Technical-Appendix (for manuscript 2102.5)

Effect of Product Assortment Changes on Customer Retention

This appendix provides further details specifically on the category level analysis discussed in the above article.

1. Graphical Representation of the Store Level Data

Figure 1: Inter-delivery times across households

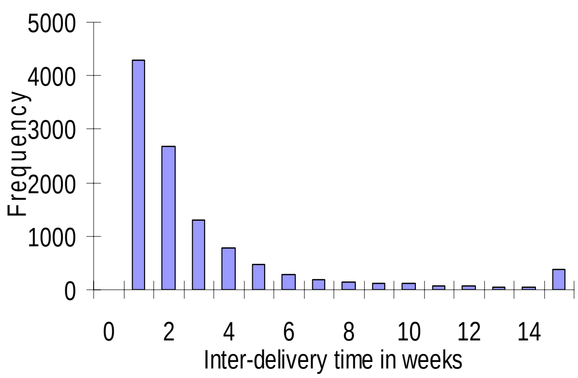


Figure 2: Purchase amounts across households

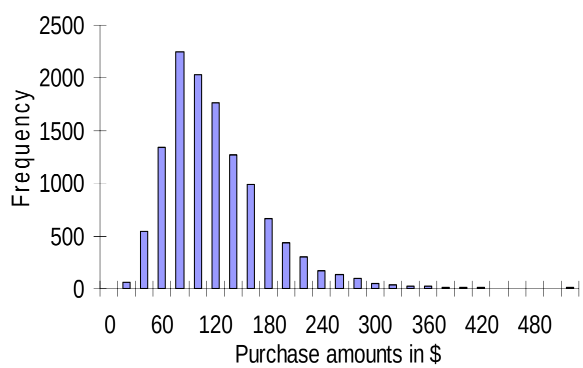


Figure 1 is a histogram of the inter-delivery times for and Figure 2 is a histogram plot of the purchase amounts (of the entire shopping basket) for each of the 1,218 households. The inter-delivery times range from one week to about 51 weeks. More than 92% of these times are less than or equal to 8 weeks, and about 3.5% lie beyond 14 weeks. Purchase amount observations span from little more than a dollar up to $675, though the bulk of purchases (about 95%) lie within the $40 to $220 band. Figures 1 & 2 also illustrate the vast degree of customer heterogeneity. While a large portion of the inter-delivery times is less than or equal to 1 week

(38.9%), a substantial portion is greater than 3 weeks (25.0%). Also, the average purchase amount per household varies from about $16 to $538.

2. The Category Level Data

We observe purchases in 147 product categories[1] with the average number of categories purchased per occasion being 20 (the median and mode being 19 and 18 respectively). It is interesting to note that earlier studies in the context of the brick-and-mortar grocery stores reported similar figures; Spiggle (1987) reports that on an average there are 22 categories on the shopping lists of consumers (the mode being 18). For our category level analysis we selected the top 75 categories in terms of number of purchases and combined all other categories into one (naming it '*all others*', a 76th category). The top 75 categories account for more that 90% of purchases, both in terms of number of purchases and the dollars involved.

Table 1 lists average dollar amounts (per household per purchase occasion) and total dollar amounts (in the span of the dataset) spent in each of the 76 categories, the percentage cuts in items affected in various categories, and the percentage revenue (prior to the cuts) accounted for by the items that were removed from each category. The average amounts range from $1.78 to $15.99, the largest being in diapers, baby food, (fresh) meat, and the (fresh) produce categories. The produce category is also the highest revenue earner, with close to 10% of the total revenue observed in the dataset. Surprisingly, the category 'all others' stands next in terms of overall revenues (8.2%). However, no category in 'all others' except for cigarettes (0.83%) accounts for more than half a percentage point (the next highest being 'diet aids' at 0.33%). This

[1] The category classification used is as defined by the online grocery service but does not differ significantly from that of Information Resources Inc (IRI).

illustrates the wide spreads in grocery revenues by category, or simply that most categories are only a small portion of the whole store. Table 1 also lists the percentage of items cut in each category. These cuts range from 24% to 91%, the average across categories being 57%. Apparently, managers followed a crude heuristic of eliminating slow moving/low revenue items, for while on average 57% of items were eliminated, they accounted for a mere 20.2% of the revenues.

Table 1:The product categories, total & average expenditures and extent of assortment reductions

Cat No.	Category Description	Total Sales in $ (A)	% (A)	Average Amount per purchase occasion ($)	% item Reduction in a category	% revenue deleted items
1	MARKET PRODUCE	139093	10.1%	13.10	36%	4.7%
2	DAIRY MILK/DRINKS/HALF & HALF/CREAM	35950	2.60%	3.83	31%	2.3%
3	GROC BREAD PACKAGED	34933	2.53%	4.12	47%	11.6%
4	GROC CEREAL/BREAKFAST FOODS/GRANOLA SNACKS	60746	4.39%	8.09	45%	10.2%
5	MARKET MEAT	112756	8.15%	15.59	43%	7.2%
6	GROC SODA/BOTTLED TEA	50188	3.63%	7.21	55%	13.5%
7	MARKET DELI	48400	3.50%	7.10	52%	5.6%
8	DAIRY CHEESE PRE-PACKAGED	26550	1.92%	4.44	58%	28.3%
9	DAIRY JUICE/DRINK	26138	1.89%	4.52	48%	8.8%
10	GROC JUICE/DRINK NONREFRIGERATED	32960	2.38%	5.91	72%	31.9%
11	DAIRY YOGURT/PUDDING/GELATIN	22296	1.61%	4.17	39%	15.5%
12	GROC SNACKS/DIPS	26670	1.93%	5.11	64%	19.5%
13	FROZEN ICE CREAM/NOVELTIES/YOGURT/SHERBET	26250	1.90%	5.73	73%	39.5%
14	DAIRY EGGS/SUBSTITUTES	8110	0.59%	1.78	55%	6.4%
15	DAIRY MARGARINE/BUTTER	10577	0.76%	2.40	52%	16.8%
16	GROC CRACKERS/PROCESSED CHEESE	22422	1.62%	5.22	58%	19.5%
17	HH PAPER TOWELS	18510	1.34%	4.31	38%	6.6%
18	HH TOILET PAPER	15827	1.14%	3.73	44%	8.4%
19	GROC WATER	17735	1.28%	4.56	43%	5.3%
20	MARKET BAKERY (FRESH)	16217	1.17%	4.10	55%	10.1%
21	GROC COOKIES	19192	1.39%	4.92	75%	31.1%
22	GROC SOUP CANNED & MIXES	16223	1.17%	4.38	54%	18.5%
23	HH LAUNDRY PRODUCTS	28941	2.09%	7.80	57%	25.8%
24	GROC PASTA DRIED	8891	0.64%	2.45	59%	22.9%
25	FROZEN VEGETABLES	13453	0.97%	3.77	46%	24.8%
26	GROC SPAGHETTI SAUCE/GRATED CHEESE	12525	0.91%	3.96	62%	27.5%
27	PET FOOD & PRODUCTS	29848	2.16%	9.19	52%	31.2%
28	GROC CANNED VEGETABLES/TOMATOES/BEANS	8857	0.64%	2.83	51%	20.5%
29	FROZEN BREAKFAST	10250	0.74%	3.31	48%	27.0%
30	HH BAGS	15663	1.13%	5.22	42%	14.8%
31	GROC PREPARED FOODS BEANS/PASTA/MEAT/SAUCES	9769	0.71%	3.30	64%	32.3%
32	DAIRY COTTAGE CHEESE/SOUR CREAM/DIPS	6792	0.49%	2.34	53%	18.0%
33	FROZEN MEALS	27220	1.97%	9.70	52%	18.8%
34	HH PAPER PLATES/CUPS/NAPKINS	10506	0.76%	3.69	53%	17.5%
35	GROC CONDIMENTS/MEAT SAUCES/MARINADES	7595	0.55%	2.83	69%	22.7%
36	HH DISHWASHING PRODUCTS	8601	0.62%	3.28	38%	15.4%
37	HH CLEANERS BATH/KITCHEN/GLASS	11722	0.85%	4.49	55%	22.8%
38	HH FACIAL TISSUES	8378	0.61%	3.45	44%	5.1%
39	GROC CANNED SEAFOOD	8057	0.58%	3.58	58%	9.6%

40	GROC COFFEE	15011	1.09%	6.92	70%	34.2%
41	GROC ETHNIC MEXICAN	6749	0.49%	3.28	69%	22.1%
42	GROC RICE/DRIED BEANS	5497	0.40%	2.91	65%	28.3%
43	HBA SOAP BAR/LIQUID	7316	0.53%	3.97	57%	13.8%
44	GROC HOT DOGS/SAUSAGE PRE-PACKAGED	7715	0.56%	4.37	61%	27.6%
45	GROC CANDY/GUM/MARSHMALLOWS	8067	0.58%	4.52	79%	33.6%
46	GROC SUGAR	4755	0.34%	2.78	38%	9.1%
47	FROZEN JUICE/DRINK	7645	0.55%	4.43	24%	3.7%
48	HBA ORAL HYGIENE	8652	0.63%	5.06	80%	37.9%
49	GROC FRUIT DRIED/SNACKS	6201	0.45%	3.62	46%	11.9%
50	MARKET SEAFOOD	14833	1.07%	8.56	58%	9.6%
51	GROC CANNED FRUIT	5117	0.37%	3.04	60%	23.5%
52	BABY DIAPERS	26325	1.90%	15.99	73%	44.2%
53	GROC SALAD DRESSINGS & TOPPINGS	5030	0.36%	3.11	83%	58.5%
54	GROC JAM/JELLIES/SPREADS/HONEY	5134	0.37%	3.20	57%	27.4%
55	GROC OLIVES/PICKLES/PEPPERS/RELISH	4413	0.32%	2.91	74%	35.0%
56	BABY ACCESSORIES	8229	0.59%	5.79	70%	18.4%
57	GROC PANCAKE/WAFFLE MIX/SYRUP	4740	0.34%	3.56	57%	25.1%
58	GROC BAKERY/SNACKS PRE-PACKAGED	5659	0.41%	4.25	68%	27.9%
59	GROC SALT/SEASONING/SPICE/SAUCE MIX	5404	0.39%	4.20	84%	62.3%
60	GROC DELI PREPACKAGED	5970	0.43%	4.43	70%	31.3%
61	BABY FOOD/FORMULA	20489	1.48%	15.51	41%	14.1%
62	GROC PEANUT BUTTER	3383	0.24%	2.57	50%	38.1%
63	GROC OIL COOKING	6249	0.45%	4.84	63%	33.0%
64	GROC BACON/BREAKFAST SAUSAGE	5536	0.40%	4.41	46%	6.8%
65	HBA HAIR PRODUCTS	6711	0.49%	5.38	91%	70.3%
66	FROZEN MEAT	8295	0.60%	6.91	60%	22.6%
67	GROC BAKING NEEDS/CHOC/NUTS	4069	0.29%	3.65	78%	37.5%
68	GROC APPLESAUCE	2393	0.17%	2.07	54%	12.5%
69	FROZEN PIZZA	5644	0.41%	4.85	72%	31.2%
70	GROC TEA	4337	0.31%	3.89	69%	24.7%
71	DAIRY DOUGH/BREAD/BISCUITS	3673	0.27%	3.51	52%	23.5%
72	HH ALUMINUM FOIL/WAX PAPER/FOOD WRAP	3468	0.25%	3.38	30%	13.7%
73	FROZEN BAKERY/WHIPPED TOPPINGS	3245	0.23%	3.31	69%	32.2%
74	FROZEN POTATOES/ONION RINGS	2637	0.19%	2.74	71%	23.3%
75	DAIRY PASTA/SAUCES/TOPPINGS	6913	0.50%	7.52	51%	16.7%
76	All OTHERS	112812	8.16%	5.15	82%	50.4%
	T O T A L	**1383123**	**100%**	**5.50**	**57%**	**20.0%**

3 Category level models

As mentioned earlier, we observe Q_{ht} (the entire shopping basket amount) in dollars, where *h* indexes households and *t* indexes observations for a household. We also observe the composition of this shopping basket in terms of amount spent on different items across product categories. Again, following an approach similar to that of the store level models, we build a joint model of category incidence and category amounts by considering (1) the marginal distribution of the purchase probabilities across categories and (2) the distribution of the overall shopping basket amount across categories, conditional on a purchase being made in the category.

The impact of assortment reductions at the category level is modeled by introducing two assortment variables $ASSORTF_h^c$ and $ASSORTNF_h^c$. The first is the extent of reduction in assortment from amongst the household *h*'s favorite items in product category *c*, and the second is the reduction from among non-favorite items[2]. These variables are defined over the interval [0,1] with '0' indicating no cuts in assortment and '1' indicating a 100% cut in assortment. By definition, if a household did not face an assortment cut, these variables take a value 0. These variables provide a simple yet useful operationalization of the assortment cut across the favorite/non favorite dimensions.

The other covariates we use for the category level models are the quarterly dummies $QTR1_{ht}^c, QTR2_{ht}^c$ and $QTR3_{ht}^c$ for the first, second and third quarters respectively, the natural logarithm of one lag of the purchase amount in a category, Q_{ht-1}^c, and the natural logarithm of the time elapsed (in days) since the last purchase in that category, $TIME_{ht}^c$.

[2] 'Favorite item' is defined as the set of all items purchased by a particular household prior to assortment reduction. Similarly, 'non favorite item' is the set of all available items that were not purchased by the household prior to assortment reduction.

3.1 Category incidence model

The probability π_{ht}^{c} that a household *h* on purchase occasion *t* buys into a particular category *c* is given by the following logit transformation,

$$\pi_{ht}^{c} = \frac{\exp(\omega_{ht}^{c})}{1+\exp(\omega_{ht}^{c})}$$

$$c = 1,2,\ldots,C \text{ (the total number of categories)} \tag{1}$$

where ω_{ht}^{c} is further specified as,

$$\begin{aligned}\omega_{ht}^{c} = {} & \omega_{h}^{c} + \omega_{1}^{c}Q_{ht-1}^{c} + \omega_{2}^{c}TIME_{ht}^{c} + \omega_{3}^{c}ASSORTF_{h}^{c} + \omega_{4}^{c}ASSORTNF_{h}^{c} \\ & + \omega_{5}^{c}QTR1_{ht}^{c} + \omega_{6}^{c}QTR2_{ht}^{c} + \omega_{7}^{c}QTR3_{ht}^{c}\end{aligned} \tag{2}$$

the variables are as defined earlier and $\omega_{h}^{c}, \omega_{1}^{c}, \omega_{2}^{c}, \omega_{3}^{c}, \omega_{4}^{c}, \omega_{5}^{c}, \omega_{6}^{c}, \omega_{7}^{c}$ are the parameters to be estimated. We assume that the decision to buy in a category by a household is independent of other categories. Thus we have *C* independent category incidence models, where *C* = 76 is the total number of categories. The parameters $\omega_{3}^{c}, \omega_{4}^{c}$ specify the effect of assortment cuts on the purchase probabilities across various product categories, where ω_{3}^{c} is the effect of reduction in favorite items on the probability of purchase in category *c*, while ω_{4}^{c} is the effect of reduction in non-favorite items. If the cuts in non-favorite items were seen as ‘clutter reduction,’ in turn easing the purchase process for the consumer, we would expect purchase probabilities to go up with a reduction in non-favorite items, i.e. $\omega_{4}^{c} > 0$, because we know that too much choice can be demotivating (Iyengar and Lepper 2000). However, if reduction in non-favorite items led to the assortment being viewed as too ‘simplistic’ and offering ‘limited variety,’ we would expect reduced purchase probabilities with increased reductions ($\omega_{4}^{c} < 0$). The result may depend on the nature of the categories involved. For example there may be some categories (such as ‘candies’) where the consumers inherently desire a large assortment, even though some of the items in the category are never purchased. Similarly there may be some other categories, such as analgesics, where consumers are known not to desire a large variety. On the other hand, we would expect a reduction in purchase probabilities across categories when one’s favorite items are removed, i.e., for a majority of the categories we would expect $\omega_3^c < 0$.

The parameter ω_h^c in equation (2) can be viewed as specifying the baseline purchase probability for each household for category *c*. We specify a normal heterogeneity distribution for this parameter,

$$\omega_h^c \sim \ Normal(\overline{\varpi}^c, \xi^{2^c}) \tag{3}$$

where $\overline{\varpi}^c$ & ξ^{2^c} are the mean and variance of the heterogeneity distribution.

3.2 Category purchase amount model

Category purchase amounts are modeled conditional on the overall shopping basket amounts; thus, we model the spread of the entire shopping basket amount [which in turn is specified by equations (6), (7) and (8)] across the categories purchased. We assume this spread to be distributed according to a Dirichlet distribution. Thus we have a Dirichlet distribution for each purchase occasion *t* (for each household *h*) with the number of Dirichlet parameters equal to the number of categories purchased by that particular household on that purchase occasion. Expressed in equation form,

$$\mathbf{SHARE}_{ht}^{N_{ht}} \sim \ Dirichlet\left(N_{ht}, \eta_{ht}\right) \tag{4}$$

where N_{ht} is the number of product categories bought by household h on purchase occasion t, and $\mathbf{SHARE}_{ht}^{N_{ht}}$ is a N_{ht} dimension vector of the ratio of category dollar amount spent to the overall shopping basket amount by household h on purchase occasion t. The vector $\boldsymbol{\eta}_{ht} = (\eta_{ht}^{c=1}, \eta_{ht}^{c=2}, \ldots, \eta_{ht}^{c=C})$, where η_{ht}^{c} is proportional to category c's share of the total shopping basket for household h purchase occasion t when category c is purchased. The relative magnitudes of these parameters indicate the relative shares of the respective categories in the overall shopping basket whenever the categories co-occur. These parameters vary across households as well as observations and are further specified as follows,

$$\eta_{ht}^{c} = \exp\begin{pmatrix} \eta_h^c + \eta_1^c Q_{ht-1}^c + \eta_2^c TIME_{ht}^c + \eta_3^c ASSORTF_h^c + \eta_4^c ASSORTNF_h^c \\ + \eta_5^c QTR1_{ht}^c + \eta_6^c QTR2_{ht}^c + \eta_7^c QTR3_{ht}^c \end{pmatrix}$$
$$c = 1,2,\ldots,C \tag{5}$$

where the variables are as specified earlier and $\eta_h^c, \eta_1^c, \eta_2^c, \eta_3^c, \eta_4^c, \eta_5^c, \eta_6^c, \eta_7^c$ are the parameters to be estimated.

Analogous to the category incidence model (section 2.2), the parameters η_3^c, η_4^c specify the effect of assortment cuts on the 'share-of-the-basket' across various product categories: η_3^c is the effect of reduction in favorite items on the share of category c in the overall shopping basket, while η_4^c is the effect of reduction in non-favorite items on the share of category c in the overall shopping basket. Values of $\eta_4^c > 0$ indicate that reduction in non-favorite items in category c leads to a larger share-of-the-basket for that category, and values $\eta_4^c < 0$ imply a reduced share-of-the-basket. Similarly values of $\eta_3^c > 0$ and $\eta_3^c < 0$ imply respective increases and decreases in the share-of-the-basket with cuts in the favorite items in a category.

The parameter η_h^c in equation (5) can be viewed as specifying the baseline share-of-the-basket for each household for category *c*. A higher value of η_h^c for a category implies that the category has a higher baseline share-of-the-basket whenever the category is part of the shopping basket. We specify a normal heterogeneity distribution for this parameter,

$$\eta_h^c \sim \ Normal(\overline{\eta}^c, \rho^{2^c}) \tag{6}$$

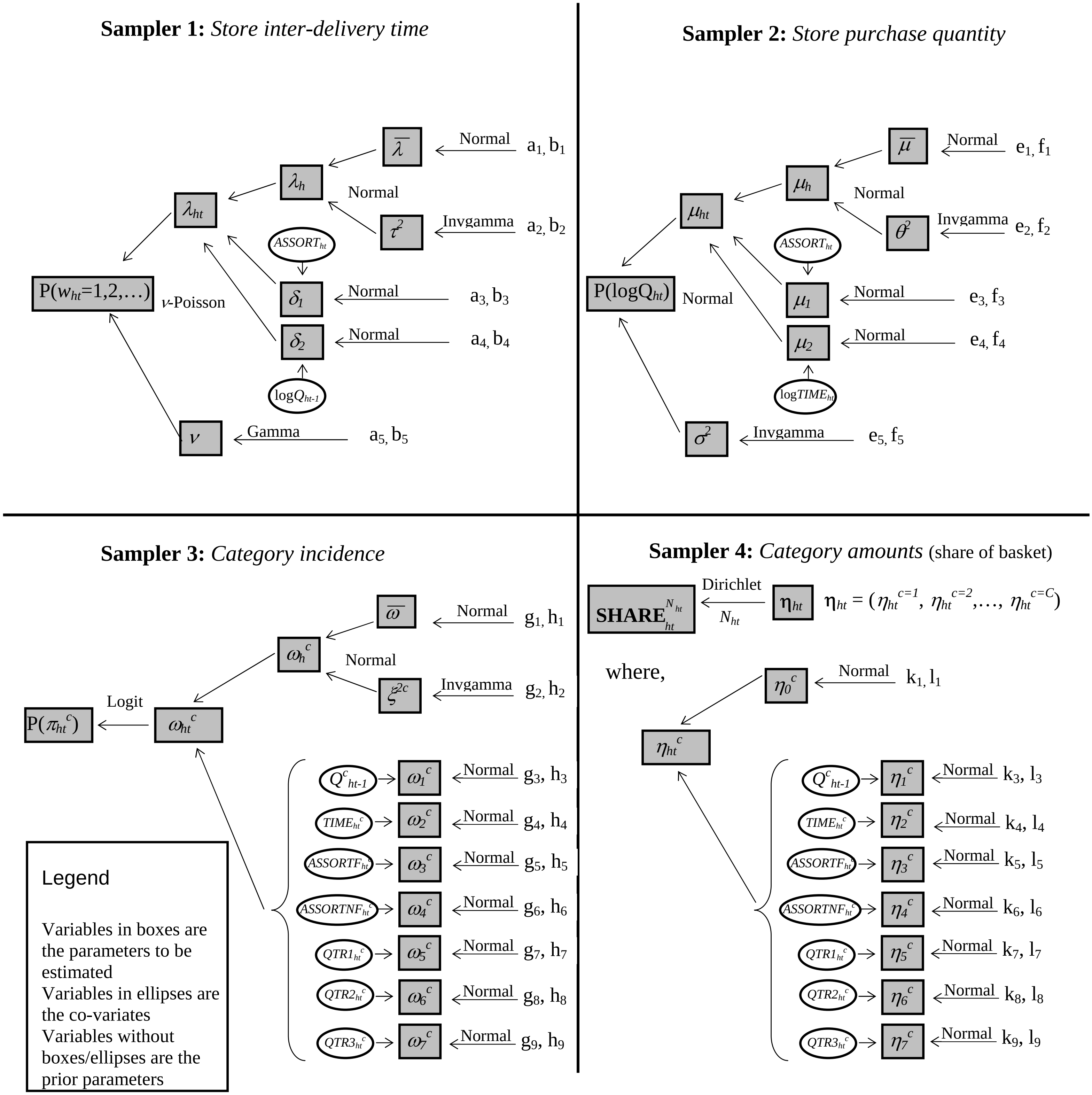
where $\overline{\eta}^c$ & ρ^{2^c} are the mean and variance of the heterogeneity distribution.

The two category level models (category incidence and category purchase amount models) provide interesting opportunities to study the impact of assortment reductions on the category purchase behavior. For example, assortment reductions may cause some categories simultaneously to have lower purchase probabilities and higher shares-of-the-basket. This outcome indicates that when faced with less choice, though the households buy into those categories less often (reduced probability of purchase), they devote a larger share of their overall expenditure to the category when they do make a purchase.

4 Prior specification and the full conditionals

There are four models to be estimated, the two store level models and the two category level models. We use a hierarchical Bayes approach using four MCMC samplers (Casella and George, 1992; Gelfand and Smith, 1990) to estimate these models. Figure 3 is a schematic representation of all the four samplers and the following subsections lay out the prior parameter specifications associated with the two category level models and the full conditional distributions used in the estimation algorithm.

Figure 3 A schematic representation of the four MCMC samplers



4.1 Prior specifications

Table 2: The priors used in the estimation of the two category models

Parameter	Priors
$\overline{\varpi}^{c}$	Normal(0,100)
ξ^{2c}	Inverse gamma(2.5,0.5)
$\omega_1^c, \omega_2^c, \omega_3^c, \omega_4^c$	Normal(0,100)
$\omega_5^c, \omega_6^c, \omega_7^c$	Normal(0,10)
$\eta_0^c, \eta_1^c, \eta_2^c, \eta_3^c, \eta_4^c$	Normal(0,100)
$\eta_5^c, \eta_6^c, \eta_7^c$	Normal(0,10)

Table 2 provides the priors used in the estimation of the two category level models. The first category level model (the category incidence model, section 3.1), corresponding to *Sampler 3*, is given by equations (1), (2) and (3). We need to specify prior distributions for the parameters $\overline{\varpi}^{c}$ & ξ^{2c} (equation 2) and $\omega_1^c, \omega_2^c, \omega_3^c, \omega_4^c, \omega_5^c, \omega_6^c, \omega_7^c$ (equation 2), c = 1,2,…,C (the number of categories). The $\overline{\varpi}^{c}$ & ξ^{2c} parameters specify the heterogeneity distribution over the ω_h^c's, which can be viewed as specifying the baseline category purchase probability for each household for category *c*. We employ conjugate priors for these parameters and, keeping in line with our lack of sufficient prior information on this heterogeneity distribution, we use relatively 'diffuse' priors [normal(0,100) and inverse gamma(2.5,0.5) respectively].

The parameters ω_1^c & ω_2^c specify the effect of one lag of category purchase amount (the amount spent in that category when a purchase was last made in the category) and the category inter-delivery time respectively on purchase probabilities for a category. One would expect these effects to vary depending on the inter-delivery times and the amounts involved and *a priori* there are no strong reasons to believe either way. Our prior [a normal(0,100)] reflects this prior belief over these parameters. We also set a similar prior for ω_3^c, ω_4^c which specifies the effect of assortment reductions in the favorite/non-favorite items on purchase probabilities, again recognizing that *a priori* there is little information on the strength of these effects. For the seasonality effect given by ω_5^c, ω_6^c & ω_7^c (in terms of quarter1, quarter2 and quarter3 of the year), except perhaps during the holiday season (Thanksgiving/Christmas) where one would expect higher consumption of certain product categories (for example cooking oil, butter etc) we have little prior information to form sharp priors. Again, the prior on these parameters [a normal(0,10)] reflects this belief of ours.

The second category level model (the category purchase amount model, section 3.2), corresponding to *Sampler 4*, is given by equations (4), (5) and (6). The parameters in this model are analogous to those in the category incidence model, and we use similar reasoning to specify our prior distributions for these parameters. For computational convenience, in the estimation we do not specify a heterogeneity distribution over the η_h^c parameter and instead have the same parameter η_0^c for all the households.

4.2 The full conditionals

The estimation proceeds by running four separate MCMC samplers (Casella and George, 1992; Gelfand and Smith, 1990) - two for the store level models (Sampler 1 & 2) and two for the category level models (Sampler 3 & 4).

The first sampler corresponds to equations (2), (4) & (5) in the main article modeling the inter-delivery times. The second sampler corresponds to equations (6), (7) & (8) in the main article modeling the purchase amounts at the store level. The third sampler corresponding to equations (1), (2) & (3) in this technical-appendix models the category purchase incidence for each category. Finally, the fourth sampler corresponds to equations (4), (5) & (6) in this technical appendix and models the category purchase amounts conditional on the total shopping basket amount.

The convergence diagnostics for these MCMC samplers was carried out using the software: Bayesian Output Analysis Program (BOA) Version 0.5.0. We provide below a shortened version of the full conditional distributions associated with these four samplers. For the interested reader the full set of conditional distributions can be obtained from the authors.

Sampler 1 proceeds by recursively drawing from the following set of full conditional distributions.

$$\Pi\left[\lambda_h / \{w_{ht}\}, \{\lambda_{ht}\}, \overline{\lambda}, \tau^2\right] \qquad where \quad \lambda_{ht} = \exp\left(\lambda_h + \delta_1 ASSORT_{ht} + \delta_2 \log Q_{ht-1}\right)$$

$$\Pi\left[\overline{\lambda} / \{\lambda_h\}, \tau^2, a_1, b_1\right]$$

$$\Pi\left[\tau^2 / \{\lambda_h\}, \overline{\lambda}, a_2, b_2\right]$$

$$\Pi\left[\nu / \{w_{ht}\}, \{\lambda_{ht}\}, a_5, b_5\right]$$

$$\Pi\left[\delta_1 / \{w_{ht}\}, \{\lambda_{ht}\}, a_3, b_3\right]$$

$$\Pi\left[\delta_2 / \{w_{ht}\}, \{\lambda_{ht}\}, a_4, b_4\right]$$

Sampler 2 proceeds by recursively drawing from the following set of full conditional distributions.

$$\Pi\left[\mu_h / \{Q_{ht}\}, \{Z_{ht}\}, \sigma^2, \overline{\mu}, \theta^2\right] \qquad where \quad Z_{ht} = \mu_1 ASSORT_{ht} + \mu_2 \log TIME_{ht}$$

$$\Pi\left[\overline{\mu} / \{\mu_h\}, \theta^2, e_1, f_1\right]$$

$$\Pi\left[\theta^2 / \{\mu_h\}, \overline{\mu}, e_2, f_2\right]$$

$$\Pi\left[\sigma^2 / \{\mu_{ht}\}, \{Q_{ht}\}, e_5, f_5\right] \qquad where \quad \mu_{ht} = \mu_h + \mu_1 ASSORT_{ht} + \mu_2 \log TIME_{ht}$$

$$\Pi\left[\mu_1 / \{Q_{ht}\}, \{Y_{1ht}\}, \sigma^2, e_3, f_3\right] \qquad where \quad Y_{1ht} = \mu_h + \mu_2 \log TIME_{ht}$$

$$\Pi\left[\mu_2 / \{Q_{ht}\}, \{Y_{2ht}\}, \sigma^2, e_4, f_4\right] \qquad where \quad Y_{2ht} = \mu_h + \mu_1 ASSORT_{ht}$$

Sampler 3 proceeds by recursively drawing from the following set of full conditional distributions for *each* product category (c = 1,2,…,C).

$$\Pi\left[\omega_h^c / \{\omega_{ht}^c\}, \{I_{ht}^c\}, \xi^{2^c}, \overline{\omega}^c\right]$$

where

$$\omega_{ht}^c = \omega_h^c + \omega_1^c Q_{ht-1}^c + \omega_2^c TIME_{ht}^c + \omega_3^c ASSORTF_h^c + \omega_4^c ASSORTNF_h^c + \omega_5^c QTR1_{ht}^c + \omega_6^c QTR2_{ht}^c + \omega_7^c QTR3_{ht}^c$$

I_{ht}^c is an indicator variable indicating a purchase in category c by household h on purchase occasion t.

$$\Pi\left[\overline{\omega}^c / \{\omega_h^c\}, \xi^{2^c}, g_2, h_2\right]$$

$$\Pi\left[\xi^{2^c} / \{\omega_h^c\}, \overline{\omega}^c, g_2, h_2\right]$$

$$\Pi\left[\omega_1^c / \{\omega_{ht}^c\}, \{Q_{ht-1}^c\}, \{I_{ht}^c\}, g_3, h_3\right]$$

$$\Pi\left[\omega_2^c / \{\omega_{ht}^c\}, \{TIME_{ht}^c\}, \{I_{ht}^c\}, g_4, h_4\right]$$

$$\Pi\left[\omega_3^c / \{\omega_{ht}^c\}, \{ASSORTF_{ht}^c\}, \{I_{ht}^c\}, g_5, h_5\right]$$

$$\Pi\left[\omega_4^c / \{\omega_{ht}^c\}, \{ASSORTNF_{ht}^c\}, \{I_{ht}^c\}, g_6, h_6\right]$$

$$\Pi\left[\omega_5^c / \{\omega_{ht}^c\}, \{QTR1_{ht}^c\}, \{I_{ht}^c\}, g_7, h_7\right]$$

$$\Pi\left[\omega_6^c / \{\omega_{ht}^c\}, \{QTR2_{ht}^c\}, \{I_{ht}^c\}, g_8, h_8\right]$$

$$\Pi\left[\omega_7^c / \{\omega_{ht}^c\}, \{QTR3_{ht}^c\}, \{I_{ht}^c\}, g_9, h_9\right]$$

Sampler 4 proceeds by recursively drawing from the following set of full conditional distributions for *each* product category (c = 1,2,…,C).

$$\log\Pi\left[\eta_0^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, \rho^{2^c}, \overline{\eta}^c, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$

where

$$\eta_{ht}^c = \exp\left(\eta_0^c + \eta_1^c Q_{ht-1}^c + \eta_2^c TIME_{ht}^c + \eta_3^c ASSORTF_h^c + \eta_4^c ASSORTNF_h^c + \eta_5^c QTR1_{ht}^c + \eta_6^c QTR2_{ht}^c + \eta_7^c QTR3_{ht}^c\right)$$

$$\Pi\left[\eta_1^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_3, l_3, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$

$$\Pi\left[\eta_2^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_4, l_4, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$

$$\Pi\left[\eta_3^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_5, l_5, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$

$$\Pi\left[\eta_4^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_6, l_6, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$
$$\Pi\left[\eta_5^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_7, l_7, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$
$$\Pi\left[\eta_6^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_8, l_8, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$
$$\Pi\left[\eta_7^c / \{\eta_{ht}^c\}, \{I_{ht}^c\}, k_9, l_9, \{\mathbf{SHARE}_{ht}^{N_{ht}}\}\right]$$

Many of the conditional distributions in the above four samplers are from non-conjugate distributions. We employ a random walk Metropolis Hastings algorithm to obtain draws from these distributions, using a normal/truncated normal proposal density with its variance used as a tuning constant.

5 The Estimated Coefficients (*Category level models*)

The category level models are specified by equations (1), (2), (3) [the Category incidence model] and (4), (5) & (6) [the Category purchase amount model]. The parameters estimated in these models are $\omega_h^c, \omega_1^c, \omega_2^c, \omega_3^c, \omega_4^c, \omega_5^c, \omega_6^c, \omega_7^c$ & $\overline{\omega}^c, \xi^{2^c}$ for the category incidence model and $\eta_0^c, \eta_1^c, \eta_2^c, \eta_3^c, \eta_4^c, \eta_5^c, \eta_6^c, \eta_7^c$ & $\overline{\eta}^c, \rho^{2^c}$ for the category purchase amount model. The superscript *c* denotes the category, of which we use 76 in total (75 categories plus “All others” category). In line with the focus of this work and due to constraints on space we do not report all these estimates, and instead concentrate on coefficients (ω_3^c, ω_4^c) and (η_3^c, η_4^c). Again, these parameters specify the effect of assortment reductions in favorite items and non-favorite items on the category incidence probability and category share-of-the-basket respectively[3]. Also, we do not discuss individual categories in order to maintain our broader focus.

[3] For the interested reader, the complete set of estimated coefficients can be obtained from the corresponding author.

Table 3 reports the posterior means for these coefficients $(\omega_3^c, \omega_4^c, \eta_3^c, \eta_4^c)$. It also indicates whether or not 95% of the posterior mass for these estimates lies on one side of zero. To aid the visualization of these results, the same information is also presented in Figures 4 & 5.

These figures visually depict these coefficients (Figure 4 for ω_3^c, ω_4^c and Figure 5 for η_3^c, η_4^c) for each of the 76 categories. This depiction groups the categories based on the sign of these coefficients (as measured by whether 95% of posterior mass lies on either side of zero; the shaded regions in Table 3) and whether the coefficient is positive or negative. The horizontal axis in these figures is the coefficient on $ASSORTF_h^c$ (the extent of assortment reduction in favorite items), ω_3^c for the category incidence model (Figure 4) and η_3^c for the category purchase amount model (Figure 5). The vertical axis is the coefficient on $ASSORTNF_h^c$ (the extent of assortment reduction in non-favorite items), ω_4^c for the category incidence model (Figure 4) and η_4^c for the category purchase amount model (Figure 5).

Table 3 Parameter estimates, Category level models[4]

Cat No.	Category Description	ω_3	ω_4	η_3	η_4
1	MARKET PRODUCE	-0.4583	-1.8397	-0.2732	0.2301
2	DAIRY MILK/DRINKS/HALF & HALF/CREAM	1.2409	-0.8199	-0.3426	0.2394
3	GROC BREAD PACKAGED	-0.3636	0.4424	0.0091	0.1512
4	GROC CEREAL/BREAKFAST FOODS/GRANOLA SNACKS	-0.4046	0.0003	0.2171	-0.0287
5	MARKET MEAT	-1.1699	0.5480	0.2366	0.2520
6	GROC SODA/BOTTLED TEA	-0.2667	0.1356	-0.1861	0.1356
7	MARKET DELI	-0.9457	-0.6468	-0.0732	0.1267
8	DAIRY CHEESE PRE-PACKAGED	-0.2209	0.0721	-0.0713	-0.0060
9	DAIRY JUICE/DRINK	-1.3423	0.6661	0.0721	0.2136
10	GROC JUICE/DRINK NONREFRIGERATED	-0.0425	0.4189	-0.0766	0.0782
11	DAIRY YOGURT/PUDDING/GELATIN	0.9642	-0.2370	0.0239	0.2617
12	GROC SNACKS/DIPS	-0.1542	-0.9880	-0.1626	0.0276
13	FROZEN ICE CREAM/NOVELTIES/YOGURT/SHERBET	0.4759	-0.3694	-0.0715	0.1117
14	DAIRY EGGS/SUBSTITUTES	0.5260	-0.1324	-0.0320	0.1155
15	DAIRY MARGARINE/BUTTER	-0.5326	-0.2275	-0.1229	0.2514
16	GROC CRACKERS/PROCESSED CHEESE	-0.1820	-0.3248	-0.1247	0.2044
17	HH PAPER TOWELS	-1.1413	0.5616	0.0442	0.3717
18	HH TOILET PAPER	-0.1627	0.0522	0.0119	0.1155
19	GROC WATER	-0.3654	1.0750	0.4116	0.1400
20	MARKET BAKERY (FRESH)	-0.0828	-0.4108	-0.0235	0.3068
21	GROC COOKIES	-0.1222	-0.1313	-0.0881	0.2644
22	GROC SOUP CANNED & MIXES	-0.6164	0.9008	0.1488	0.0306
23	HH LAUNDRY PRODUCTS	-0.1855	-0.0535	-0.0182	-0.0177
24	GROC PASTA DRIED	-0.0889	0.0837	-0.1366	0.1683
25	FROZEN VEGETABLES	-0.6732	-0.0850	0.3138	-0.0634
26	GROC SPAGHETTI SAUCE/GRATED CHEESE	-0.0453	-1.2077	-0.0873	0.0978
27	PET FOOD & PRODUCTS	-3.1378	1.6346	-0.1842	0.2820
28	GROC CANNED VEGETABLES/TOMATOES/BEANS	0.1832	-0.5984	-0.2026	0.2699
29	FROZEN BREAKFAST	0.1210	-0.8095	0.0794	0.1239
30	HH BAGS	-0.0257	-0.1346	-0.1590	0.2349
31	GROC PREPARED FOODS BEANS/PASTA/MEAT/SAUCES	-0.0319	-0.3114	0.1823	-0.0604
32	DAIRY COTTAGE CHEESE/SOUR CREAM/DIPS	-0.6360	-0.8080	0.0475	0.0817
33	FROZEN MEALS	0.0159	-0.6245	0.0058	0.1850
34	HH PAPER PLATES/CUPS/NAPKINS	0.0844	-0.6706	0.0607	0.0849
35	GROC CONDIMENTS/MEAT SAUCES/MARINADES	-0.1146	-0.2954	0.1448	-0.1186
36	HH DISHWASHING PRODUCTS	0.1630	-1.2359	-0.1183	0.1452

[4] Posterior means reported; shaded cell indicates that the 95% posterior interval does not contain 0.

37	HH CLEANERS BATH/KITCHEN/GLASS	-0.3171	-0.9032	0.0611	-0.0339
38	HH FACIAL TISSUES	0.4650	-0.3781	0.1323	0.1842
39	GROC CANNED SEAFOOD	-0.4875	-0.8240	0.0488	0.1216
40	GROC COFFEE	-1.0261	-0.3675	-0.2839	0.2418
41	GROC ETHNIC MEXICAN	-1.0340	-0.0192	0.0467	-0.0734
42	GROC RICE/DRIED BEANS	0.1252	-0.1483	0.0156	-0.0527
43	HBA SOAP BAR/LIQUID	0.2016	-1.0279	0.2813	0.1756
44	GROC HOT DOGS/SAUSAGE PRE-PACKAGED	0.0522	0.0014	0.2127	0.1660
45	GROC CANDY/GUM/MARSHMALLOWS	0.0384	-1.5582	0.1861	0.0850
46	GROC SUGAR	0.3445	-2.5700	0.0993	0.0802
47	FROZEN JUICE/DRINK	0.8273	-0.3932	-0.1567	0.4410
48	HBA ORAL HYGIENE	-0.0624	-0.7174	-0.3229	0.3091
49	GROC FRUIT DRIED/SNACKS	-0.2474	-0.9859	-0.2363	0.1661
50	MARKET SEAFOOD	-0.2232	0.7277	-0.1789	0.0696
51	GROC CANNED FRUIT	-0.1459	-0.3058	0.1092	0.1703
52	BABY DIAPERS	-1.6020	-0.0229	-0.0844	0.1498
53	GROC SALAD DRESSINGS & TOPPINGS	-0.2655	0.0233	0.0257	0.1288
54	GROC JAM/JELLIES/SPREADS/HONEY	-0.3880	-1.0777	0.0080	0.1190
55	GROC OLIVES/PICKLES/PEPPERS/RELISH	0.2739	-0.0714	-0.0110	0.1974
56	BABY ACCESSORIES	0.4870	0.5055	0.0989	0.2349
57	GROC PANCAKE/WAFFLE MIX/SYRUP	0.0302	-0.0032	0.0591	0.0680
58	GROC BAKERY/SNACKS PRE-PACKAGED	-1.4305	-0.2753	-0.0407	0.2239
59	GROC SALT/SEASONING/SPICE/SAUCE MIX	-0.3111	-0.6516	-0.3138	-0.0548
60	GROC DELI PREPACKAGED	-0.0475	0.7791	0.0133	0.1598
61	BABY FOOD/FORMULA	-3.2402	3.0400	-0.2008	0.2283
62	GROC PEANUT BUTTER	-0.1683	-0.0061	0.0636	-0.0495
63	GROC OIL COOKING	-0.2738	-0.3874	-0.1029	0.1204
64	GROC BACON/BREAKFAST SAUSAGE	0.0117	-0.1913	0.1196	0.1368
65	HBA HAIR PRODUCTS	-0.8115	0.4047	0.0513	-0.1431
66	FROZEN MEAT	0.8807	-0.0641	0.4848	-0.0459
67	GROC BAKING NEEDS/CHOC/NUTS	-0.2141	-0.7687	0.0496	0.1166
68	GROC APPLESAUCE	-0.0436	-2.6346	0.0261	0.0519
69	FROZEN PIZZA	-0.0430	-1.0167	-0.1751	0.0665
70	GROC TEA	-0.4287	-0.8857	-0.0145	-0.0425
71	DAIRY DOUGH/BREAD/BISCUITS	-0.8728	-1.4472	0.1542	0.2522
72	HH ALUMINUM FOIL/WAX PAPER/FOOD WRAP	-0.7856	-0.6027	0.0853	0.2558
73	FROZEN BAKERY/WHIPPED TOPPINGS	-0.0778	-1.9122	-0.0961	0.1700
74	FROZEN POTATOES/ONION RINGS	-0.3494	0.9793	-0.0910	0.1157
75	DAIRY PASTA/SAUCES/TOPPINGS	-0.3784	0.6411	0.0392	0.4586
76	All OTHERS	-0.9406	0.0837	-0.0036	-0.0861

Figure 4 Visual depiction of estimates from the Category incidence model

ω_4

$\omega_3 < 0,\ \omega_4 > 0$

5 MARKET MEAT
9 DAIRY JUICE/DRINK
17 HH PAPER TOWELS
22 GROC SOUP CANNED & MIXES
27 PET FOOD & PRODUCTS
61 BABY FOOD/FORMULA

$\omega_3 \cong 0,\ \omega_4 > 0$

3 GROC BREAD PACKAGED
10 GROC JUICE/DRINK NONREFRIGERATED
19 GROC WATER
50 MARKET SEAFOOD
56 BABY ACCESSORIES
60 GROC DELI PREPACKAGED
74 FROZEN POTATOES/ONION RINGS
75 DAIRY PASTA/SAUCES/TOPPINGS

$\omega_3 > 0,\ \omega_4 > 0$

$\omega_3 \cong 0,\ \omega_4 \cong 0$

4 GROC CEREAL/ BREAKFAST FOODS/ GRANOLA SNACKS
6 GROC SODA/BOTTLED TEA
8 DAIRY CHEESE PRE-PACKAGED
14 DAIRY EGGS/SUBSTITUTES
18 HH TOILET PAPER
21 GROC COOKIES
23 HH LAUNDRY PRODUCTS
24 GROC PASTA DRIED
30 HH BAGS
38 HH FACIAL TISSUES
42 GROC RICE/DRIED BEANS
44 GROC HOT DOGS/SAUSAGE PRE-PACKAGED
47 FROZEN JUICE/DRINK
51 GROC CANNED FRUIT
53 GROC SALAD DRESSINGS & TOPPINGS
55 GROC OLIVES/ PICKLES/ PEPPERS/ RELISH
57 GROC PANCAKE/WAFFLE MIX/SYRUP
62 GROC PEANUT BUTTER
64 GROC BACON/BREAKFAST SAUSAGE
66 FROZEN MEAT

$\omega_3 < 0,\ \omega_4 \cong 0$

15 DAIRY MARGARINE/BUTTER
25 FROZEN VEGETABLES
41 GROC ETHNIC MEXICAN
52 BABY DIAPERS
58 GROC BAKERY/SNACKS PRE-PACKAGED
65 HBA HAIR PRODUCTS
72 HH ALUMINUM FOIL/ WAX PAPER/ FOOD WRAP
76 ALL OTHERS

$\omega_3 > 0,\ \omega_4 \cong 0$

11 DAIRY YOGURT/ PUDDING/ GELATIN

ω_3

$\omega_3 \cong 0,\ \omega_4 < 0$

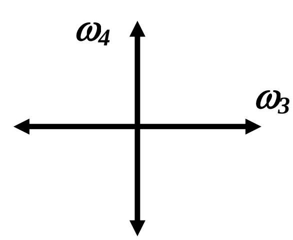
1 MARKET PRODUCE
12 GROC SNACKS/DIPS
16 GROC CRACKERS/PROCESSED CHEESE
20 MARKET BAKERY (FRESH)
26 GROC SPAGHETTI SAUCE/GRATED CHEESE
28 GROC CANNED VEGETABLES/ TOMATOES/ BEANS
29 FROZEN BREAKFAST
31 GROC PREPARED FOODS BEANS/ PASTA/ MEAT/ SAUCES
33 FROZEN MEALS
34 HH PAPER PLATES/ CUPS/ NAPKINS
35 GROC CONDIMENTS/MEAT SAUCES/ MARINADES
36 HH DISHWASHING PRODUCTS
37 HH CLEANERS BATH/KITCHEN/GLASS
39 GROC CANNED SEAFOOD
43 HBA SOAP BAR/LIQUID
45 GROC CANDY/GUM/MARSHMALLOWS
46 GROC SUGAR
48 HBA ORAL HYGIENE
49 GROC FRUIT DRIED/SNACKS
54 GROC JAM/JELLIES/SPREADS/HONEY
59 GROC SALT/ SEASONING/ SPICE/ SAUCE MIX
63 GROC OIL COOKING
67 GROC BAKING NEEDS/CHOC/NUTS
68 GROC APPLESAUCE
69 FROZEN PIZZA
70 GROC TEA
73 FROZEN BAKERY/WHIPPED TOPPINGS

$\omega_3 < 0,\ \omega_4 < 0$

7 MARKET DELI
32 DAIRY COTTAGE CHEESE/SOUR CREAM /DIPS
40 GROC COFFEE
71 DAIRY DOUGH/BREAD/BISCUITS

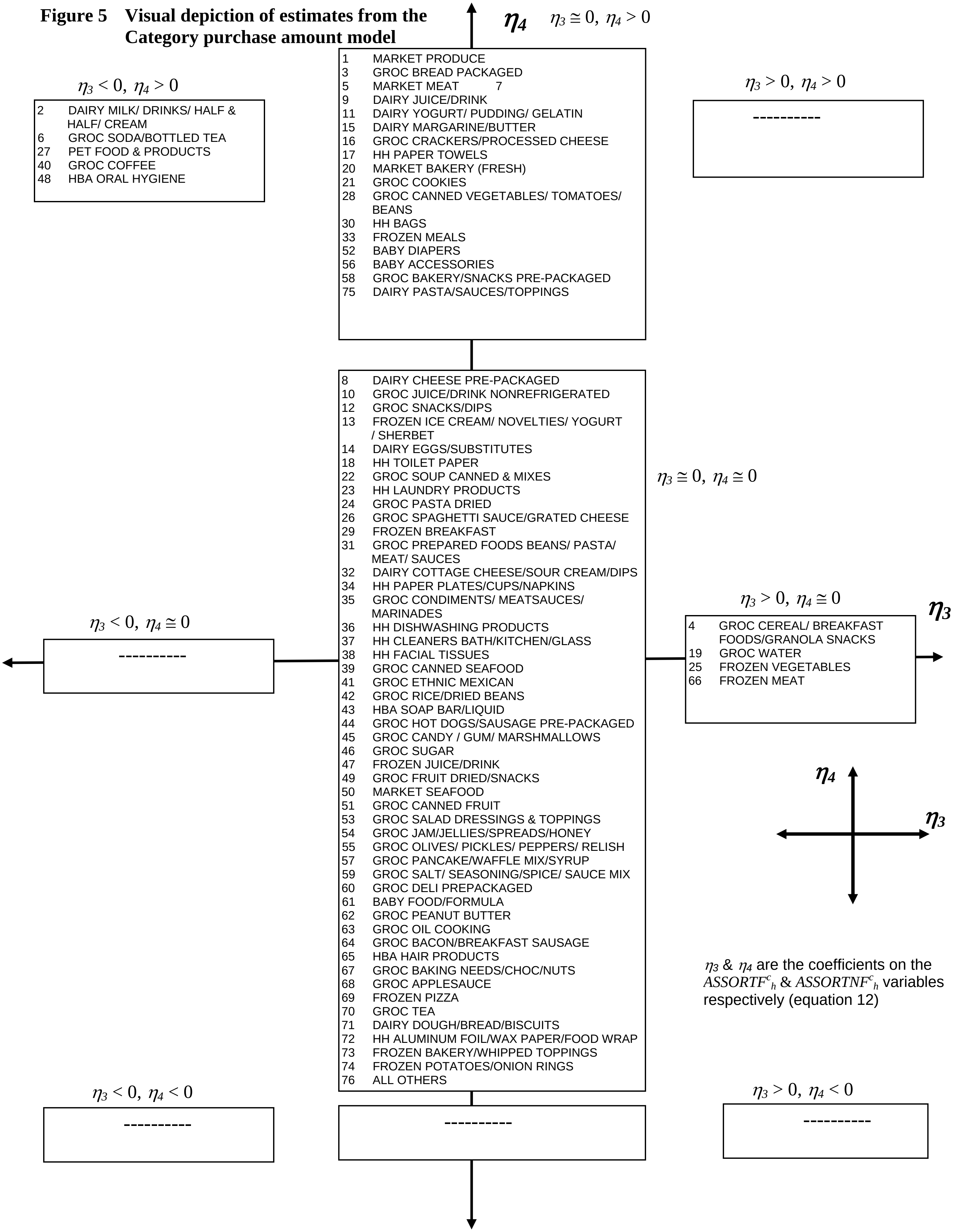
$\omega_3 > 0,\ \omega_4 < 0$

2 DAIRY MILK/DRINKS/HALF & HALF/ CREAM
13 FROZEN ICE CREAM/ NOVELTIES/ YOGURT/ SHERBET



ω_3 & ω_4 are the coefficients on the $ASSORTF^c_h$ & $ASSORTNF^c_h$ variables respectively (equation 9)

Figure 5 Visual depiction of estimates from the Category purchase amount model



Thus each of these figures has nine boxes in which the categories are arranged. For example, the central box in both these figures has categories for which the assortment reductions (both in the favorite and non-favorite items) have no effect on the category incidence probabilities (Figure 4) as well as no effect on the share-of-the-basket for that category (Figure 5). Similarly the top right box has categories in which the reduction in the favorite as well as non-favorite items both lead to an increased category incidence (Figure 4); and lead to an increased share-of-the-basket for that category (Figure 5).

Recall that the store level results (section 4.1 in the main article) indicate that reduction in assortments led to a decrease in purchase frequency (increased inter-delivery times) as well as a decrease in purchase amounts. So, as mentioned earlier, at the overall store level the reduction in assortment led to lost sales for the grocer. However, considering the category specific results (Figures 4 & 5), note that not all the categories lie in the lower left quadrant, and note also that they are spread out. These results imply that, although at the overall store level the assortment cut led to reduced sales, this trend was not uniformly observed across all categories. Choosing a subset of categories for analysis and ignoring purchases in all the other categories could give different results than those at the overall store level. This again points to the importance of considering *all* of the categories or an appropriately sized random sample if the objective is to study the impact of assortment reductions on the overall store.

Most of the categories lie in boxes along the vertical axis (the central column), 55 in Figure 4 (the category incidence model) and 67 in Figure 5 (the category share-of-the-basket model), implying that for a majority of categories reduction in favorite items caused no changes in the category purchase incidence probabilities nor the category's share-of-the-basket (this effect being more pronounced in the share-of-the-basket). A study by Broniarczyk, Hoyer &

McAlister 1998 found that reduction in assortment had “little” effect on consumer perceptions of assortment as long as the “favorite” items were available and the category space (in the context of a Brick-and-Mortar grocer) was held constant. We find here that, for a majority of the categories, even reductions of favorite items had no effect on the category (in terms of purchase incidence and share-of-the-basket). Surprising results are for the categories in the right column boxes (three categories in Figure 4 and four in Figure 5). For these categories, a reduction in favorite items actually increased the purchase incidence probability/share-of-the-basket. Even so, it is difficult to compare inferences of the two studies in this regard, as the definitions of “favorites” are quite different. Our “favorites” are products that were observed to have been purchased, while Broniarczyk et al. (1998) surveyed consumers to determine the product of highest preference.

The effect of assortment cuts for non-favorites on purchase incidence was similar to that for favorites in that for majority of the categories these cuts led to a decrease in purchase incidence. However, there were many more increases in share-of-basket for non-favorites than there were for favorites. Cuts in favorites engendered share-of-basket increases for four categories (right column of Figure 5), while cuts in non-favorites engendered share-of-basket increase for 23 categories (top row of Figure 5). Why would purchase amounts increase after non-favorite items were eliminated? Iyengar and Lepper (2000) showed that choice among too many items is de-motivating and can reduce sales. We imagine that, in the online setting, consumers would tend not to browse overly cluttered categories (those with many non-favorites) but would simply search for a specific planned purchase. If the number of items was not overwhelming (where many non-favorites have been eliminated), we envision that a greater number of consumers would browse, buying some items on impulse in addition to the planned purchase.

While cuts in non-favorite items increased the category share-of-basket for 23 categories, none of the categories saw a decline in the share-of-basket on account of cuts in non-favorites (no category in the bottom row boxes of Figure 5). This result implies that, though many categories have reduced purchase incidence probability on account of cuts in non-favorite items (Figure 4), the household does not reduce the relative amounts allocated to these categories *when they do purchase in the category*. The effects on the categories are either close to zero in the relative amounts or increases in the relative amounts (middle row and top row boxes of Figure 5 respectively). Managers making decisions on assortment cuts must worry more about the adverse impact of cuts on category purchase probabilities rather than the impact on re-allocation of shopping basket amounts within the set of categories.

6 Category 'Sales' vis-à-vis Category Purchase Frequencies

An interesting insight is revealed in examining the results from the category level models and the purchase frequencies observed in various categories. The more frequently purchased categories were less adversely affected in terms of the impact on sales, category incidence, and share of basket. We can use our category level coefficient estimates to arrive at a estimate of the net % impact on category incidence, share of basket and "sales" for a household *h* on purchase occasion *t* conditional on the store inter-purchase time being delayed and store purchase basket shrinking (as is the case in our store results). These estimates when regressed with the mean inter-purchase times observed in various categories prior to assortment cuts lead to an average additional loss of 2.23% (s.e. of 0.71%) in category ‘sales’ for each additional day of category inter-purchase time[5].

[5] The equivalent losses in category incidence and share of basket were 2.18% (s.e. of 0.70%) and 0.04% (s.e. of 0.01%) respectively. Further details of this analysis can be obtained from the authors on request.