An Empirical Model of Shopping Costs and Di¤erentiated Stores estimated with Micro Data*

Howard Smith
Department of Economics, Oxford University

February 18, 2003

Abstract

This paper estimates a model of supermarket choice where consumers make multiple store cisits in a goven shopping period. This is estimated using a monthly panel of over 5000 consumers' shopping decisions and a dataset of store characteristics. The model is capable of predicting the number of stores visited and the level of expenditure in several categories (fresh produce, grocery, medical, etc.) although in the present version we aggregate expenditure. The model may be used to assess the e¤ect of multi-store shoppers on …rms pricing and product slection decisions.

1. Introduction

THIS PAPER ESTIMATES a model of shopping choice incorporating several features of relevance to the grocery market: multi-store shopping, multi-category shopping, and shopping costs—i.e. costs of shopping at more than one outlet. These features of retail markets may in‡uence the pricing and product selection decisions of …rms (see Klemperer(1992)). Consequently, modelling multi-shop shoppers may result in better analysis of market power, dynamic pricing behaviour, and product selection decisions such as product range. Recent years have

*Very preliminary. Comments welcome. The paper represents an initial attempt to bring a model of shopping costs to the data. Some interesting aspects of the data such as categories of grocery expenditure, and persistence over time, remain to be incorporated into the modelling.

seen supermarkets extending product range outside traditional grocery items; the motive may be to reduce the prevalence of multi-store shopping which in turn may enhance the market power of the firms. The present paper represents an initial estimation of a simple model of shopping costs; later versions of the paper will develop the model. There is no counterfactual analysis in the present version; in later versions we aim to use the model to analyse the consequences of product range expansion and reductions in multi-store shopping.

The consumer demand model is based on quadratic utility. Utility is a quadratic function of the quantity of groceries at each store. Absent interaction effects, therefore, there is a benefit to shopping at more than a single store. Interaction effects are modelled, however, allowing the possibility that groceries at any two stores are perfect substitutes. The consumer has per-store costs of shopping. When choosing how many stores to visit the consumer trades off the variety-increasing effects of visiting an extra store with the shopping cost conequences of doing so. The level of demand for groceries at the chosen stores is derived from the quadratic utility model, which generates linear demands based on consumer and store characteristics.[1]

The model is estimated using three sources of data: a consumer panel giving shopping decisions, a panel of prices charged by stores, and a dataset of store characteristics. The consumer survey consists of a panel of shopping decisions from over 5500 households who record the exact outlets of shopping choice and the expenditure at each store on each of a number of categories (e.g. health, fresh food, alchololic beverages etc.). This is recorded for each month over a 20 month period. We have detailed information on the households' characteristics. The price data is for a basket of items for the main firms in the market over the 20 months. The store characteristics data gives location, store size, firm, and some other store characteristics.

There are several related papers. Klemperer (1992) develops a models of firms product selection decisions in the presence of shopping costs. Hendel (1999) estimates a model of multi-product choice in which consumers choose between alternatives involving combinations of several products. Nevo (2000) and BLP (1995) estimate charactersitics-baed models of discrete choice—although these differ from the present paper in that they use macro (as opposed to household-level) data, and study situations in which only one product is chosen.

The rest of the paper is organised as follows. In section 2 we present the data.

[1] This demand model may be extended to handle demand for groceries by category; the present implementation omits this.

In section 3 we outline the model. Section 4 presents the estimated results and section 5 concludes.

2. Data

The data are taken from A. C. Nielsen's *Homescan* survey which gathers data from households participating in the scheme. Households scan in their shopping purchases, recording items bought, price paid, and outlet of purchase. The grocery demand data is supplied monthly, aggregated into six categories—fruit and vegetables, health, grocery, meat etc. The period covers 20 months in 1998-2000. We have detailed household characteristics information.

Table 1 gives an example of the data: responses from two households for month 7. These are households A and B in panels A and B respectively. The top row of each panel gives household characteristics. As shown, household A has three children (in the age groups shown), two adults, a car, and income level $35-42k. The shopper is aged 39. For this household shopping is concentrated in only two stores. In contrast household B visits ten stores. Evidently, household B is either a low shopping cost household, or faces a diverse choice set that encourages multiple outlet shopping. The table shows how household expenditure is divided between the six di¤erent categories of expenditure.

Table 2 displays the shopping habits of the entire[2] household panel for the second month of the survey. There is considerable variation in the number of stores visited from just one up to a maximum of thirteen. That month only 11% of shoppers go to a single outlet but a majority go to less than …ve. Table 3 shows how expenditure and demand vary across outlets. Ranking outlets by amount spent per monthly period, the …rst has a mean expenditure of £112 (and a maximum of £498) while the second has a mean of £28 and a maximum of £192.

The same survey provides price data (at a company level) for 100 items by …rm and month. This data is presented in Table 3. The table suggests that variation over time within …rms clearly exceeds the variation across …rms; Chart 1 also presents this pattern.

Store characteristics data, from the *Institute of Grocery Distribution* (London, UK) are obtained for each store in the choice set and are presented in Table 4. For each store we have location, size, …rm, and parking.

[2] In this preliminary stage we exaine a regional subset of the country in which there are 2119 households.

3. Shopping Model

3.1. Utility Framework

We present a framework for the consumer's store-choice problem in a given time period. For initial analysis the model is limited in two ways: (i) we use a single aggregate commodity instead of the expenditure categories; and (ii) we limit analysis to the consumer's two highest-expenditure outlets. These limitations will be relaxed in later versions of the work.

A choice j is a combination of either one ($K_j = 1$) or two ($K_j = 2$) outlets. Household i's valuation of choice j be given by:

$$\begin{aligned} U_j^i &= u(x_j, z_j, \xi_{F(j)}, p_j, \tau_i, \theta) + \varepsilon_j^i \\ &= u_j^i + \varepsilon_j^i. \end{aligned}$$

where x_j is a K_j vector of (units of) grocery consumption for each of the K_j outlets chosen. Let $(z_j, \xi_{F(j)})$ represent observed and unobserved outlet characteristics for choice j. p_j is a K_j-vector of prices, and τ^i represents observed and unobserved household characteristics. $\theta = (\beta, \phi, \alpha)$ is model parameters (to be de...ned in the rest of this section). ε_j^i is idiosyncratic utility and u_j^i is "mean" utility.

We divide u_j^i into two parts: (1) utility from groceries consumed (the bracketed part in the following equation) and (2) utility from the shopping experience (the remaining terms):

$$u_j^i = \alpha_0^i \left[g(x_j, z_j, \xi_{F(j)}^i, \tau_i, \alpha) + y - p_j x_j \right] + h(K_j, z_j^s, \theta^i)$$

where α_0 is a scaling term, $g()$ is direct gross utility from x_j. α is parameters in gross utility, z_j^s is "shopping experience" related store characteristics (e.g. distance, store size, etc.), K_j is the number of stores visited and ϕ^i is household i's shopping costs.

3.2. Conditional Demand

We specify that direct utility function $g()$ is a quadratic form in terms of demand at each outlet $x_j = (x_1, ..., x_K)$. If only one outlet is visited ($K_j = 1$) then it is given by the simple quadratic form:

$$g() = \frac{1}{\alpha_1} \left\{ \gamma(\alpha, \tau^i, z_1) x_1 - \frac{1}{2} x_1^2. \right\}$$

where $\gamma()$ is a "taste shifter" which allows variation in the taste for groceries at the store in question to depend on store and household characteristics. Grocery utility, while increasing in $x_1,$ does so at a diminishing rate determined by parameter α_1. If two outlets are visited ($K_j = 2$) direct utility is given by the quadratic form:

$$g() = \sum_{k=1}^{K_j} \frac{1}{\alpha_1} \left\{ \gamma(\alpha, \tau^i, z_k) x_k - \frac{1}{2} x_k^2 \right\} - \frac{\mu}{\alpha_1} x_1 x_2$$

where μ is a "substitutability" parameter, with zero substitutability when $\mu = 0$.

Demand for groceries x_1 is given by maximising u_j^i with respect to x_1 giving

$$g_{x_1} - p_1 = 0.$$

If i chooses only a single outlet ($K_j = 1$) then this condition gives i's optimal demand x_1^i as:

$$x_1^i = \gamma(\alpha, z_1, \tau^i, \upsilon^i) - \alpha_1 p_1.$$

Where i chooses two stores ($K_j = 2$) demand x_k^i is (for $k = 1, 2$):

$$x_k^i = \gamma(\alpha, z_k, \tau^i, \upsilon_k^i) - \alpha_1 p_k - \mu x_{-k}^i \tag{1}$$

where x_{-k} is the demand for shopping in the other outlet.

In the above demand functions $\gamma()$ is the intercept. We use a simple linear form for $\gamma()$:

$$\gamma_k() \equiv \bar{\gamma}_j^i + \upsilon_k^i = \gamma^1 z_k^1 + \gamma^2 \tau^i + \upsilon_k^i$$

where υ_k^i is a consumer speci…c unobserved utility term and the $\gamma = (\gamma^1, \gamma^2)$ parameters are the e¤ect of outlet and household characteristics on demand.

Demand equations may be written:

$$\begin{aligned} x_1^i &= \bar{\gamma}_1 - \alpha_1 p_1 - \mu x_2^i + \upsilon_1^i \\ x_2^i &= \bar{\gamma}_2 - \alpha_0 p_2 - \mu x_1^i + \upsilon_2^i. \end{aligned}$$

As usual there is an endogeneity problem in which the unobserved utility is correlated with prices, requiring the use of instruments for prices. We estimate parameters assuming that the error terms $\upsilon_1^i, \upsilon_2^i$ are uncorrelated with prices of supermarket …rms not patronised by individual i. While prices of the …rms are correlated, there is no reason for the individual's utility from the chosen …rm's

products to be correlated with the price of …rms he is not patronising. This identi…cation assumption gives the moment conditions for estimation:

$$E\left(\upsilon_2^i|p^{-i}\right) = E\left(\upsilon_2^i|p^{-i}\right) = 0$$

where p^{-i} is the prices of …rms not patronized by individual i.

3.3. Store Choice

Utility from choice j is

$$u(j) = \alpha_0\left[g(z_j, \tau^i, \hat{\alpha}, \hat{\upsilon}^i) + y - p_j x_j\right] + z_j\beta^i - \phi^i(K_j - 1) + \upsilon_j^i$$

where parameters β^i and ϕ^i carry an i superscript as they depend on observed household characteristics τ^i as follows:

$$\begin{aligned} \phi_m^i &= \bar{\phi}_m + \phi_m^1\tau^i \\ \beta_m^i &= \bar{\beta}_m + \beta_m^1\tau^i \end{aligned}$$

where β_m^1 are parameters describing the e¤ect of observed household charactersitics on taste for the mth characteristic and ϕ_m^1 give the e¤ect of household characteristics on shopping costs.

We de…ne the unobserved part of utility for choice υ_j^i as follows:

$$\upsilon_j^i = \xi_{F(1)} + \varepsilon_1^i + (K_j - 1) * (\xi_{F(2)} + \varepsilon_2^i)$$

and $(\varepsilon_1^i, \varepsilon_2^i)$ is a pair of (independent) iid Type–1 EV deviates from each outlet ($k = 1$ and $k = 2$) and $\xi_{F(1)}$, $\xi_{F(2)}$ are estimated as *firm* dummies. Where only one outlet is visited ($K_j = 1$) the unobserved terms ε_2^i and $\xi_{F(2)}$ do not appear.

We estimate the choice model in a second stage, after the estimation of the parameters of the conditional demand system (in the manner described in the previous subsection). The gross utility from groceries $g(.)$ can be computed for any j-combinations using parameters and …tted unobserved utility terms $(\hat{\alpha}, \hat{\upsilon}^i)$ from the quadratic demand system described in the previous subsection.

To avoid a dimensionality problem arising from computing the utility of a large number of combinations of alternative outlet pairs, we simplify the estimation by *conditioning* on the choice of the “other” store. This requires us to assume that nothing unobserved in the choice “conditioned on” is correlated with an unobservable in the conditional model.

For each individual i, we observe he visits either one or two stores (everyone goes to at least one outlet per period). Initially let's suppose he visits two stores ($K = 2$). Randomly take one of his store choices and call it the "…rst" outlet a. Then, conditioning on that, the model gives the probability of the outlet chosen for his "second" store b. Conditioning on a to get utility of b give a we have (with constant conditioned-on terms from outlet a cancelled out):

$$u\left(b|a\right) = \alpha_0 \left[g(z_j, \tau^i, \hat{\alpha}, \hat{v}^i) + y - p_j x_j\right] + \phi^i + z_b\beta + \xi_{F(b)} + \varepsilon_b^i$$

where ϕ^i is the …xed disutility associated with visiting any store. If the shopper only visits one store we treat this as the "outside good option" for the second store, setting:

$$u\left(b|a\right) = 0.$$

Consequently, increases in shopping costs ϕ^i increase the probability of not patronizing a second outlet (in the given period). ϕ^i is identi…ed because some people go to only one store and others to two.

In similar fashion we can obtain the utility $u\left(a|b\right)$ of the "…rst" chosen outlet conditioned on the …rst (although here the ϕ^i term cancels out as everyone visits at least one outlet.

The model is estimated maximising a likelihood based on the conditional probabilities for each consumers observed (a^i, b^i):

$$\Pr(a^i|b^i) = \frac{\exp u(a|b)}{\Sigma_l \exp u(l|b)} \qquad \Pr(b^i|a^i) = \frac{\exp u(b|a)}{\Sigma_l \exp u(l|a)}$$

4. Results

The estimated results for the conditional demand model are presented in Table 5. Two speci…cations are estimated. A simple OLS model is reported in the …rst two columns. The IV model is presented in the next two. All parameters are signi…cantly di¤erent from zero. The instrumentation method used makes hardly any di¤erence to the estimated price parameter. The …rm dummy terms—in the …rst 15 rows—are signi…cant and their variation across …rms shows the importance of …rm as a determinant of expenditure. The signs of parameters on price and outlet size are as expected. Household size, car ownership and income have the expected signs. (Time dummies are not reported). The substitution parameter is about -0.4 suggesting that grocery products are not perfect substitutes (which would imply a value of -1).

The estimated results for the choice part of the estimation are shown in Table 6. The …rst 15 rows (f1...f15) are the …rm dummies which again are signi…cant and varied. The shopping cost parameter is identi…ed (as expected its sign is negative) and the e¤ect of income is to increase the magnitude of the shopping cost (in the income*shopping cost interaction). Store size has a positive parameter. If the model is interpreted strictly, this suggests that people value increases in store size intrinsically, and not just because of the improvement of the utility from the groceries that may be bought in larger stores (this latter e¤ect is picked up in the conditional demand functions). The distance variable has a negative parameter. Households with a car, however, have a lower cost of distance.

5. Conclusions

The paper estimates a model of consumer choice of supermarket outlet, endogenizing the levels of expenditure and the number of stores visited. The model is a …rst run on the data and can be extended in a number of directions.

References

[1] Klemperer, P. (1992) "Equilibrium Product Lines: Competing Head-to-Head may be Less Competitive" *American Economic Review* Vol 82 No 4 pp740-755.

[2] Berry S., Levinsohn J., and A Pakes (1995) "Automobile Prices in Market Equilibrium" *Econometrica* Vol 63, pp 841-890.

[3] Hendel, I., (1999) "Estimating Multiple Discrete Choice Models: An Application to Computerization Returns" *Review of Economic Studies* Vol 66. No 2. pp423–446.

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

Table 1: Panel A

H1: hsize=5 children=3 social=1 work<30hrs car use=1 children: 1 is 0-5, 2 are 5-10, age 39, $35-42k

		Expenditure ($)								
Week	Store ID	All	Veg	Meat	Liquor	Bakery bi	Grocery	Health	Store type	Firm
Per 07										
9817	1211	193.662	8.062	49.5465	0	17.9075	80.388	37.758	F	SAINSBURYS
9819	1121	17.342	0	7.7865	0	0.7975	2.3635	6.3945	F	MARKS & SPENCER
PER08										
9821	1211	228.071	5.771	69.2665	0	19.053	85.927	47.4005	F	SAINSBURYS
9821	1121	52.9975	0.6525	37.6565	0	5.6115	4.0455	2.871	F	MARKS & SPENCER
9823	1181	13.2965	0	2.465	0	2.146	8.6855	0	F	SAFEWAY
9822	2421	10.411	0	2.6825	0	3.77	3.9585	0	F	NORTH TAYSIDE S'MKT
9821		5.2925	0	0	0	5.2925	0	0	S	SPECIALIST

Table 1: Panel B

H2: hsize = 3, children=0, so			Expenditure ($)								
Week	Store ID	v5	All	Veg	Meat	Liquor	Bakery bi	Grocery	Health	Store Type	Firm
PER07											
9817	2421		90.76	3.25	35.12	0.00	6.25	38.54	3.26	F	NORTH TAYSIDE S'MKT
9817	1212		67.85	5.39	24.97	0.00	10.76	13.46	6.45	F	SAINSBURYS
9817			47.72	13.17	14.47	0.00	0.00	0.00	0.00	S	SPECIALIST
9817			15.21	0.00	10.80	0.00	0.67	0.00	2.31	C	CORNER STORES
9817			15.01	2.45	0.00	0.00	0.00	0.00	9.96	O	OTHER STORES
9817	1181		14.69	1.84	1.67	0.00	0.64	1.29	8.40	F	SAFEWAY
9819	1211		9.96	0.00	7.37	0.00	1.45	1.15	0.00	F	SAINSBURYS
9819	8016		8.87	3.23	0.00	0.00	2.76	1.29	1.60	F	TESCO METRO
9817	126		2.26	0.00	0.00	0.00	0.00	0.00	2.26	F	MARKS & SPENCER
9819	166		0.35	0.00	0.00	0.00	0.35	0.00	0.00	F	SAFEWAY
Per 08											
9822	1212		88.41	11.31	42.82	0.00	7.89	13.15	11.47	F	SAINSBURYS
9821	2421		63.64	8.21	36.08	0.00	1.15	14.95	3.26	F	NORTH TAYSIDE S'MKT
9824			60.96	0.00	2.44	15.23	0.00	2.16	2.16	O	OTHER STORES
9821	1181		39.15	4.36	6.44	0.00	0.00	5.61	22.74	F	SAFEWAY
9824			32.94	24.65	1.86	0.00	3.19	0.00	0.00	S	SPECIALIST
9824	76		10.12	2.84	1.44	0.00	1.04	2.45	2.35	F	GATEWAY
9821	1211		4.73	1.02	0.00	0.00	0.29	0.00	3.42	F	SAINSBURYS
9821	126		1.13	0.00	0.00	0.00	0.00	0.00	1.13	F	MARKS & SPENCER
9824			1.00	0.00	0.00	0.00	1.00	0.00	0.00	C	CORNER STORES

Table 2
Occurrence of Multiple Shopping Choices over a Monthly Period
Data from second month of the survey

# stores	All Shopping			Shopping over £5			Shopping at chains		
1	247	11.7	11.7	247	11.7	11.7	247	12.4	12.4
2	363	17.1	28.8	537	25.4	37.0	873	43.7	56.1
3	401	18.9	47.7	518	24.5	61.5	354	17.7	73.8
4	393	18.5	66.3	378	17.9	79.4	239	12.0	85.8
5	265	12.5	78.8	220	10.4	89.7	139	7.0	92.7
6	192	9.1	87.8	120	5.7	95.4	71	3.6	96.3
7	120	5.7	93.5	52	2.5	97.9	46	2.3	98.6
8	65	3.1	96.6	24	1.1	99.0	11	0.6	99.1
9	36	1.7	98.3	13	0.6	99.6	10	0.5	99.6
10	23	1.1	99.3	4	0.2	99.8	4	0.2	99.8
11	8	0.4	99.7	3	0.1	100.0	2	0.1	99.9
12	5	0.2	100.0	1	0.0	100.0	1	0.1	100.0
13	1	0.0	100.0	0	0.0	100.0	0	0.0	100.0
hh:	2119			2117			1997		

Table 3

Price of a Shopping Basket (£)				
	Mean	**Min**	**Max**	**SDEV**
ASDA Wa	24.22	26.30	0.62	5.31
Morrison	24.05	26.34	0.80	5.27
Safeway	24.34	27.27	1.37	5.38
Somerfiel	24.64	27.13	0.85	5.40
Sainsbury	24.39	27.18	0.86	5.35
Tesco	24.86	27.34	0.73	5.44

#of items in basket=24

Table 4

	Northern				Southern			
	#	mean size	sd size	mean park	#	mean size	sd size	mean park
Aldi	47	8024	351	1	15	8109	309	1.00
ASDA/Walmart	46	41259	11947	0.98	33	44546	9775	1.00
Budgen	0				67	6468	2809	0.00
Coop	350	4953	6237	0.35	278	4629	6789	0.26
Iceland	99	4542	974	0.14	195	4820	1175	0.28
KS	184	8124	4222	0.94	70	7271	5468	0.79
Lidl	24	8000			35	8000		
M&S	37	43089	38521	0.05	104	39730	32497	0.10
Morrison	50	35030	7637	0.90	2	38025	5342	1.00
Netto	62	6825	1802	0.48	16	7599	821	0.44
Safeway	98	17622	9583	0.82	91	22119	8684	0.86
Sainsbury	35	37999	20632	0.97	168	29192	16542	0.98
Somerfield	44	12228	4421	0.95	130	8803	6168	0.88
TESCO	58	21974	12073	0.76	179	26631	16921	0.58
Waitrose	0				79	15440	4025	0.7
	1134				1462			
Total stores:	2596							

Table 5--Parameters in Conditional Demand

Parameters: Conditional Demand	OLS		IV	
	Estimate	SE	Estimate	SE
f0	0.044	0.049	0.044	0.070
f1	0.499	0.080	0.575	0.124
f2	4.627	0.273	4.598	0.375
f3	1.476	0.092	1.386	0.146
f4	1.093	0.052	1.124	0.077
f5	0.622	0.053	0.631	0.078
f6	1.037	0.052	1.090	0.077
f7	0.785	0.061	0.813	0.091
f8	0.178	0.061	0.150	0.090
f9	4.205	0.271	4.226	0.373
f10	1.012	0.055	1.063	0.082
f11	4.507	0.275	4.508	0.378
f12	4.955	0.276	4.920	0.379
f13	3.959	0.275	3.990	0.378
f14	5.015	0.280	4.960	0.384
f15	1.836	0.058	1.801	0.087
price	-0.1079	0.011	-0.108	0.015
store size	0.140	0.005	0.150	0.007
sigma	-0.397	0.003	-0.409	0.003
hsize	0.701	0.007	0.729	0.010
income	0.031	0.001	0.030	0.001
car	0.326	0.024	0.327	0.035
	R=0.34	85400	R=0.34	85400

(time dummies not reported)

Table 6:
Parameters: Choice Model

Parameter	Estimate	S.E.
f1	-3.626	0.043
f2	-0.204	0.020
f3	-4.374	0.055
f4	-3.314	0.019
f5	-2.968	0.018
f6	-3.095	0.018
f7	-2.729	0.028
f8	-3.502	0.028
f9	-0.432	0.022
f10	-2.609	0.022
f11	-0.885	0.018
f12	-0.558	0.017
f13	-1.396	0.022
f14	-0.417	0.017
f15	-3.645	0.028
scaling	0.025	0.000
shopping cost	-0.618	0.030
income*shopping cost	-0.054	0.001
size	0.083	0.002
dist	-0.161	0.003
car*dist	0.155	0.003
	#obs=	42700