix: A General Multi-Store-Category Model

In this Appendix we show how the model we estimate can be derived from a more general framework of multi-store and multi-category demand. At the most general level, the consumer chooses for each category $k \in \{1, ..., K\}$ in every store $j \in \{1, ..., J\}$, how much quantity q_{jk} to purchase, subject to his budget constraint:

$$\begin{aligned} \max_{\mathbf{q} \geq 0} \quad & V(\mathbf{q}, \theta, X) \\ \text{s.t.} \quad & \mathbf{p}'\mathbf{q} \leq y, \end{aligned}$$

where $\mathbf{q} = (q_{11}, ..., q_{JK}, q_0)$ denotes the quantity vector for all store/category combinations, and the outside option is q_0. The price vector $\mathbf{p} = (p_{11}, ..., p_{JK}, 1)$ is defined analogously (the price of the outside good is normalized to one). θ is a vector of parameters to be estimated and X a vector of observable store, category and consumer characteristics. y denotes the consumer's income.

In this setting corner solutions in quantity are likely to arise and they can originate from two sources. Either the consumer does not visit a particular store and hence cannot purchase any positive quantity there. Or the consumer might visit the store, but decides not purchase any quantity in a specific category. Dealing with the choice over $J \times K$ quantities with possible corner solutions for many of the quantities makes this a difficult demand system to estimate and we hence impose a set of restrictions based on the data patterns described in Section 2 of the paper. Specifically, we assume that the cost of visiting more than two stores is prohibitively high, so that no consumer wishes to visit a third store in a given week and that consumers only purchase at one store within a given category.

With these restrictions on the utility function, we can re-write the optimization problem in the following way:

$$\begin{aligned} \max_{c} \max_{d} \max_{q \geq 0} \quad & V(c, d, q, \theta, X) \\ \text{s.t.} \quad & p_d' q + q_0 \leq y \end{aligned}$$

This formulation allows us to break up the problem into a discrete choice between (pairs of, or single) stores c, a discrete choice of store for each category d and a continuous quantity choice in each category q. To derive equation (1) in Section 3 of the paper this formulation also assumes utility is additively separable in the variable utility derived from purchasing a specific basket of goods and shopping costs and that variable utility is linear in the outside good. Substituting the budget constraint into the variable utility function for the quantity of the outside option q_0 yields (1).

E Category Aggregation

In this Appendix we discuss the assumptions under which we can aggregate demand from product to category level, and show how this can underpin the utility function we use in the paper. Our derivation follows closely the established literature on aggregation (see Deaton and Muellbauer (1980)). We adopt the approach of using separability restrictions on preferences (see Gorman (1953) as opposed to that of assuming colinear prices (Hicks (1946)). This approach is commonly used in papers that require aggregation to estimate models in an AIDS framework (see for example Hausman and Zona (1994)).

Let us first define some notation. Let there be H products and let the quantities bought by a consumer be $\mathbf{x} = (x_1, x_2, x_3, ..., x_H)$. These can be grouped by category using $\mathbf{x}_k$ so that we may write $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_k, ..., \mathbf{x}_K)$. In a similar way let product level prices be $\mathbf{p} = (p_1, p_2, p_3, ..., p_H) = (\mathbf{p}_1, ..., \mathbf{p}_k, ..., \mathbf{p}_K)$. To distinguish category level from product level prices we use ρ_k to denote the category level price index (note this deviates from the notation in the main text). We denote category aggregate quantities $(q_1, ..., q_K)$. Let $\mathbf{x}_{-k}$ denote a consumption vector for products not in category k. Weak separability for category k requires $(\mathbf{x}_k^1, \mathbf{x}_{-k}) \succeq (\mathbf{x}_k^0, \mathbf{x}_{-k}) \Rightarrow (\mathbf{x}_k^1, \mathbf{x}_{-k}^*) \succeq (\mathbf{x}_k^0, \mathbf{x}_{-k}^*)\ \forall \mathbf{x}_{-k}^*$ i.e. the quantities an agent consumes of products in other categories $(\mathbf{x}_{-k}^*)$ does not change the preferences a consumer has between any two bundles in category k (here, $\mathbf{x}_k^1$ and $\mathbf{x}_k^0$). This in turn implies that the consumer's problem may be written:

$$\max_{\mathbf{x}} u = U[v_k(\mathbf{x}_k), \mathbf{x}_{-k}] \text{ subject to } y = \mathbf{p}\mathbf{x} \tag{54}$$

where $v_k(\mathbf{x}_k)$ is a category-specific utility function for category k and y is the consumer's overall budget.

We can now divide the consumer's problem into two stages: a "first stage" inter-category budget allocation decision in which budget y_k is allocated to category k and a number of independent "second stage" problems in which the utility $u_k = v_k(\mathbf{x}_k)$ from category k is maximized given the budget y_k. The indirect utility for the second stage problem is

$$\psi_k(y_k, \mathbf{p_k}) = \max_{\mathbf{x}_k} v_k(\mathbf{x}_k) \text{ subject to } \mathbf{p}_k\mathbf{x}_k = y_k.$$

The first stage decision of how much budget to allocate to category k can be characterized as a decision of how much category-specific utility u_k to enjoy, i.e. if category k is weakly separable (without saying anything about the other categories) we can rewrite (54) as

$$\max_{u_k, \mathbf{x}_{-k}} U[u_k, \mathbf{x}_{-k}] \text{ subject to } y = e_k(u_k, \mathbf{p}_k) + \mathbf{p}_{-k}\mathbf{x}_{-k} \tag{55}$$

where the category level expenditure function (dual to the indirect utility function) is substituted in place of the category k budget.

If the agent's preferences over products in category k are homothetic then we can write the category specific indirect utility

$$\psi_k(y_k, \mathbf{p_k}) = y_k/\rho_k(\mathbf{p_k}) \tag{56}$$

where $\rho_k(\mathbf{p_k})$ is the lower-stage price index which must be homogeneous of degree one. From this expression it follows that $u_k = e_k(u_k, \mathbf{p}_k)/\rho_k(\mathbf{p_k})$ and hence (by rearranging) the amount budgeted is $e_k = u_k\rho_k(\mathbf{p_k})$ which allows us to replace the expenditure function inside the budget constraint expression in (55) above so that the "first stage" decision can be rewritten

$$\max_{u_k, \mathbf{x}_{-k}} U[u_k, \mathbf{x}_{-k}] \text{ subject to } y{=}u_k\rho_k(\mathbf{p_k}) + \mathbf{p}_{-k}\mathbf{x}_{-k}.$$

From this it follows that u_k plays the role of a category-level quantity aggregate and $\rho_k(\mathbf{p_k})$ as a category level price index. Using (56), the category quantity q_k is obtained by dividing category expenditure y_k by the price index $\rho_k(\mathbf{p_k})$.

We can now derive the utility function we use in the paper. The derivation above extends easily to the case of weak separability between all categories. Under this assumption, we can write overall utility as being composed of category-level sub-utilities.

$$u = U[v_1(\mathbf{x}_1), ..., v_k(\mathbf{x}_k), ...]$$

where $\mathbf{x}_k$ denotes the vector of quantities of goods within category k and $v_k()$ denotes the sub-utility function for category k. Under the additional assumption of homotheticity (within each category), the indirect utility in each category is given by $y_k/\rho_k(\mathbf{p_k})$, where y_k is category-level expenditure and $\rho_k(\mathbf{p_k})$ is a price index. $y_k/\rho_k(\mathbf{p_k})$ can be interpreted as a category-level quantity index.[48] We can now think of the utility-function across categories as

$$u = U[v_1(\mathbf{x}_1), ..., v_k(\mathbf{x}_k), ...] = U[q_1, ..., q_k, ...] = \mu_d' q - 0.5q'\Lambda q + \alpha q_0 \tag{57}$$

This function defines variable utility across all categories (and the outside option). Substituting the budget constraint for q_0 and adding the shopping cost term, we obtain equation (2) in the main text.

References

[1] Deaton A. and J. Muellbauer (1980) *Economics and Consumer Behavior*, Cambridge University Press.

[2] Gorman, W. (1953) "Community preference fields", *Econometrica*, 21, 63-80.

[3] Hicks, J. (1946) *Value and Capital*, 2nd ed., Oxford: Clarendon Press.

[48] Under the assumption of additive (strong) separability between categories, we can replace the assumption that product-level utility is homothetic, with the weaker assumption that product-level utilities have a Generalized Gorman Polar Form.

F Online Appendix: Variables used in Moments

- $[Z^Q_{itcjk}]$: household i size hz_i, time dummies T_t (2 years, 3 quarters), price p_{tjk}, price for categories $k' \neq k$ (for which we estimate cross effects), log store j size (sz_j), indicator that there are two stores in shopping choice c, $1_{[n(c)=2]}$, firm dummies (eight of the nine firms in footnote (21)) and a constant.

- $[Z^D_{itcjk}]$: as Z^Q_{itcjk} but without time dummies T_t.

- $[Z^I_{itc}]$: distance $dist_{itc}$, two-stop shopping indicator $1_{[n(c)=2]}$, distance squared $(dist_{itc})^2$, interaction of distance and the two-stop shopping indicator $1_{[n(c)=2]}$, mean price across categories and stores for shopping choice c at time t, and mean price across categories and stores for shopping choice c at time t divided by per capita income.

G Likelihood Function

In this Appendix we derive the likelihood function for the model. The observed choice outcome is the triple (c, d, q): a discrete shopping choice c (up to two stores), a vector $d = (d_1, .., d_K)$ which indicates the store $j \in c$ chosen for each category $k = 1, .., K$, and a K-vector of continuous choices $q = (q_1, \ldots, q_K)$ for each category. The likelihood of this choice outcome at parameters θ for an individual consumer, written $L(c, q, d|\theta)$, is given by the probability that his unobserved tastes (ν, ϵ) are in the region that rationalize the choice (c, d, q) given the taste density $f(\nu|\theta)$ and the type-1 extreme value distribution for ϵ.

We proceed in four steps. First, in section G.1 we express the variable utility specification in an alternative but equivalent way that facilitates the derivation of the likelihood. Second, in section G.2, we derive the set of unobserved tastes ν that are consistent with utility maximization given the category choices (d, q) at shopping choice c. Third, in section G.3, we derive the likelihood of the observed category decisions (d, q) treating the shopping choice (c) as exogenous. Finally, in section G.4, we derive the joint likelihood of the triple (c, d, q) allowing the shopping choice (c) to be endogenous. Throughout we consider a single consumer-week observation and so to avoid clutter we can suppress (i, t) subscripts from the notation.

The likelihood is a generalization of likelihoods derived previously for two separate groups of models: (i) those that consider corner solutions for products (or product categories) but do not allow for store choices (Kim et al (2002) and Wales and Woodland (1982)), and (ii) discrete-continuous models that consider only a single continuous choice but do not allow for zero expenditures (Dubin McFadden (1984), Smith (2004)).

G.1 Variable Utility and Category-Store Choices

In this section we present the variable utility specification in our paper (i.e. the first two terms in (1)) in an alternative but equivalent form that facilitates derivation of the likelihood. Given the specification assumptions of our model, outlined in subsection 3.2, we can write the consumer's variable utility from the choice (q, d, c) as

$$\begin{aligned} u(q,d,c) &= (\mu_d - \alpha p_d)' q - 0.5 q' \Lambda q \\ &= \sum_{k=1}^{K} \sum_{j \in c} [(\mu_{jk} - \alpha p_{jk}) 1(d_k = j)] q_k - 0.5 \sum_{k=1}^{K} \sum_{k'=1}^{K} \Lambda_{kk'} q_k q_{k'} \end{aligned}$$

where $1(d_k = j)$ is an indicator that takes the value 1 if the consumer chooses store j for category k and 0 otherwise.

Decomposing the taste coefficient μ_{jk} into its deterministic $(\bar{\mu}_{jk})$ and random (ν^{μ}_{jk}) components we have

$$\sum_{k=1}^{K} \sum_{j \in c} [(\bar{\mu}_{jk} + \nu^{\mu}_{jk} - \alpha p_{jk}) 1(d_k = j)] q_k - 0.5 \sum_{k=1}^{K} \sum_{k'=1}^{K} \Lambda_{kk'} q_k q_{k'} \tag{58}$$

Recall that α is a random term. It is convenient to write the random effects in (58) terms of a single unobserved taste term for each (j, k) so we define $\nu^{w}_{jk} = \nu^{\mu}_{jk} - \alpha p_{jk}$ which implies variable utility is given by

$$\sum_{k=1}^{K} \sum_{j \in c} [(\bar{\mu}_{jk} + \nu^{w}_{jk}) 1(d_k = j)] q_k - 0.5 \sum_{k'=1}^{K} \sum_{k=1}^{K} \Lambda_{kk'} q_k q_{k'} \tag{59}$$

or, equivalently, in terms of store-category quantities (written q_{jk}, where $\Sigma_{j \in c} q_{jk} = q_k$):

$$\sum_{k=1}^{K} \sum_{j \in c} [(\bar{\mu}_{jk} + \nu^{w}_{jk}) q_{jk} - 0.5 \sum_{k'=1}^{K} \sum_{k=1}^{K} \Lambda_{kk'} (\Sigma_{j \in c} q_{jk})(\Sigma_{j \in c} q_{jk'}). \tag{60}$$

Maximization of (60) with respect to q_{jk} for all k and all $j \in c$ yields the same outcome as maximization of (59) with respect to $q = (q_1, \ldots, q_K)$ and $d = (d_1, .., d_K)$. Thus when the shopping choice has two stores $(n(c) = 2)$ the utility function implies that consumers use one store per category as an outcome of utility maximization (with respect to quantity q_{jk} in each store $j \in c$) not an outcome of a constraint on the maximization of utility. This in turn implies that the chosen quantities q_{jk} satisfy the Kuhn-Tucker conditions for the maximization of (60) with respect to q_{jk}, subject to $q_{jk} \geq 0$, for all k and all $j \in c$. We use these conditions in the next section.

G.2 Unobserved variable utility tastes implied by choice (c, d, q)

We now derive the implications of observed choice (c, d, q) and utility maximization for unobserved tastes ν^{w}_{jk} for $j \in c$.

When demand at store j for category k is positive $(q_{jk} > 0)$ the first order condition for maximization of equation (60) with respect to q_{jk} is satisfied. This implies that the taste shock ν^{w}_{jk} has a unique value $\bar{\nu}^{w}_{jk}$ given by

$$\bar{\nu}^{w}_{jk} = -\bar{\mu}_{jk} + \Lambda_{kk} (\Sigma_{j \in c} q_{jk}) + 0.5 \sum_{k' \neq k} \Lambda_{kk'} (\Sigma_{j \in c} q_{jk'}). \tag{61}$$

If alternatively the consumer's observed quantity choice for category k at store $j \in c$ is zero $(q_{jk} = 0)$ the first order condition is not satisfied but the Kuhn-Tucker conditions imply the derivative of utility (60) with respect to quantity q_{jk} is not positive which gives the following upper limit for the category-store taste shock

$$\nu^w_{jk} \leq -\bar{\mu}_{jk} + \Lambda_{kk}(\Sigma_{j\in c} q_{jk}) + 0.5 \sum_{k'\neq k} \Lambda_{kk'}(\Sigma_{j\in c} q_{jk'}). \tag{62}$$

The conditions (61) and (62) together give the following restrictions on unobserved tastes ν^w_{jk} for $j \in c$ that are consistent with the observed choices (c, q, d) and utility maximization:

$$\bar{\nu}^w_{jk} = -\bar{\mu}_{jk} + \Lambda_{kk}(\Sigma_{j\in c} q_{jk}) + 0.5 \textstyle\sum_{k'\neq k} \Lambda_{kk'}(\Sigma_{j\in c} q_{jk'}) \qquad \text{if } q_{jk} > 0$$

and

$$\nu^w_{jk} \in \mathcal{A}_{jk} \qquad \text{if } q_{jk} = 0$$

where $\mathcal{A}_{jk}$ is the set of values for ν^w_{jk} that are consistent with zero demand for (j, k), i.e.

$$\mathcal{A}_{jk} = \{\nu^w_{jk} | \ \nu^w_{jk} \leq (-\bar{\mu}_{jk} + \Lambda_{kk}(\Sigma_{j\in c} q_{jk}) + 0.5 \sum_{k'\neq k} \Lambda_{kk'}(\Sigma_{j\in c} q_{jk'}))\}. \tag{63}$$

Thus we have a unique point value for (j, k) with positive demand and a set $\mathcal{A}_{jk}$ of values for (j, k) with zero demand.

G.3 Likelihood when shopping choice c is exogenous

We now use the restrictions on tastes derived in the last section to derive a likelihood $L(q, d|\theta)$ for the observed choice (q, d) assuming that the consumer's shopping choice c is exogenous. Let

$$\nu^w_c = (\nu^w_{j1}, ..., \nu^w_{jK})_{j\in c}. \tag{64}$$

Suppose the consumer is observed to have positive demands for a number l of store-category (j, k) combinations and let $\nu^{(1)}_c$ denote the l-vector of unobserved tastes $\bar{\nu}^w_{jk}$ for these. Let $\nu^{(2)}_c$ be taste shocks for remaining (j, k) combinations, i.e. those with zero demand. The vector of category-store taste shocks ν^w_c in (64) can be written

$$\nu^w_c = (\nu^{(1)}_c, \nu^{(2)}_c).$$

Let the joint density be $f(\nu^{(1)}_c, \nu^{(2)}_c|\theta)$. The likelihood that a consumer selects (q, d) combines a probability density component for the unobserved taste elements in $\nu^{(1)}_c$ (which each have a unique value $\bar{\nu}^w_{jk}$ given utility maximization at parameters θ) and a probability mass component for the unobserved taste elements in $\nu^{(2)}_c$ (which each have a range of possible values $\mathcal{A}_{jk}$ given utility maximization at parameters θ). The likelihood is given by integrating the probability density f over the range of possible values for $\nu^{(2)}_c$ at the unique value of $\nu^{(1)}_c$ i.e.

$$L(q, d|\theta) = \int_{\mathcal{A}(c,q,d)} f(\bar{\nu}^{(1)}_c, \nu^{(2)}_c|\theta) \text{abs}[J] \ d\nu^{(2)}_c \tag{65}$$

where $\bar{\nu}_c^{(1)}$ is the vector of unobserved store-category taste shocks that satisfy the first order condition in (61) for all (j,k) with positive demand. Note that abs$[J]$ is the absolute value of the Jacobian for the transformation from $\bar{\nu}_{jk}^w$ to q_{jk} for all the errors in the vector $\bar{\nu}_c^{(1)}$. (From (61) the elements in matrix J are given by the second order utility terms in the quadratic utility, i.e.: $\partial\bar{\nu}_{jk}^w/\partial q_{jk'} = \Lambda_{kk'}$, etc.). Finally $\mathcal{A}(c,q,d)$ is the set of values for unobserved tastes $\nu_c^{(2)}$ that are consistent with utility maximization given choice (c,q,d) and is defined using (63) as follows

$$\mathcal{A}(c,q,d) = \times_{(jk)\in\{(jk)|\; q_{jk}\, =\, 0, j\, \in\, c\}}\mathcal{A}_{jk}$$

where $\times$ denotes the Cartesian product of the sets. The likelihood (65) is identical in form to equation (7) on page 234 in Kim et al. (2002) and equation (9) on page 266 in Wales and Woodland (1982).

G.4 Likelihood when shopping choice c is endogenous

We now derive the likelihood $L(c,q,d|\theta)$ for the observed shopping triple (c,d,q) allowing shopping choice c to be endogenous. Unlike the treatment in subsection G.3 we must now consider shopping costs Γ_c and the variable utility from stores $j \notin c$. The observed shopping choice c depends on the full set of consumer tastes ν defined as follows

$$\nu = \{(\nu_{j1}^w, ..., \nu_{jK}^w)_{j\in\mathcal{J}}, \nu^\Gamma\}. \tag{66}$$

Let the joint density of these be $f(\nu)$. The probability of observing shopping choice c given unobserved tastes ν is given by equation 18. Rewriting this in terms of ν we have

$$P_c(\nu|\theta) = \frac{\exp(w(c,p_c,\nu_c^w) + \Gamma_c(\nu^\Gamma))}{\sum_{c'\in\mathcal{C}}\exp(w(c',p_{c'},\nu_{c'}^w) + \Gamma_{c'}(\nu^\Gamma))} \tag{67}$$

where $\nu_{c'}^w$ are as defined in (64) for any $c' \in \mathcal{C}$.

We use the restrictions derived in section G.2 to determine the set of values of ν_c^w (for the chosen c) that are consistent with the observed choice (q,d) and the taste parameters. As in section G.3 we write $\nu_c^w = (\nu_c^{(1)}, \nu_c^{(2)})$ where $\nu_c^{(1)}$ denotes the vector of unobserved tastes ν_{jk}^w for (j,k) combinations (for $j \in c$) with positive demand and let $\nu_c^{(2)}$ be taste shocks for (j,k) combinations (for $j \in c$) with zero demand. As well as ν_c^w equation (66) includes ν_{jk}^w for $j \notin c$ and shopping cost shocks ν^Γ. We group these together as $\nu^{(3)} = ((\nu_{jk}^w)_{\forall jk|j\notin c}, \nu^\Gamma)$. The full vector of taste shocks ν in (66) is therefore

$$\nu = (\nu^{(1)}, \nu^{(2)}, \nu^{(3)}). \tag{68}$$

Let the joint density of these be $f(\nu^{(1)}, \nu^{(2)}, \nu^{(3)}|\theta)$.

The joint (discrete-continuous) likelihood that the consumer selects shopping choice c and category choices (q,d) is given by integrating $P_c(\nu|\theta)f(\nu)$ over the range of ν that are consistent with the consumer making a category choice (q,d) at c. This implies the restrictions for $\nu^{(1)}$ and

$\nu^{(2)}$ that we derived in subsection G.3. The likelihood is therefore

$$L(c, q, d|\theta) = \int_{\nu^{(3)}} \int_{\nu^{(2)} \in \mathcal{A}(c,q,d)} P_c(\bar{\nu}^{(1)}, \nu^{(2)}, \nu^{(3)}|\theta) f(\bar{\nu}^{(1)}, \nu^{(2)}, \nu^{(3)}|\theta) \text{abs}[J] d\nu^{(2)} d\nu^{(3)} \tag{69}$$

where $\bar{\nu}^{(1)}$, abs$[J]$, and $\mathcal{A}(c, q, d)$ are defined as in section G.3. This likelihood resembles the standard likelihood expression $\int P_c(\nu) f(\nu) d\nu$ for a mixed logit model for the probability of discrete choice c in the sense that we integrate the choice probability over the density for ν. The difference is that we do not integrate over all possible values of ν: we fix some of them (namely $\nu^{(1)}$) to the unique value ($\bar{\nu}^{(1)}$) that is implied by utility maximization given non-zero observed category demands and we restrict others (namely $\nu^{(2)}$) to the set of values $\mathcal{A}(c, q, d)$. Thus instead of obtaining the probability for a discrete choice c we obtain the probability expression for the discrete-continuous choice of (c, q, d). This likelihood extends those derived in the discrete-continuous literature (Dubin and McFadden (1984), Haneman (1984), Smith (2004)) to the case of multiple-continuous demands with corner solutions.

H Online Appendix: First-Order Condition for Prices

H.1 Profit maximization in terms of product prices

This subsection demonstrates that the first-order condition (50) can be derived from the assumption of profit maximization at the level of product prices. Let p_{fh} denote the price of product h in firm f, where each product belongs to some category k.[49] We express this as $h \in k$. The profit of firm f is $\pi_f(\bar{p})$, where $\bar{p} = (p_{fh})_{\forall f, \forall h}$, the vector of all product prices in all firms.

The usual first-order conditions for profit maximization by firm f are that

$$\tfrac{\partial \pi_f(\bar{p})}{\partial p_{fh}} = 0 \text{ for all } h. \tag{70}$$

Suppose that we can aggregate the firm's demand to the category level Q_{fk} and to write it as a function of category price indices p_{fk}. Category price indices p_{fk} are functions of the product prices p_{fh} so that we can write the function $p_{fk}(\bar{p}_{fk})$ where $\bar{p}_{fk} = (p_{fh})_{h \in k}$ is the subvector of $\bar{p}$ containing only the prices of products $h \in k$ owned by f. Then profit can be written

$$\pi_f(\bar{p}) = \Sigma_{k=1}^{K} Q_{fk}(p)(p_{fk} - mc_{fk}), \tag{71}$$

where $p = (p_{fk})_{\forall f, \forall k}$ is the vector of category-specific price indices and mc_{fk} is the marginal cost. (To simplify the notation, we assume $\chi_f = 1$ in this discussion. See Section 6.1.) Using (70) and

[49] Appendix C discusses the construction of price indices based on store-time specific product prices p_{jht}. Since we look at profit maximization at the weekly level, we suppress the t subscript in the current discussion.

(71) we arrive at a first-order condition in terms of the category price index p_{fk}:

$$
\begin{aligned}
0 &= \Sigma_{h\in k}\tfrac{\partial \pi_f(\bar{p})}{\partial p_{fh}} \\
&= \Sigma_{h\in k}\tfrac{\partial}{\partial p_{fk}}\left[\Sigma_{k'=1}^{K} Q_{fk'}(p)(p_{fk'}-mc_{fk'})\right]\tfrac{\partial p_{fk}(\bar{p}_{fk})}{\partial p_{fh}} \\
&= \left[Q_{fk}(p)+\Sigma_{k'=1}^{K}\tfrac{\partial Q_{fk'}(p)}{\partial p_{fk}}(p_{fk'}-mc_{fk'})\right]\Sigma_{h\in k}\tfrac{\partial p_{fk}(\bar{p}_{fk})}{\partial p_{fh}} \\
&= Q_{fk}(p)+\Sigma_{k'=1}^{K}\tfrac{\partial Q_{fk'}(p)}{\partial p_{fk}}(p_{fk'}-mc_{fk'}) \qquad (72)
\end{aligned}
$$

where the last line follows because $\Sigma_{h\in k}\frac{\partial p_{fk}(\bar{p}_{fk})}{\partial p_{fh}} \neq 0$. Reintroducing χ_f, dividing by $\partial Q_{fk}/\partial p_{fk}$ and p_{fk}, and rearranging, we get (50).

H.2 Consumer Group Specific Price Indices

This subsection demonstrates that the first-order condition (50) holds when we allow price indices to vary across consumer groups to reflect different purchasing patterns in households of different size and social class. For simplicity we use the general case of i-subscripts which allows for price indices for individual consumers. The price index is a weighted average of product prices (see Appendix C)

$$p_{ifk} = \sum_{h\in k} w_{ih} p_{fh} \qquad (73)$$

where $\sum_{h\in k} w_{ih} = 1$. To allow for a common shift to all product prices given by the scalar ρ_{fk} the price index can be written

$$p_{ifk} = \sum_{h\in k} w_{ih}(p_{fh}+\rho_{fk}) \qquad (74)$$

where $\rho_{fk}=0$ at equilibrium prices and

$$\Sigma_{h\in k}\frac{\partial p_{ifk}}{\partial p_{fh}} = \frac{\partial p_{ifk}}{\partial \rho_{fk}} = 1 \text{ for all } i. \qquad (75)$$

The i-specific category demands are $Q_{ifk}(p_i)$ where $p_i = (p_{ifk})_{\forall f,\forall k}$ is the vector of firm-category price indices. Profit is

$$\pi_f(\bar{p}) = \Sigma_i \Sigma_{k=1}^{K} Q_{ifk}(p_{ifk})(p_{ifk}-mc_{fk}). \qquad (76)$$

As we saw in the previous subsection, profit maximization implies

$$
\begin{aligned}
0 &= \Sigma_{h\in k}\tfrac{\partial \pi_f(\bar{p})}{\partial p_{fh}} \\
&= \Sigma_{h\in k}\Sigma_i\tfrac{\partial}{\partial p_{ifk}}\left[\Sigma_{k'=1}^{K} Q_{ifk}(p_i)(p_{ifk'}-mc_{fk'})\right]\tfrac{\partial p_{ifk}}{\partial p_{fh}} \\
&= \Sigma_i\tfrac{\partial}{\partial p_{ifk}}\left[\Sigma_{k'=1}^{K} Q_{ifk'}(p_i)(p_{ifk'}-mc_{fk'})\right]\Sigma_{h\in k}\tfrac{\partial p_{ifk}}{\partial p_{fh}} \\
&= \Sigma_i\left[Q_{ifk}+\Sigma_{k'=1}^{K}\tfrac{\partial Q_{ifk'}}{\partial p_{ifk}}\tfrac{\partial p_{ifk}}{\partial \rho_{fk}}(p_{ifk'}-mc_{fk'})\right] \\
&= Q_{fk}+\Sigma_{k'=1}^{K}\Sigma_i\tfrac{\partial Q_{ifk'}}{\partial \rho_{fk}}(p_{ifk'}-mc_{fk'})
\end{aligned}
$$

where the first line is from (70), the fourth uses (75), and $Q_{fk} = \Sigma_i Q_{ifk}$. Dividing by $\frac{\partial Q_{fk}}{\partial \rho_{fk}} = \Sigma_i \frac{\partial Q_{ifk'}}{\partial \rho_{fk}}$ we have

$$Q_{fk}\left(\frac{\partial Q_{fk}}{\partial \rho_{fk}}\right)^{-1} + (p_{fk} - mc_{fk}) + \frac{\Sigma_{k'} \frac{\partial \pi_{fk'}}{\partial \rho_{fk}}}{\frac{\partial Q_{fk}}{\partial \rho_{fk}}} = 0$$

where p_{fk}=$\Sigma_i \tilde{w}_{ifk} p_{ifk}$ in which $\tilde{w}_{ifk} = \frac{\partial Q_{ifk}}{\partial \rho_{fk}} \Big/ \frac{\partial Q_{fk}}{\partial \rho_{fk}}$. Note that $\frac{\partial p_{fk}}{\partial \rho_{fk}} = 1$ which allows us to replace $\frac{\partial Q_{fk}}{\partial \rho_{fk}}$ with $\frac{\partial Q_{fk}}{\partial p_{fk}}$. Rearranging, reintroducing χ_f, and dividing by p_{fk}, we get expression (50).

I Online Appendix: Profit Margin Calculations

In this Appendix we explain how we calculate the profit margin figures which are reported in Table 8 We begin with the calculations using firm-level data covering all grocery categories and then discuss the calculations using data specific to the the milk category which uses the same method.

The Competition Commission (CC) reports two profit margin figures that we use to derive profit margin estimates. The first figure is "gross retail margins" m_r defined as the difference between the retailer's annual total revenue and its annual total wholesale cost divided by annual revenue (using the supermarkets' accounts). The CC reports gross retail margins in the range $0.24 - 0.25$ depending on firm (CC(2000) Table 8.19). The second figure is "gross manufacturer margins" m_m defined as the difference between manufacturer revenues and supplier operating costs (excluding labor costs) as a proportion of manufacturer revenues. The CC reports gross manufacturer margins of 0.25 and 0.36 depending on the sample of firms used (CC(2000) Paragraph 11.108 and CC(2008) Appendix 9.3 Paragraph 11).

Let us begin by deriving a lower bound to the profit margins from these external data. To do this we assume double marginalilzation, i.e. assume that all payments to manufacturers are of the form of a marginal (or "linear") wholesale price and the retailer optimizes against this price plus its own marginal costs. Under this assumption the manufacturer's marginal costs are not relevant to the retailer when setting retail prices so that we can ignore the CC's information on the manufacturer's margins. If linear prices are used in relations between supermarkets and manufacturers (as double marginalization implies) then the gross retail margin m_r is equivalent to the retailers margin over wholesale prices. To obtain the lower bound to the profit margin we combine (i) the assumption of double marginalization, with (ii) the assumption that all of the retailer's labour costs are marginal, and (iii) the lower end of the range of the figures (noted above) from the CC for m_r (i.e. 0.24). The CC reports that the ratio of labour costs to wholesale price costs is 9:83 (see CC(2000), Paragraph 10.3) which implies labor costs are $\frac{9}{83}\% = 10.8\%$ of wholesale costs. This implies we should adjust the retail gross margins reported above using the formula $m = 1 - 1.108(1 - m_r)$ which gives 0.16. This is the lower bound figure presented in Table 8.

Now we derive an upper bound to profit margins using the external data. To do this we assume that there is efficient retail pricing so that the manufacturer's marginal cost is relevant to the retailer when setting prices. To obtain an upper bound to margins we combine (i) the assumption of efficient pricing, with (ii) the assumption that *none* of labour costs are marginal, and (iii) the *upper* end of the range of figures (noted above) from the CC for both m_r and m_m. With assumptions (i) and (ii) the overall vertical profit margin as a proportion of retail prices is given by the formula $m = m_r + (1 - m_r)m_m$ where m is the overall margin, m_r is retail margin and m_m is the manufacturer's margin. Assumption (iii) is that we use the higher of the gross margins figures from the CC for both retailers and manufacturers in this formula, i.s. $m_r = 0.25$ and $m_m = 0.36$. Together this gives the upper bound figure of $m = 0.52$ that appears in Table 8.

In the case of the milk category the CC reports gross retail margins in the range 0.28-0.30 and gross manufacturer margins in the range 0.04-0.05 (see CC (2008) Appendix 9.3, Paragraphs 12 and 15). Using the same method as in the previous two paragraphs these figures imply margin estimates for the milk category ranging from 0.20 (using $m = 1 - 1.108(1 - m_r)$ for $m_r = 0.28$) to 0.34 (using $m = m_r + (1 - m_r)m_m$ for $m_r = 0.30$ and $m_m = 0.05$).

The lower and upper bounds are conservative because it is likely that some intermediate proportion of labour costs is marginal and because where the CC present a range of figures for gross margins we have (under assumption (iii)) selected them to generate the widest bounds.

J Online Appendix: Alternative Specifications

Table 13: Estimated Parameters: Alternative Specifications

	Ind p_{jk}		Alt $dist_{ij}$		Alt $\Lambda_{kk'}$		Ind $\Lambda_{kk'}$	
	Est.	Std. Er.	Est.	Std. Er.	Est.	Std. Er.	Est.	Std. Er.
A: Store-category Taste Effects								
β_{01}	2.228	0.194	2.181	0.080	2.195	0.236	2.252	0.055
β_{02}	1.475	0.049	1.456	0.031	1.497	0.076	1.477	0.052
β_{03}	1.129	0.040	1.125	0.014	1.000	0.023	1.215	0.032
β_{04}	2.035	0.124	1.972	0.047	1.891	0.042	2.068	0.096
β_{05}	2.783	0.208	2.731	0.152	2.823	0.131	2.863	0.226
β_{06}	1.113	0.033	1.088	0.016	1.181	0.038	1.129	0.026
β_{07}	2.359	0.133	2.617	0.113	2.419	0.127	2.566	0.099
β_1	0.469	0.003	0.437	0.004	0.449	0.004	0.415	0.004
β_2	0.450	0.011	0.482	0.011	0.472	0.018	0.469	0.013
Scale of Taste Shocks (ν):								
σ_1	0.188	0.032	0.320	0.018	0.213	0.030	0.197	0.039
σ_2	0.977	0.022	1.046	0.027	0.894	0.033	0.973	0.026
σ_3	0.580	0.016	0.731	0.026	0.579	0.037	0.751	0.039
σ_4	1.186	0.023	1.091	0.024	1.096	0.029	1.077	0.020
σ_5	–	–	–	–	–	–	0.225	0.045
B: Second-Order Quadratic Parameters $\Lambda_{kk'}$								
Λ_{11}	22.254	6.828	21.116	3.305	20.153	3.019	21.542	2.248
Λ_{22}	11.851	0.655	12.093	0.375	11.840	3.499	11.996	0.746
Λ_{33}	3.919	0.286	4.035	0.182	3.601	0.198	4.330	0.218
Λ_{44}	11.485	1.087	11.276	0.412	10.214	0.420	11.651	0.866
Λ_{55}	16.851	2.558	16.808	1.164	16.554	1.060	17.219	2.049
Λ_{66}	4.315	0.204	4.348	0.144	4.489	0.241	4.369	0.204
Λ_{77}	9.105	0.902	10.677	0.674	9.127	0.792	10.094	0.653
Λ_{88}	13.633	0.392	14.427	0.568	15.572	0.775	14.395	1.407
Λ_{57}	1.879	0.121	2.032	0.081	–	–	2.070	0.115
Λ_{18}	1.558	0.083	1.444	0.182	–	–	1.107	0.605
Λ_{23}	0.286	2.473	0.201	0.681	–	–	0.382	1.269
Λ_{23}	0.394	1.709	0.870	0.998	–	–	0.554	0.346
Λ_{23}	0.061	1.030	0.113	0.152	–	–	0.160	0.656
Λ_{14}	–	–	–	–	0.717	0.426	–	–
Λ_{27}	–	–	–	–	-0.033	1.213	–	–
Λ_{38}	–	–	–	–	-0.321	0.172	–	–
λ_1	–	–	–	–	–	–	0.241	0.052
λ_2	–	–	–	–	–	–	-0.033	0.062
C: Price Parameters								
α_1	1.759	0.056	1.908	0.021	1.938	0.033	1.967	0.043
α_2	34.374	2.351	32.315	2.544	37.584	3.503	20.961	5.116
D: Shopping Costs								
γ_{11}	8.477	1.135	8.762	0.849	8.660	1.250	9.206	1.237
γ_{12}	0.447	0.028	0.436	0.026	0.457	0.029	0.402	0.024
γ_{21}	11.636	2.045	11.188	1.613	12.773	2.340	13.142	2.273
γ_{22}	-0.386	0.031	0.385	0.029	0.422	0.032	0.366	0.028

Notes: Parameters are estimated using 6000 consumer-week observations. Standard errors are corrected for simulation noise as detailed in Section 4. Time and firm-category fixed effects are not reported. The specifications are described in subsection 6.6.