"Shopping cost and brand exploration in online grocery"
by Andrea Pozzi
Web Appendix

A-1 Robustness

A-1.1 Sensitivity to initial conditions

The very definition of brand exploration requires the identifications of the set of brands that are already known to a consumer at t_0. I construct this set using the first few months of purchase data which are not then used for the estimation. Choosing the length of the period used to calibrate initial conditions imposes a trade-off. On the one hand, the longer the amount of data allocated to recovering the set of known brands the more accurate this will be. On the other hand, extending the length of the calibration window reduces the number of observations left for the estimation. For the estimates reported in Table 7 of the main text I implicitly assumed that three months was a long enough period to credibly recover the set of cereal brands explored by the household in the past. In reality, this assumptions surely generates spurious instances of brand exploration, although it is not obvious that this bias would be at all correlated with the shopping channel chosen for the trip. Nevertheless, I present below evidence that results are not overly sensitive to different assumptions.

A first concern relates the fact that I retrieve the set of brands already known to a household using a limited span of observation and only using purchases occurring at one specific grocery chain. Ideally, this set should be constructed using the entire shopping history of the household at all the stores visited. In Panel A of Table A-1 I asses the practical relevance of this shortcoming using an additional piece of scanner data: a subset of observation from the HomeScan panel, collected by AC Nielsen. Each household in this sample is equipped at home with a scanning device through which they can record all their purchases at every grocery chain visited.[1] The sample I obtained covers over 89,000 households in the period between January 1998 and December 2007, although not all the households are in the sample for the entire span. 85,827 households shopped for breakfast cereals at least once for a total of 2,752,802 trips. The number of households engaging in online shopping of grocery is limited, which explains why this data source could not be used for the core analysis.

The HomeScan panel allows to recover the set of known brands using a much longer shopping history, potentially leading to accuracy gains. For purchases generated by households active between June 2004 and June 2006, the difference between the number of brand trials counted using a three months calibration span and the number resulting from using the full previous history for the calibration is 71,194 cases. This implies that mislabeled brand exploration instances make up 3% of the overall number of trips. Some of these households, however, may not have been in the sample very long before June 2004 which would reduce the gap between using three months or the entire history to define brand exploration.

[1] See Aguiar and Hurst (2007); Broda and Weinstein (2010); Einav, Leibtag, and Nevo (2010) for more detailed descriptions of this data source.

A more powerful test implies computing the same statistics for a subsample of households who have a fairly long record of past purchases prior to September 2004. The second column of Panel A conditions on household having been in the sample for at least four years as of June 2006. This implies that their shopping history can be recovered from at least two years of past purchases. Now the share of misclassified brand trials grows, although not dramatically. The two criteria would lead to different conclusions in 10% of the cases. This figure maintains stable if we strengthen the requirement focusing on households who have been the sample for at least 6 or 8 years.

Panel B, instead relies on the same single-grocer data used for the rest of the analysis and checks whether coefficients obtained in columns (1) and (3) of Table Table 7 of the main text are robust to changing the length of the period used to construct the set of brands initially known to the household. In particular, I compare the default length of three months with alternative choices: one month, six months, and one year. These changes do not affect the main result: online shopping is still associated with lower rates of brand and UPC exploration. The final column replicates specification of column (1) of Table Table 7 of the main text considering only purchases of cereal brand introduced in the market between June 2004 and June 2006.[2] Purchase of these brands could not have occurred before the beginning of the sample; therefore, initial conditions do not affect the classification of brand trials. The negative correlation between Internet purchase and probability of brand exploration is even stronger in this case.

A-1.2 Alternative explanations

Other dimensions of the difference between online and in-store shopping can contribute to explain the main result in Table Table 7 of the main text in addition to those explored in the paper. To begin with, there is a striking difference between the size of online and in-store trips, with the former being on average much larger. Column (1) in Table A-2 shows that this does not suffice to explain difference in brand trial. The gap in exploration between the two channels shrinks by half once I control for the size of the trip but the online channel is still associated with a significant reduction in the probability of trying a new brand.

Online and traditional shopping channel also radically differ on characteristics and effectiveness of promotion strategies. This could drive my results since promotional activities are an effective way to induce trials of new brands. A first issue concerns coupons whose use is massive in the cereal product category (Nevo and Wolfram, 2002). The fact that paper coupons can only be redeemed in-store could be behind the higher propensity to new trials on the traditional channel. However, coupons do not play a major role in the supermarket chain under analysis: less than 2% of the transactions involve use of paper coupons. The chain tries to foster use of the loyalty card linking discounts to the membership card.[3] More-

[2]Brands in this sample include both altogether new cereal, such as KRAFT South Beach Diet, and the so called "limited edition" cereals which are introduced and sold for a short period of time exploiting contemporaneous release of movies or books (e.g. Kellogg's Incredibles Incrediberry, Post The Polar Express).

[3]Prior to the beginning of my sample, issuing of coupon books had even been discontinued by the grocer and all the promotions were linked to use of the loyalty card. Coupons books were later reintroduced but

over, use of paper coupons is stronger among the elderly (Aguiar and Hurst, 2007) which only represent a small fraction of my sample (as we would expect, given that it only includes people who have tried the online shopping channel).

The specification in column (2) includes a dummy for the purchased item being on discount. Price promotions are highlighted both online and in-store by a red price tag and the icon of the loyalty card. The effectiveness and salience could differ across channels. Brands on promotion are more likely to be tried, as expected. The interaction between the promotion and the channel dummy is positive and significant: if anything, promotions are more effective in drawing exploration on the Internet. This suggests that a bias from not accounting for price promotions would go against my findings.

The two channels differ even more when we consider promotional activity other than price discounts. The store setting offers the possibility of end-caps display or of offering free trials; this is not possible online. To control for in-store promotion of specific brands I include brands fixed effects (column 3) and brand-week fixed effects (column 4) since promotion generally take place on a weekly basis. In both cases the gap in exploration between the two shopping channels does not fade away.

Another alternative explanation revolves around the role played by kids. Kids are potential triggers of brand exploration in cereals, as they follow fashion and peers' example. If we assume that children accompany parents and influence them in the store but do not assist to Internet orders, this might explain the difference in the rate of brand exploration across the two channels. However, brand exploration is not limited to kids' cereal. The mean of the "new trial" dummy is .36 for kids cereal, only slightly above the overall figure of .34. Moreover, the gap between online and in-store rates of exploration is not dramatically more pronounced for kids' cereals than it is for other types of cereals. Online exploration in kids cereals is .28, while in-store is .39, an 11 percentage points gap. The gap for the sample without kids' cereals is 9 percentage points (.25 online vs. .34 in-store). I perform two robustness checks of the main descriptive exercise. In column (5) I exclude from the sample all the trips where kids' cereals were purchased to find that the qualitative results do not change. Finally, in column (6) I interact the online dummy with the census variable "family". This should give a measure of how much of the effect is truly driven by families showing up at different rates online and in-store. The interaction dummy is not significantly different from zero.

A-2 Estimation

In this section, I discuss details of the estimation procedure presented in Section IV. Once again, I present the procedure used for the specification with random coefficients. The routine including only fixed coefficient is just a simplified case of it.

---

do not play a major role. To get an idea, I collected weekly booklets for a single price area between April and October 2008. Those booklets are mailed as general advertising and list price promotions for the week. They also contain few paper coupons that can be cut out and used by the customer. Out of the 27 weekly booklets I collected, only three offered a coupon for a breakfast cereal brand.

Table A-1: Probability of exploration: Robustness to the construction of the set of previously known brands.

Panel A: Divergences in the number of new trials for different length of window used to calibrate the initial conditions.

Length of the period used for the calibration	*Full sample*	*Households in the sample for at least 4 years*	*Households in the sample for at least 6 years*	*Households in the sample for at least 8 years*
Num. of divergences	71,194	26,394	19,174	10,946
% of divergences	3%	10%	10%	10%
Num. of households	47,250	5,827	3,984	1,853

Panel B: Effect of online shopping on the probability of new trial for different length of window used to calibrate the initial conditions.

	Dependent Variable= New brand				Dependent Variable= New upc				Dep.Var.= New brand
Length of the period used for the calibration	*3 months*	*1 month*	*6 months*	*1 year*	*3 months*	*1 month*	*6 months*	*1 year*	
online	-.077**	-.066**	-.113**	-.102**	-.110**	-.062**	-.080**	-.088**	-.24**
	(.0045)	(.0041)	(.0043)	(.0043)	(.0046)	(.0069)	(.0048)	(.0047)	(.0320)
N	163,324	177,618	143,771	100,377	123,064	132,064	96,053	75,554	4,512
Sample restriction									New introduced brands

Notes: In *Panel A* a *divergence* is defined as an instance where purchase of a brand is labeled as an exploration event when calibrating the set of brands known to the household using a three-months window; whereas it would not qualify as such if the set had been constructed relying on the entire prior shopping history available.
Panel B replicates specifications from columns (4) and (6) of Table Table 7 of the main text. Trial of a new brand is defined as purchase of a brand not purchased in the past x months, where x is set to a different value in each column. The last column only consider purchases of cereal brands newly introduced to the market. In this case case no assumptions are needed to define brand trial: it occurs at the first observed purchase of the brand by the household. All specifications include household and day of the week fixed effects. Standard errors (in parentheses) are clustered at the household level. Significance levels :$*$: 5% $**$: 1%

Table A-2: Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)
online	-.042**	-.084**	-.068**	-.132**	-.063**	-.116**
	(.003)	(.005)	(.004)	(.007)	(.005)	(.034)
trip size	.000					
	(.001)					
promotion		.024**				
		(.003)				
promotion× online		.014**				
		(.005)				
family × online						.000
						(.001)
Household f.e.	X	X	X		X	
Brand f.e.			X			
Brand×Week f.e.				X		
Obs.	163,324	163,324	163,324	121,148	114,926	117,062
Sample	all	all	all	top 25 brands	no kid cereals	demo

Notes: Trial of a new brand is defined as purchase of a brand not bought in the previous 3 months. Trips size is measured as the worth of the trip in dollars. Standard errors in parentheses.Column 5 excludes all trips involving purchase of kids'cereals. Column 6 only includes households for which demographic information is available. Columns 1-5 are linear probability models, Column 6 is a probit with marginal effects reported. Standard errors clustered at the household level. Significance levels :$*$: 5% $**$: 1%

There are three sets of parameters of interest: i) γ the vector of parameters of the channel selection equation; ii) $\beta_i, i = 1, \ldots, N$ the vectors of random coefficients in the brand equation. Each household i is characterized by two random coefficients: one for its upc stickiness online and another one for its upc stickiness in-store; iii) $\tilde{\beta}$ the fixed coefficients of the demand equation. Data consist of a vector of observed channel choices c for each trip of every household and on a matrix C of observables that are supposed to influence the latent utility of shopping on the Internet. On the demand side, I observe a sequence of product choices summarized in the vector y, which has an entry for each household-product-trip triplet. Such entry is a zero if the specific product was not purchased by that household in that period and a one if the product was purchased in that occasion. The matrix X contains shifters of the utility from purchase.

Let $X = (\overline{X}\tilde{X})$ be the partition of the matrix of observables for the demand equation into characteristics for which I estimate a random coefficient $(\overline{X})$ and the characteristics for which coefficients are fixed $(\tilde{X})$. Furthermore, I define $X_{y_{it}}$ as a matrix including only the characteristics of the particular product chosen by household i at time t.

The likelihood function for household i is

$$L_i(\gamma, \mu_i, \sigma_\mu, \beta_i, \Sigma) = \prod_{t=1}^{T_i} (P^{onl}_{iy_{it}t} \Phi(C_{it}\gamma + \mu_i))^{c_{it}} (P^{store}_{iy_{it}t}(1 - \Phi(C_{it}\gamma + \mu_i)))^{(1-c_{it})} \quad \text{(A-1)}$$

where Φ is the cdf of the standard normal and $P^{onl}_{iy_{it}t}$ is the probability that household i makes the product choice we observe in the data, given that the trip is online. It takes the form

$$P^{onl}_{iy_{it}t} = \int \frac{\exp \overline{X}_{iy_{it}t}\beta_i + \tilde{X}_{iy_{it}t}\tilde{\beta} + \xi_{it}}{\sum_k \exp \overline{X}_{ikt}\beta_i + \tilde{X}_{ikt}\tilde{\beta} + \xi_{it}} g_1(\xi) d\xi \quad \text{(A-2)}$$

Similarly

$$P^{store}_{iy_{it}t} = \int \frac{\exp \overline{X}_{iy_{it}t}\beta_i + \tilde{X}_{iy_{it}t}\tilde{\beta} + \xi_{it}}{\sum_k \exp \overline{X}_{ikt}\beta_i + \tilde{X}_{ikt}\tilde{\beta} + \xi_{it}} g_2(\xi) d\xi \quad \text{(A-3)}$$

The probabilities in equations A-2 and A-3 take the well known logit form. However, the unobserved shock ξ_{it} has to be integrated out. Due to the dependance between ξ and θ, the marginal distribution of ξ will be different (although Normal in both cases) depending on the channel choice.

Finally, the overall likelihood is

$$\mathbb{L} = \prod_{i=1}^{I} \int \int L_i f_1(\beta) f_2(\mu) d\beta d\mu \quad \text{(A-4)}$$

Although it would be possible to estimate the model using simulated maximum likelihood, Bayesian techniques prove more efficient, in terms of computation time, to handle the rich heterogeneity structure of the model. Below I describe the routine followed to estimate the model.

In each iteration of the MCMC routine, I draw from the conditional distribution of each of

the parameters. In addition, unobserved shocks are also treated as parameters; also drawn from their conditional distributions. Finally, each iteration involves a draw of the hyperparameters of the model.
In sum, the procedure requires drawing from the following conditional densities in each iteration.

$$\mathbb{K}(c^*|c, C, \mu_i) \quad \text{(A-5)}$$
$$\mathbb{K}(\gamma|c^*, C) \quad \text{(A-6)}$$
$$\mathbb{K}(\mu_i|\sigma_\mu) \quad \text{(A-7)}$$
$$\mathbb{K}(\sigma_\mu|\mu_i) \quad \text{(A-8)}$$
$$\mathbb{K}(\xi_{it}|y_i, X_i, \theta_{it}, \Sigma) \quad \text{(A-9)}$$

$$\mathbb{K}(\beta_i|y_i, X_i, b, W, \tilde{\beta}, xi_{it}) \quad \text{(A-10)}$$
$$\mathbb{K}(b|y, X, \beta_i, W) \quad \text{(A-11)}$$
$$\mathbb{K}(W|b, \beta_i) \quad \text{(A-12)}$$
$$\mathbb{K}(\tilde{\beta}|\beta_i, xi_{it}) \quad \text{(A-13)}$$
$$\mathbb{K}(\Sigma|\theta, \xi) \quad \text{(A-14)}$$

Each step is commented in detail below

Draw of the c^*_{it}

To estimate the parameters of the channel selection equation, I start to apply data augmentation and draw the latent utility of online shopping, c^*. The distributional assumption made over θ_{it} in equation (5), implies that the c^*_{it} are distributed according to a truncated normal with mean $C_{it}\gamma + \mu_i$ and variance 1. The support of the truncated normal is $(-\infty, 0]$ if the trip occurs in a store, and $[0, +\infty)$ if the trip is made online.

Draw of γ

The full conditional posterior on γ in equation A-6 is normal with mean $\bar{\gamma}$ and variance matrix V. Given the previous step gave us values for the latent utility of equation (2), this case becomes analogous to obtaining the posterior for regression parameters in a linear model. Therefore, we have

$$\gamma \sim N((C'C)^{-1}C'c^*, (C'C)^{-1}) \quad \text{(A-15)}$$

where c^* is the vector of the stacked c^*_{it}.

Draw of μ

For each agent, the random effect μ_i is drawn its conditional density in equation A-7, which is a normal distribution with mean 0 and variance σ_μ.

Computation of θ

Conditional on the draws of c^* and γ, I can compute the implied values for the residuals of the channel selection equation as $\theta_{it} = c^*_{it} - X\gamma - \mu_i$.

Draw of σ_μ

The density of σ_μ conditional on the other parameters is an inverse gamma distribution, $IG(\hat{a_1}, \hat{a_2})$, where

$$\hat{a_1} = \frac{N + 2a_1}{2} \tag{A-16}$$

$$\hat{a_2} = \frac{\sum_{i=1}^{N} \mu_i^2 + 2a_2}{(2 + N)} \tag{A-17}$$

where a_1 and a_2 are hyperparameters of the Inverted Gamma prior and N is the total number of random effects to be drawn (i.e., the number of households).

Draw of ξ_{it}'s

Draws of the time-individual specific shocks to the valuation of unknown brands are conditional to the simulated residuals from the channel selection probit. The conditional density in equation A-9 can be written as proportional to the following expression.

$$\mathbb{K}(\xi_{it}|y_i, X_i, \theta_{it}, \Sigma) \propto L(y_{it}, X_{it}|\beta_i, \tilde{\beta}, \xi_{it}) * \phi(\xi_{it}|\Sigma, \theta_{it}) \tag{A-18}$$

where the first part of the expression is the likelihood of the data given the parameters and $\phi(\cdot)$ is, given the prior in equation (5), a normal distribution with mean $\delta + \frac{\rho}{\sigma_\xi}\theta_{it}$, and variance $1 - \rho$.

We can't draw directly from this distribution, therefore; draws are simulated using a MH algorithm. In each iteration k I make one draw ν from $\phi(\xi_{it}|\Sigma, \theta_{it})$ and construct a "trial draw" $\xi_{it}^{trial} = \xi_{it}^{k-1} + \eta\nu$. The probability that this draw is accepted depends on the ratio between the posterior evaluated at the current and at the trial value, which provides a way of updating through the data. In particular, I draw a scalar u from a Uniform distribution and assess whether

$$u < \mathbb{K} = \frac{L(y_{it}, X_{it}|\beta_i, \tilde{\beta}, \xi_{it}^{trial}) * \phi(\xi_{it}^{trial}|\Sigma, \theta_{it})}{L(y_{it}, X_{it}|\beta_i, \tilde{\beta}, \xi_{it}^{k-1}) * \phi(\xi_{it}^{k-1}|\Sigma, \theta_{it})} \tag{A-19}$$

If that is the case, I accept the trial value and $\xi_{it}^k = \xi_{it}^{trial}$; otherwise, $\xi_{it}^k = \xi_{it}^{k-1}$.

It is evident that draws of the M-H algorithm will be correlated; I follow the literature in thinning the chain (retaining one out of every ten draws) to ease this problem. The parameter η is a scalar specified by the researcher and adjusted in each iteration to keep the acceptance rate close to the "optimal" one specified by Gelman, Roberts, and Gilks (1996), which limits such correlation.

Draw of β^i

For each household i, the random coefficients vector β_i is drawn from the conditional distribution in equation A-10, which can be expressed as follows Each β^i's should be drawn from its conditional distribution

$$\mathbb{K}(\beta_i|y_i, X_i, b, W, \tilde{\beta}, xi_{it}) \propto L_i(y_i, X_i|\beta_i, \tilde{\beta}, \xi_{it})\phi(\beta^i|b, W) \tag{A-20}$$

where L_i is the likelihood function for the household and ϕ is the normal density assumed as prior. Once again, drawing directly from this posterior it is not possible and we have to use a M-H algorithm. The procedure is analogous to the one described in the previous step.

Draw of b and W

Conditional on all the β_i's and W, b is normally distributed with mean $\bar{\beta}$ and variance/covariance matrix $\frac{W}{N}$), where $\bar{\beta} = \frac{1}{N}\sum \beta_i$.
Conditional on all the β_i's and b, W is distributed as an Inverse Wishart with $N+2$ degrees of freedom and inverse scale matrix $\frac{2*I+N*S}{2+N}$). N is the number of households in the sample and $S = \frac{1}{N}\sum(\beta^i_t - b_t)(\beta^i_t - b_t)'$.

Draw of $\tilde{\beta}$

The vector of fixed coefficients $\tilde{\beta}$ is drawn by the conditional density in equation A-13, which can be expressed as proportional to the likelihood of the pooled data times the (normal) prior on the vector of coefficients. If the variance of the prior is sufficiently large (flat prior) we can assume

$$\mathbb{K}(\tilde{\beta}|\beta_i, xi_{it}) \propto L(y|\beta, \xi_{it}) \tag{A-21}$$

Draws from this density have to be approximated using the M-H algorithm.

Draw of Σ

The variance-covariance matrix of the bivariate normal in equation (5) is drawn from an inverse Wishart distribution with 2+R degrees of freedom and scale matrix $\frac{2*I+R\hat{\Sigma}}{2+R}$, where $\hat{\Sigma}$ is the variance matrix implied by the draws of θ_{it} and ξ_{it} in that iteration.

Computation of δ

In every iteration, we can compute $\delta = 1/R\sum_i\sum_t(\xi_{it} - \rho\theta_{it})$, where R is the total number of trips in the sample and ρ is the correlation between ξ and θ.
In the specifications where δ is parameterized as in equation (6), the parameters (α's) are recovered through Bayesian regression.

$$\delta = \alpha_0 + \alpha_1 Internet + \alpha_2 Weekend + o \rightarrow \delta = \alpha' Z + o, \quad o \sim N(0, \sigma_o^2) \tag{A-22}$$

We know that $\delta_{it} = (\xi_{it} - \rho\theta_{it})$, put a normal prior on α and proceed as follows

- Draw σ_o^2 from an inverse chi-square with parameters $R - 3$ (with three being the dimension of the vector of coefficients and R the number of observations) and $s^2 = \frac{(\delta - Z\tilde{\alpha})'(\delta - Z\tilde{\alpha})}{(R-3)}$
- Draw α from $N(\tilde{\alpha}, \sigma_o^2(Z'Z + A)$, where $\tilde{\alpha} = (Z'Z + A)^{-1}(Z'Z\hat{\alpha} + A\bar{\alpha})$

$\hat{\alpha}$ is the vector obtained by plain OLS in the model in A-22, $\bar{\alpha}$ is the initial prior and the matrix A determines the weight put on the prior. I choose a diffuse prior by picking $A = 0.1 * I$.

References

Aguiar, M., and E. Hurst (2007): "Life-Cycle Prices and Production," *American Economic Review*, 97(5), 1533–1559.

Broda, C., and D. Weinstein (2010): "Product Creation and Destruction: Evidence and Price Implications," *American Economic Review*, 100(3), 691–723.

Einav, L., E. Leibtag, and A. Nevo (2010): "Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?," *Quantitative Marketing and Economics*, 8(2), 207–239.

Gelman, A., G. Roberts, and W. Gilks (1996): "Efficient Metropolis jumping rules," in *Bayesian Statistics 5*, ed. by J. Bernardo, J. Berger, A. Dawid, and A. Smith. New York: Oxford University Press.

Nevo, A., and C. Wolfram (2002): "Why Do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals," *RAND Journal of Economics*, 33(2), 319–339.