STRUCTURAL DEMAND ESTIMATION WITH VARYING PRODUCT AVAILABILITY

Hernán A. Bruno¹
Naufel J. Vilcassim²

London Business School
Regent’s Park, NW1 4SA
London, United Kingdom.
ph:+44-20-7000-7000
fax: +44-20-7000-8601

January 24, 2008

⁰We would like to thank Céline Bonnet and conference participants at the INRA-IDEI “Empirical Industrial Organization of the Food Retailing Industry” for useful discussions. We also thank J.P. Dube, Bart Bronnenberg, and Pradeep Chintagunta for their helpful comments and especially Georg Toufar for making the data available. All errors remain our own.

¹Doctoral Candidate at London Business School, email: hbruno@london.edu
²Professor of Marketing, London Business School
STRUCTURAL DEMAND ESTIMATION WITH VARYING PRODUCT AVAILABILITY

Abstract

Demand estimation based on discrete choice theory has become an important tool in marketing and empirical Industrial Organization. In particular, estimation techniques have been developed that allow the use of individual-based discrete choice methods in situations where only aggregate data are available (e.g. Berry et al. 1995). In essence, these methods allow making inferences on the distribution of individual preferences over products or attributes by integrating purchase probabilities over heterogeneous individuals and relating them to the aggregate market shares observed in the data. However, if not all products were available in every store and/or on every purchase occasion, the observed market share will be a convolution of two different –albeit related– factors: consumer choice and the probability of finding the products available in the store. Failing to account for the varying degree of availability would produce incorrect demand parameter estimates and would lead to wrong inferences about consumer preferences, competitive reactions, etc.

This paper develops a model that extends the current methodology to account for varying levels of observed product availability, although the actual store/trip product assortments faced by the consumer are not observed by the researcher. The model parameters are estimated by simulating potential product assortment vectors by drawing multivariate Bernoulli vectors consistent with the observed aggregate level of availability. We show that our model is a generalization of the traditional random coefficients multinomial logit and similar estimation methodologies can be used. Furthermore, results from a Monte Carlo experiment show that utility parameters can be recovered correctly using the proposed estimation method, notwithstanding what causes product availability to vary. The model is applied to the UK chocolate confectionery market, focusing on the convenience store channel. We compare the parameter estimates to those obtained from not accounting for varying availability and analyze the pricing and competitive implications.

Keywords: Structural Modeling, Availability, Retailing
1 Introduction

Demand analysis has been recognized as the “most basic tool of empirical Industrial Organization” (Ackerberg et al. 2005) and has become an essential building block of comparative static analysis of market changes and their implications on firms’ actions, consumer welfare and public policy. Furthermore, understanding the strategic factors that affect demand has provided an important forecasting tool to assess the effects of marketing actions such as product introductions, category management or different levels of advertising investment. In particular, the development of demand estimation methods based on individual-level discrete-choice models using aggregate data has allowed researchers to infer the distribution of consumer preferences by observing aggregate realizations of demand, product and demographic characteristics, and other external factors affecting demand (Berry et al. 1995).

In discrete-choice based aggregate-data demand models, consumers are assumed to have preferences that can be described by a quasilinear indirect utility function over products or product characteristics. Estimates of the parametric distribution of utility can be obtained by making use of a convenient one-to-one relationship between the mean utility level and the observed market share (Berry 1994) and a given numerical inversion relationship (Berry et al. 1995), together with simulation methods (Pakes 1986). This estimation strategy has been extremely popular and has been applied to different settings in Industrial Organization and Marketing by academics and practitioners alike. Nevo (2000, 2001), for example, studied the ready-to-eat cereal industry using a random coefficients logit model and obtained demand estimates and their relation to demographic characteristics of markets. This, in turn, allowed him to conduct policy simulations regarding the ownership structure of the industry. Aggregate data discrete choice models have also been used to study the strategic behaviour of players in a vertical channel (Sudhir 2001, Chintagunta 2002), personal computer demand (Hui 2004) and detailing behaviour (Chintagunta and Desiraju 2005).

One key underlying assumption of aggregate choice models is that of complete availability of all products over the consumer choice occasions. This assumption, which may be realistic in certain choice situations (e.g. mobile telephony services), may not hold in other situations (e.g. grocery products). With market level data (cf. Dube 2004), where sales are aggregated
across households and stores, part of the observed variation in market shares could be explained by variation in distribution coverage, as well as assortment variation or stock-outs in different retailers or zones. In such cases, one would be ascribing differences in market shares to variations in consumers’ taste and behaviours, even though there might be a significant contribution to the observed variance from different levels of availability.

Consider a category with three products: A, B and C. While A and B are available in every single shopping trip, C is only available with a 60% probability. Evidently, 60% of the shopping trips would involve a choice over three alternatives and the remaining 40% would involve a choice over two alternatives. If one ignores this varying level of availability, the difference in market share would be ascribed to differences in preferences and would be captured by the product preference parameters. This would generate a bias in the estimates for the characteristics (or the product intercept) of product C. In addition, it would also bias the estimates for the inherent preferences for the other products (because they would seem relatively more ‘desirable’ than they really are) and the price coefficient (because changes in distribution may be captured by variation in prices). This problem is made worse if preferences are heterogeneous and the availability of product C changes across markets or over time.

Previous studies have proposed demand estimation methods and brand choice models that incorporate availability information using individual level data (which in this context means observing all available alternatives and the chosen alternative by the individual agent). One of the first studies to recognize that demand is a combination of both choice and availability was proposed by Jeuland (1979), who integrated a model of availability and a model of multibrand choices by heterogeneous consumers. In random utility model with transaction data at the individual-level, correcting for availability amounts to using the appropriate choice set for each customer purchase (e.g. Campo et al. 2003). However, if we only observe aggregate data, we do not directly observe individual choices or the set of alternatives available at the individual shopping occasions.

The issue of varying product availability has been absent from the reviews of the literature on structural modeling (see Dube et. al. 2005 Chintagunta et al. 2005, Kadiyali et al. 2001). One possible exception to this is the work of Tenn (2006). Although Tenn’s work deals with the issue of store heterogeneity in promotions, its model could be applied to the problem of varying
product availability provided some of the assumptions made there are applicable. Specifically, Tenn’s analysis uses the fact that most of the time only one product is promoted in the product category analyzed (premium ice cream), which allows him to identify the marginal distribution of promotional activity. Applying this assumption to the situation of varying product availability would require that at any point in time, only a single product has less than 100% availability – something that clearly does not hold in our data. An additional assumption is that the promotions are the same across all SKU’s of a given brand. In many categories where size or product form effects are present (for example, when only one size (32 oz.) or one form (powder) of a detergent is on promotion), so that aggregating across SKUs to the brand level is not appropriate, applying Tenn’s model would be computationally challenging, if not infeasible. We contend that the method proposed here would be more suitable for the situation with many products (both brands and SKUs) in the category as it is computationally simpler. The objective of this paper is thus to present a tractable model that can account for varying levels of product availability and which can be easily estimated a la BLP.

The contribution in this chapter is to (i) recognize that ignoring varying levels of availability when using aggregate-level data as mentioned above can lead to incorrect inferences about consumer preferences and, more importantly, (ii) provide a method to incorporate aggregate measures of product availability to statistically correct the demand estimates. Our model can be considered an extension of the well-known discrete choice based models with aggregate data. The main challenge is that, although we observe the aggregate level of distribution, we do not observe the set of products available in a particular store. Hence, we use simulation methods in order to model the different possible assortment of products that individual stores may have had on a given shopping trip.

Accounting for product availability can be approached in at least two different ways. First, the retailer decision to stock a product can be endogeneized and the vertical arrangement between the manufacturer and the retailer can be modeled explicitly in terms of parameters to be estimated. The data can then be used to infer the nature of the vertical relationship and the stocking behaviour of retailers. However, explicitly defining the vertical game would introduce further modelling assumptions regarding the retailer’s assortment and inventory decisions that generate the observed levels of availability. For instance, we may assume that retailers strate-
gically set the inventory levels taking into consideration the stockout probability. Alternatively, one might assume that retailers assort products by maximizing variety or profits, given a shelf-space constraint. Taking this approach would call for information on shelf space opportunity costs or shelf space capacity constraints. Hence, endogeneizing availability in an equilibrium framework in the vertical channel would entail non-trivial challenges in terms of modelling and data requirements.

We take a second approach and provide a method that uses the observed levels of product availability to correct for the influence of varying availability on market share. Taking this reduced-form approach to the problem allows us to obtain demand estimates, given the observed data on product characteristics, prices, market shares and aggregate availability, without having to address the challenges mentioned above. Under the conditions of our analysis, we are able to rely on the main results of discrete-choice demand analysis as in Berry et al. (1995).

The aggregate availability pattern observed in the data can be the result of several factors. First, stores make assortment decisions and more highly preferred products may be more widely available. Empirical studies have shown a convex relationship between distribution and market share (Reibstein and Farris 1995). That is to say, highly distributed products also have higher market shares per point of sale. In general, there seems to be a two-way causality in which market share and distribution positively affect each other. Second, the inverse can also be true: more popular products can be out-of-stock, leading to lower levels of availability. Finally, in the case of new products, retailers may be uncertain about the demand of the new product and the decision to carry the product can be influenced by the decision of other stores (Bronnenberg et al. 2000 and Bronnenberg and Mela 2004).

Although we do not model the vertical decisions, and thus remain agnostic about what drives the availability pattern in the data, we show that the proposed estimation method can yet recover the utility parameters correctly. In a series of Monte Carlo experiments we generate synthetic datasets in which product availability varies according to patterns that are consistent with

---

1 The assortment decision by a retailer has been studied extensively and there is general agreement that retailers assort products with high levels of sales, reliable levels of quality, acceptable category growth and strong advertising investments (Grashof 1970, Heeler et al. 1973, Montgomery 1975, Nilsson and Host 1987, Rao and McLaughlin 1989, Davis 1994)

2 Some studies have addressed included stockouts in the models, e.g. Anupindi et al. (1998), Campo et al. (2000), Campo et al. (2003). Aguirregabiria (2005) has modeled the stockout probability as a strategic decision.
several possible retailer or manufacturer behaviour. For instance, availability can be positively or negatively correlated to the mean utility of the product. For each of those cases we show that we can correctly recover the mean utility.

To illustrate the availability issue and apply our model, we analyze data from the chocolate confectionery industry in the United Kingdom (UK). The UK market for chocolate is characterized by high per capita consumption and continuous new product introductions. Confectionery is sold through two main channels: large grocery stores (e.g. large supermarkets and hypermarkets) and small convenience stores (referred to as the “impulse” channel). These two channels are, in general, independent in the sense that consumers’ decisions to buy in one type is not likely to affect purchase decisions in the other. We focus on the impulse channel, which accounts for half of UK national sales. Our results indicate that not accounting for variations in product availability results in a significant bias in the demand parameter estimates. This effect is more pronounced for products with low levels of distribution. These biases lead to inaccurate price elasticities and a wrong assessment of the competitive environment.

In the next section we describe the UK confectionery market and provide more illustrations of the effect of changing product availability. In Section 3 we develop the model and estimation method and carry out some Monte Carlo experiments to determine the ability of the proposed estimation method to correctly recover the mean preference vector. The results of the estimation are given in section 4, together with some price elasticity analysis and computation of optimal markups. Finally, a discussion and concluding remarks are presented in section 5.

2 Accounting for Varying Levels of Availability

Observed and Conditional Market Shares

We model a market with $J$ partially differentiated competing products. Consumers can choose any one of the $J$ alternatives with respective prices $p_1, \ldots, p_J$ or choose not to purchase any (i.e. they have an outside option). At time $t$, the choice of alternative $j = 1, \ldots, J$ with a characteristics vector $x_{jt}$ provides consumer $h$ with a level of utility equal to

$$u_{hjt} = x_{jt} \beta_h - \alpha_h p_{jt} + \xi_{jt} + \epsilon_{hjt}.$$
The parameter vector $\beta_h$ and the parameter $\alpha_h$ capture, respectively, the taste for the product characteristics and the price sensitivity for consumer $h$. The term $\xi_{jt}$ captures unobserved product characteristics or unobserved market-level demand shocks—depending on the definition of $x$. We assume that the no-purchase option ($j = 0$) provides a utility level of zero. If $\epsilon$ is extreme value distributed the probability choice follows the logit form (McFadden 1974).

Heterogeneity in preferences among consumers is modeled by assuming that the variation in individual parameters follows a probability distribution such as a continuous parametric distribution, or a non-parametric mass-point distribution. To simplify the notation, we will write the individual-level parameters in terms of their mean across individuals and an individual-level deviation (noted as realization of the random variables $\nu_\alpha$ and $\nu_\beta$). Hence, we will write the parameters as $\alpha_h = \alpha + \sigma_\alpha \nu_\alpha$ and $\beta_h = \beta + \Sigma \nu_\beta$, where the scalar $\sigma_\alpha$ and the matrix $\Sigma$ are parameters. Additionally, we will write the mean utility level (that is constant across consumers) as $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ and collect the heterogeneity terms in $\mu_{hjt} = x_{jt}\Sigma \nu_\beta - \sigma_\alpha \nu_\alpha p_{jt}$.

Conditional on the parameters, the logit market share for product $j$ is the integral over the heterogeneity $\nu = [\nu_\alpha, \nu_\beta]$,

$$
\tilde{S}_{jt} = \int \frac{\exp[\delta_{jt} + \mu_{jt}(\nu)]}{1 + \sum_{k=1}^J \exp[\delta_{kt} + \mu_{kt}(\nu)]} dG(\nu; \sigma_\alpha, \Sigma).
$$

The previous expressions—though ubiquitous in the literature—rely on the assumption that all products are available when consumers make the purchase decisions. The notation $\tilde{S}_j$ represents the “share” conditional on complete availability, in contrast to the unconditional share $S_j$, defined below.

We let $a_{jt}$ denote the availability of product $j$ during purchase occasion $t$. The variable $a_{jt}$ can only take two values: 1 if the product is available and 0 if the product is not available. The set of products available will therefore be denoted by the $J$-dimensional vector $a = (a_{1t}, \ldots, a_{Jt})$. We further indexed $a_{hjt}$ to denote the availability of product $j$ at time $t$ for consumer $h$ (or more precisely, consumer-shopping trip). The expectation over preferences $\tilde{S}(a_t)$ is the share conditional on $a_t$ and hence a function of the availability vector,

$$
\tilde{S}_{jt}(a_t) = \int \frac{a_{jt} \exp[\delta_{jt} + \mu_{jt}(\nu)]}{1 + \sum_{k=1}^J a_{kt} \exp[\delta_{kt} + \mu_{kt}(\nu)]} dG(\nu; \sigma_\alpha, \Sigma).
$$
The observed market share is the expectation of the expression above over all possible availability vectors. That is, the share can be written as an expectation over all values of \(a\),

\[
S_{jt} = \sum_{\text{all } a} \tilde{S}_{jt}(a_t) \pi(a_t).
\] (2)

In equation (2) “all \(a\)” denotes all possible combinations\(^3\) of \(a_j \in \{0, 1\}\) and \(\pi(a)\) is the joint probability of observing the vector \(a\). In summary, the observed share is the result of the averaging over all the shopping trips consumers made in period \(t\), each of which involved the choice over a particular realization of the availability vector \(a\).

### An Illustration of the Availability Bias

To show the effect of different levels of availability in the estimation, we briefly revisit the example from the Introduction. Consider a market of homogeneous consumers that face a choice over alternatives A, B and C. Assume alternative A is the no-choice or outside good alternative that, naturally, is always available. As before, assume alternative B is always available and alternative C is available with probability \(\pi_C = 0.6\).

If we proceed to estimate a logit demand model regardless of the different levels of availability, we would write down market shares (we use again \(\tilde{S}\) to denote the share if all products were always available) as

\[
\tilde{S}_B = \frac{e^{\delta_B}}{1 + e^{\delta_B} + e^{\delta_C}},
\]

\[
\tilde{S}_C = \frac{e^{\delta_C}}{1 + e^{\delta_B} + e^{\delta_C}},
\]

and would then follow Berry (1994) and estimate the utility parameters as

\[
\ln \left( \frac{\tilde{S}_C}{\tilde{S}_A} \right) = x_{jt}\beta - \alpha p_{jt} + \tilde{\xi}_{jt}.
\] (3)

This estimation would yield a lower product intercept for product C than the one we would

\(^3\)Clearly, there are \(2^J\) such vectors and an equal number of terms in the summation in (2)
obtain if C’s distribution level were 100%. Conversely, the product intercept for product B would be higher. To overcome this we can write the expectation as in equation (2) over two possible availability vectors \{1,1\} and \{1,0\}. Since we are assuming that \(\pi_C\) is observed, we are not introducing new parameters but simply incorporating the availability information to correct for the market shares.

\[
S_B = \left[\frac{e^{\delta_B}}{1+e^{\delta_B}+e^{\delta_C}}\right]\pi_C + \left[\frac{e^{\delta_B}}{1+e^{\delta_B}}\right](1-\pi_C)
\]

\[
S_C = \left[\frac{e^{\delta_C}}{1+e^{\delta_B}+e^{\delta_C}}\right]\pi_C.
\] (4)

The equation above provides a relationship between market shares and utility parameters and the econometric estimation of \(\alpha\) and \(\beta\) can then proceed. However, this model cannot be linearized for estimation in a straightforward way. The corresponding equation using the market shares and the availability probabilities for product C is

\[
\ln\left(\frac{S_C}{S_A}\right) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} - \ln\left[\frac{1+e^{\delta_B}+e^{\delta_C}(1-\pi_C)}{(1+e^{\delta_B})\pi_C}\right]
\] (5)

and a slightly more cumbersome expression for product B. It should be noted that if we do not use information on the product availability \(\pi_C\), the last term in (3) is also part of the unobserved term. We would then be estimating a relationship as in (3) with

\[
\tilde{\xi}_{jt} = \xi_{jt} - \ln\left[\frac{1+e^{\delta_B}+e^{\delta_C}(1-\pi_C)}{(1+e^{\delta_B})\pi_C}\right].
\]

It is usually assumed that the unobserved product characteristics \(\xi\) are correlated with price as a result of the economic assumptions on the price setting behaviour by firms. In the expression above, we see that not correcting for varying availability would induce a correlation between the unobserved \(\xi_{jt}\) and all the regressors and the average availability in a non-linear way. Thus, traditional instrumental variable approaches that assume that price is the only endogenous variable would still yield biased estimates.
Furthermore, rewriting the expression above with (3) and (5) we obtain

\[
\ln \left( \frac{\tilde{S}_C}{\tilde{S}_A} \right) = \ln \left( \frac{S_C}{S_A} \right) - \ln \left[ \frac{1 + e^{\delta_B} + e^{\delta_C} (1 - \pi_C)}{(1 + e^{\delta_B}) \pi_C} \right].
\]

The last expression makes it clear that the observed market shares differ from the market shares given complete availability by a term that depends on the availability of products in a non-linear way. Hence, we cannot simply account for variations in availability probability by including \( \pi_C \) as an additive term. Shares depend on the availability of all the products. Therefore, it would be necessary to add availability levels for all products as explanatory variables, but that would still bring about endogeneity issues that need instrumentation. For this same reason, there is no simple way to “normalize” shares using availability information. For example, practitioners and market research firms use a ratio called rate of sales as a metric for analysis which is merely market share divided by distribution (\( S_j/a_j \) in our notation). Although this would partially compensate for cases where observed market share is driven mostly by the lack of distribution, it does not address the issue that the distribution of a product affects the sales of the competing products as well.

We argue that a correct way to recover the utility parameters would be to recover the \( \delta_B \) and \( \delta_C \) from the simultaneous equations in (4). This can be done numerically. In the more general case in which we observe the market shares \( S_1, \ldots, S_J \) and the product availability probabilities \( \pi(a) \) the demand system defines the following system of equations

\[
S_j = \sum_{a} \frac{a_j \exp(\delta_j)}{1 + \sum_{k=1}^{J} a_k \exp(\delta_k a_k)} \pi(a)
\]

The sum in (6) could potentially contain a larger number of terms, but it is always finite. It can easily be shown that such system satisfies the same assumptions found in Berry (1994) and there is indeed a one-to-one relationship between market shares \( S_j \) and the mean utilities \( \delta_s \).

However promising this approach to estimation may sound, it is clear that as the number of products \( J \) becomes large, the number of terms in the sum in (6) increases exponentially. Moreover, the discussion so far has not taken into consideration that it would be desirable to account for unobserved heterogeneity as well. Computing the integral in (2) can become computationally burdensome because we need to integrate over the unobserved heterogeneity.
and the unobserved vectors of availability. In the following section we describe a method that makes use of simulation to correct for availability in a computationally feasible way.

**Estimation Strategy**

The objective of the proposed methodology is to introduce information on the observed levels of aggregate availability in the computation of the market shares. In essence, we use the aggregate level of availability to approximate the probability of finding a particular product and thus compute the expectation in equation (6). The logit estimation of demand involves finding parameters that minimize a convenient measure of distance between the observed and the computed market shares. Berry et al. (1995) proposed using an empirical distribution of pseudo-random draws to approximate the population density and compute the expectation over consumer preferences can be computed by summing over the random draws.

In the traditional method, each draw is a particular realization of the multivariate distribution of preferences. The method we propose here involves generating vectors of product availability by taking random draws from the empirical distribution of availability. These availability vectors are then used in addition to the utility random draws to compute the market shares using equation (6). We note that this approach does not involve including additional parameters into the model.

Our data has information that can be used as a proxy for the probability of availability for each product \( \pi(a_1), \ldots, \pi(a_J) \) for each market/time-period. However, we do not observe the joint probability of the multivariate Bernoulli vector of availability \( \pi(a) = \pi(a_1, \ldots, a_J) \). In order to compute the joint probability we need to make further assumptions about the correlation among the \( a_j \)s. We assume that \( \pi(a_j) \) is independent of \( \pi(a_k) \) for \( j \neq k \). That is to say, the probability of one product being sold-out or simply not distributed is independent of whether other products are available. This would be an acceptable assumption under some conditions and we will address these issues below.

Let \( \tilde{a}_i = (\tilde{a}_i^1, \ldots, \tilde{a}_i^J) \) denote the i-th multivariate Bernoulli draw with marginal distributions given by \( \pi(a_1), \ldots, \pi(a_J) \). We can use this draw to approximate the sum of integrals in equation (2). In the spirit of Berry et al. (1995) and Nevo(2001) we can use two sets of \( I \) random draws \( \tilde{a}_i \) and \( \nu^i = (\nu^i_\alpha, \nu^i_\beta) \) to compute the sum

12
$$\tilde{S}_{jt}(\delta) = \frac{1}{I} \sum_{i=1}^{I} \frac{\bar{a}_{jti} \exp[\delta_{jt} + \mu_{jt}(\nu_i)]}{1 + \sum_{k=1}^{J} a_{kti} \exp[\delta_{kt} + \mu_{kt}(\nu_i)]}.$$  \hfill (7)

In the simplest case described in the previous section, where product B was fully available and product C was only available with 60% probability, we would draw \{1, 1\} with probability 0.6 and \{1, 0\} with probability \(1 - 0.6 = 0.4\) and then sum over the \(I\) draws.

By equating the share computed in (7) to the observed shares we obtained a system of equations analogous to those in Berry(1994), Berry et al. (1995) and Nevo (2001). If we can use a contraction mapping of the form

$$\delta^{(t+1)} = \delta^{(t)} + \log \left( S^{\text{obs}} \right) - \log \left( S(\delta^{(t)}) \right),$$ \hfill (8)

to solve iteratively for the mean utility levels \(\delta_j\) we could then proceed to define a suitable GMM estimator. We would need to show that (7) falls under the sufficient conditions outlined in Berry (1994) and Berry et al. (1995).

**Proposition.** The recursive relation in (8) defines a contraction mapping with modulus less than one.

The proof closely follows Berry et al. (1995) because the sum to compute the share of product \(j\) in (2) can always be written as the sum for all the terms in which product \(j\) is available, i.e. \(a_j = 1\). The derivatives of this sum behave in a very similar way to the derivatives of the market share function defined in Berry et al. (1995) and the proof can be followed step by step. In essence the proposition holds if the signs of the derivatives of market share obey some restrictions and some boundedness conditions can be fulfilled. The details are shown in the appendix.

This approach may resemble that of modelling consideration sets in consumer choice behaviour (e.g. Andrews et al. 1995, Mehta et al. 2003). Although related in spirit, there are very important differences between the two approaches. Introducing consideration set formation into a logit modelling allows for further econometric flexibility by probabilistically modelling the (unobserved to the econometrician) subset of brands that a consumer chooses from when making a purchase. Our approach, on the other hand, attempts to correct for the different levels of availability observed in the data. We are providing a method to use the availability information...
to obtain more accurate estimates of the utility parameters, but we are not introducing more parameters to be estimated, nor is our model intended to be more flexible than the traditional logit model when fitting the observed data. Finally, it should be noted that a consideration set formation models cited above rely on household level data and to the best of our knowledge there is no evidence that this type of analysis is feasible with aggregate level data.

**Monte Carlo Experiment**

In this section, we generate synthetic datasets to test whether the proposed method can recover parameter estimates for different availability conditions. It should be noted that the model is an extension of the random coefficients logit that corrects for availability when computing the mean utility levels. The model will produce unbiased results if it can recover the mean utility levels, and the estimation can then follow the traditional path. In this section we propose four different availability scenarios and assess whether the recovery of the mean utility is robust with respect to the nature of availability.

We consider a set of five alternatives from which a consumer can choose. The utility of alternative $j$ for individual $h$ at time $t$ is modeled as $u_{hjt} = \beta_{hj} + \alpha_{jh} p_{jt} + \xi_{jt}$. The intercept for each alternative varies across individuals and is a random normal draw with respective means $\bar{\beta} = (-1.0, -0.5, 0.0, 0.5, 1.0)$ and variance equal to 0.5. The mean price coefficient is $\alpha = -1$, with variance equal to 0.5. The term $\xi_{jt}$ is a i.i.d random draw distributed $N(0, 0.05)$. We assume that the market shares are observed for $T = 100$ periods. Availability marginal distributions $\pi(a_j)$s are produced in different ways as outlined below. In all the scenarios analyzed in this section, brand 5 is only “launched” at time period 50 and is given low availabilities (.15 and .3) in weeks 51 and 52. From week 53 onwards, the availability for brand 5 follows the same pattern as the corresponding availability of the other four brands as outlined below.

We use the mean utility level $\delta_{jt} = \beta_j + \alpha_j p_{jt} + \xi_{jt}$ (the mean utility levels is independent of $h$) to generate the prices. We want the prices to be positively correlated with the mean utility level. This correlation simulates the strategic behavior of manufacturers, who set prices based on the utility that the product provides to consumers. We use the relationship $p_{jt} = \frac{1}{2} e^{\delta_{jt}/2}$, which has no theoretical justification except that it produces realistic prices for our setting (i.e. prices that generate reasonable market shares given the choice of coefficients). The exponential
function produces prices that increase more than proportionately with utility (i.e. “quality”) which is also an appealing property.

To generate the market shares, a large set of \( N_d \) “individuals” are drawn and the utility levels computed. In the examples described below, we use \( N_d = 5000 \). In order to compute the observed shares, we first compute the probability of observing any of the \( 2^5 \) availability vectors for every time period. Then, the observed shares are obtained by first computing the share using the logit formula for all these different combinations of the availability vector. The “observed” share is computed as a weighted average using the value of \( \pi(a) \). We note that the data generation uses the exact expectation of shares over the availability and not the simulation we propose as an estimation strategy. This is feasible since we do not need to repeat the computation every time we evaluate a moment condition and the synthetic dataset has only a few products. The estimation will be performed, however, using draws from the multivariate bernoulli distribution as was outlined before, as computing the exact expectation would be computationally burdensome.

We explore four scenarios in which we make different assumptions of the process generating the observed availability and the objective is to determine how well the proposed estimation method recovers the mean utility irrespective of what cause the availability variation.

1. **Constant availability, varying across products.** This is one of the simplest situations in which biases can occur due to availability. Availability is kept constant for each product but products differ on their level of availability. In this particular experiment, the availability of products 1, 3 and 5 is 80% and the availability of products 2 and 4 is 50%.

2. **Availability varying across products and over time.** It is common for consumer goods categories to have non-constant levels of availability. This could be due to many reasons, namely, varying degrees of distribution effort, unexpected peaks in demand that leave retailers temporarily without stock or sluggish rates of adoption for new products. We model distribution as independent random draws from the uniform distribution (independence or the actual distribution is immaterial for this example, as any sequence of numbers within the unit interval would produce similar results). Products 1, 3 and 5 will have availabilities in the range \((0.6, 0.9)\) and products 2 and 4 will have availabilities in the range \((0.4, 0.7)\).
3. **Availability increasing with $\delta$.** As we mentioned in the introduction, the decision of a retailer to stock a product is sometimes related to the level of sales of the product. In this scenario we generate aggregate availability data that is proportional to the mean utility $\delta$ by simply using the formula\(^4\) \(a_{jt} = \frac{e^{\delta_{jt}}}{1+e^{\delta_{jt}}}\). This distribution pattern could be interpreted in a stylized way as the decision of the retailer to stock or not to stock based on the mean level of utility that the product has in the eyes of consumers. As we mentioned, the proposed estimation method should recover the parameters for any distribution pattern.

4. **Availability decreasing with $\delta$.** This scenario accounts for cases in which more popular products are less likely to be available. Although it is not less realistic than the previous scenarios, we produce this dataset to show that the proposed method can recover the parameters even if availability is inversely proportional to the product’s popularity.

   For each one of these four scenarios, we generate 20 synthetic datasets. We then estimate the brand intercepts, the price coefficient and the corresponding heterogeneity parameters. We estimate the parameters using 100 random draws for the heterogeneity in preferences and assortments. The heterogeneity draws are different than the draws used to construct the dataset. Each dataset is estimated using (i) the standard (labeled ‘standard’) random coefficients logit and (ii) the random coefficients logit corrected (labeled ‘corrected’) for availability that we propose in this chapter. The average of the estimates under each of the four scenarios of availability (1-4) are shown in tables 1 and 2.

   The first thing we notice from table 1 is that the proposed estimation method satisfactorily recovers the parameters for all four data generation processes. The standard method, on the other hand, underestimates the parameters in all cases. The poor performance of the standard method is a result of the non-complete availability of products, which affect market shares and biases the parameter estimates.

   Although our estimation model assumes that the availabilities are uncorrelated, this may not be the case in some situations. We explore how the Monte Carlo experiment changes when availabilities are indeed correlated. We generate 20 datasets as in the Monte Carlo exercise

\(^4\)In this context this “logit” formula provides a useful functional form and it is not related to utility maximization. The endogenous optimal decision to stock a product is the subject of future research.
### Table 1: Estimates for the standard and the corrected model for the four different scenarios.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Estimate</th>
<th>Standard 1</th>
<th>Standard 2</th>
<th>Standard 3</th>
<th>Standard 4</th>
<th>Corrected 1</th>
<th>Corrected 2</th>
<th>Corrected 3</th>
<th>Corrected 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Brand 1 (True=-1)</td>
<td>-1.29 (0.22)</td>
<td>-1.43 (0.22)</td>
<td>-1.64 (0.14)</td>
<td>-1.54 (0.11)</td>
<td>-1.03 (0.15)</td>
<td>-0.98 (0.14)</td>
<td>-0.93 (0.16)</td>
<td>-1 (0.15)</td>
<td></td>
</tr>
<tr>
<td>Het. 1</td>
<td>0.67 (0.17)</td>
<td>0.87 (0.14)</td>
<td>0.56 (0.13)</td>
<td>0.56 (0.13)</td>
<td>0.45 (0.11)</td>
<td>0.46 (0.13)</td>
<td>0.53 (0.19)</td>
<td>0.52 (0.19)</td>
<td></td>
</tr>
<tr>
<td>Het. 2</td>
<td>0.78 (0.15)</td>
<td>0.98 (0.14)</td>
<td>0.61 (0.14)</td>
<td>0.83 (0.14)</td>
<td>0.54 (0.14)</td>
<td>0.52 (0.14)</td>
<td>0.45 (0.16)</td>
<td>0.54 (0.16)</td>
<td></td>
</tr>
<tr>
<td>Brand 2 (True=-0.5)</td>
<td>-1.45 (0.16)</td>
<td>-1.37 (0.15)</td>
<td>-1.92 (0.17)</td>
<td>-1.33 (0.2)</td>
<td>-0.59 (0.16)</td>
<td>-0.48 (0.13)</td>
<td>-0.53 (0.19)</td>
<td>-0.51 (0.19)</td>
<td></td>
</tr>
<tr>
<td>Het. 3</td>
<td>0.31 (0.09)</td>
<td>0.93 (0.15)</td>
<td>0.33 (0.07)</td>
<td>0.71 (0.07)</td>
<td>0.51 (0.09)</td>
<td>0.54 (0.09)</td>
<td>0.49 (0.13)</td>
<td>0.49 (0.13)</td>
<td></td>
</tr>
<tr>
<td>Brand 3 (True=0)</td>
<td>-0.39 (0.18)</td>
<td>-0.6 (0.19)</td>
<td>-1.17 (0.18)</td>
<td>1.11 (0.2)</td>
<td>-0.06 (0.12)</td>
<td>-0.02 (0.12)</td>
<td>0.01 (0.13)</td>
<td>0.01 (0.13)</td>
<td></td>
</tr>
<tr>
<td>Het. 4</td>
<td>0.33 (0.08)</td>
<td>0.93 (0.08)</td>
<td>0.33 (0.13)</td>
<td>0.71 (0.13)</td>
<td>0.51 (0.09)</td>
<td>0.54 (0.09)</td>
<td>0.49 (0.13)</td>
<td>0.49 (0.13)</td>
<td></td>
</tr>
<tr>
<td>Brand 4 (True=0.5)</td>
<td>-0.51 (0.13)</td>
<td>-0.46 (0.2)</td>
<td>-0.39 (0.15)</td>
<td>-0.93 (0.17)</td>
<td>0.44 (0.14)</td>
<td>0.45 (0.12)</td>
<td>0.49 (0.12)</td>
<td>0.49 (0.12)</td>
<td></td>
</tr>
<tr>
<td>Het. 5</td>
<td>0.23 (0.06)</td>
<td>0.7 (0.16)</td>
<td>0.16 (0.13)</td>
<td>0.63 (0.13)</td>
<td>0.44 (0.09)</td>
<td>0.45 (0.09)</td>
<td>0.49 (0.13)</td>
<td>0.49 (0.13)</td>
<td></td>
</tr>
<tr>
<td>Brand 5 (True=1)</td>
<td>0.46 (0.15)</td>
<td>0.3 (0.13)</td>
<td>0.51 (0.12)</td>
<td>-0.68 (0.2)</td>
<td>0.93 (0.15)</td>
<td>1.07 (0.17)</td>
<td>1.01 (0.19)</td>
<td>1.03 (0.22)</td>
<td></td>
</tr>
<tr>
<td>Het. 6</td>
<td>0.22 (0.09)</td>
<td>0.52 (0.11)</td>
<td>0.5 (0.14)</td>
<td>0.55 (0.08)</td>
<td>0.41 (0.15)</td>
<td>0.48 (0.07)</td>
<td>0.47 (0.11)</td>
<td>0.51 (0.11)</td>
<td></td>
</tr>
<tr>
<td>Price (True=-1)</td>
<td>-0.91 (0.21)</td>
<td>-0.92 (0.15)</td>
<td>-1.41 (0.16)</td>
<td>-0.42 (0.17)</td>
<td>-0.97 (0.12)</td>
<td>-1.01 (0.14)</td>
<td>-0.99 (0.19)</td>
<td>-1.01 (0.19)</td>
<td></td>
</tr>
<tr>
<td>Het. Price</td>
<td>0.74 (0.21)</td>
<td>1.03 (0.15)</td>
<td>0.2 (0.16)</td>
<td>0.95 (0.17)</td>
<td>0.55 (0.12)</td>
<td>0.49 (0.14)</td>
<td>0.53 (0.19)</td>
<td>0.5 (0.19)</td>
<td></td>
</tr>
</tbody>
</table>

Described above with the added constraint that the availabilities of Brand 1 and Brand 2 are perfectly correlated. The results are shown in table 2.

In table 2 we see that, even in situations in which availabilities are correlated, the corrected method still produces more accurate estimates of the true parameters than the standard methods. Surprisingly, many of the estimates are not affected by the added correlation. Although we cannot claim that the method works for situations in which the availabilities are correlated (the model would be misspecified), the results from table 2 show that the model is flexible enough to accommodate availability patterns in which correlations are present.

We have claimed above that varying levels of availability affect the estimated price coefficient if the changes in market share (driven by changes in availability) are captured by the price coefficients. This problem is particularly problematic if availability is correlated with prices. From table 1 we see that the estimation of the price coefficient using the standard method is fairly accurate for the ‘constant’ (scenario 1) case and the ‘random’ case. However, the price
Table 2: Estimates for the standard and the corrected model for the four different scenarios when availability is correlated. In particular, we have made the availability of Brand 1 and Brand 2 identical (i.e. correlation = 1). Estimates for the mean and heterogeneity of the brand intercepts and price shown. The standard error for each estimate is show below the estimate between parentheses.

A coefficient obtained for scenarios 3 and 4 is far from the true value. In such cases, availability was built proportional to the mean utilities. Since prices are correlated with the mean utilities, the price coefficient is capturing changes in market share that are due to availability and not to consumer’s price sensitivity. On the other hand, the parameters obtained from the proposed corrected method seem are closer to the true value under all availability conditions.

This Monte Carlo simulation study shows that when product availability is not complete (less that 100%) estimation using the standard random coefficient logit method leads to biased parameter estimates. Our extension of the model can recover the mean utilities and hence produce unbiased parameters. In the next section we apply the method to a real dataset.
3 Application

Industry

We apply our estimation framework to characterize the demand system in the chocolate confectionery industry in the United Kingdom. The size of the market is GBP 3.5 billion = US$ 6.2 billion. The preferred (45%) product format are candy bars or "countlines" and the main industry players are Cadbury, Nestle and Mars.

Sales of confectionery products in the UK occur through two main channels: large grocery stores and smaller convenience or “impulse” retail stores. About half of the sales of confectionery products occur through convenience retailers, a rather high proportion in relation to other categories. Convenience stores are usually small groceries, tobacconists or newsagent which do not exceed 3000 sq. feet. Size is linked to regulation, that allow these stores to be flexible in their opening times if they do not exceed a certain maximum area. Total sales through convenience stores add up to more than GBP 20 billion a year. Around 70% of these convenience stores are non-affiliate independent stores and are less than 280 sq ft in area. The owners of this kind of convenience stores rarely own more than a few stores (typically one). Other convenience stores include petrol station forecourt stores, co-operatives and small chains. Although this type of retail outlet does not have the bargaining power that a major retail chain may have, they have very limited shelf space and are only likely to stock the more popular products.

A drop in distribution can have significant effects on sales as can be seen in figure 1. M&M and Smarties are two popular bite-sized chocolate products with very little differentiation in terms of physical attributes. The drop in sales seen around 2003 seems to coincide with falling distribution coverage and should be noted if we are estimating consumer demand. If we do not account for this drop in distribution, we could be making incorrect inferences about consumer preference and estimates of price sensitivity.

The impulse retail market for confectionery products in the UK has some attractive properties for studying the effect of distribution: high product diversity, high distribution variation, negligible income effects, no stockpiling behaviour, little price promotions and consumers are unlikely to change their consumption habits based on expectations of new product releases.
Figure 1: Sales and distribution of similar products

Data

The data were provided by IRI and cover 113 4-weekly period from June 1996 to February 2005 on confectionery products from the UK market. For each product, we observe national unit sales, national average price and national category weighted distribution. We defined the total market as the total units consumed of confectionery in the corresponding time period. We constructed the market share as the ratio of sales to the total market unit sales.

In measuring aggregate availability, we will use the all commodity weighted distribution (ACV) as the aggregate availability variable to recognize that not all stores have identical sales and traffic. We acknowledge that this is not the ideal measure of distribution (Dolan and Hayes 2005), since it is only a proxy for the probability of finding the product in a store in a given shopping trip. However, given that we are using National UK level data, this is the standard measure used by practitioners and researchers and it is widely available from market research companies. We also note that the data (including distribution) are aggregated at the brand level (See Little (1998) for a discussion of the issues of aggregation for sales and distribution measures). This is a standard procedure when the estimation is performed at the product level (as opposed to using characteristics based approach).

We expect that these shortcomings (common to many situations in which aggregate data are used) should be minimized by focusing on the impulse channels. These smaller stores are less likely to carry a large variety of SKUs. In case they stock a product, they tend to carry the single-pack, regular size, regular flavor SKU. We found that this is the case by observing data at the SKU level for the years 2002-2005. In addition, stores in the impulse channels are similar to one another in terms of sales volume and size, thus minimizing the aggregation bias in the
ACV computation.

Product characteristics were collected by direct observation of the products. We use four variables that indicate the format of the confectionery product (block, filler, indulgence and bite-sized). Filler and Indulgence formats are two types of countlines or chocolate candy bars. The two typical filler bars are Mars and Snickers. Indulgence products are usually smaller than filler bars and may contain waffle or soft nougat (e.g. Kit Kat, Mars 5 Little Ones, etc). This distinction is common among managers and the trade press and drives positioning and advertising decisions. In the UK, filler bars are sometimes advertised as “lunch replacement” food.

<table>
<thead>
<tr>
<th>Product</th>
<th>Manufacturer</th>
<th>Format</th>
<th>Weeks on market</th>
<th>Average availability</th>
<th>Average price (GBP)</th>
<th>Average share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cadburys Dairy Milk</td>
<td>Cadbury</td>
<td>Block</td>
<td>113</td>
<td>95%</td>
<td>0.53</td>
<td>9.0%</td>
</tr>
<tr>
<td>Cadburys Whole Nut</td>
<td>Cadbury</td>
<td>Block</td>
<td>113</td>
<td>86%</td>
<td>0.55</td>
<td>3.5%</td>
</tr>
<tr>
<td>Cadburys Fruit &amp; Nut</td>
<td>Cadbury</td>
<td>Block</td>
<td>113</td>
<td>90%</td>
<td>0.54</td>
<td>4.1%</td>
</tr>
<tr>
<td>Cadburys Caramel</td>
<td>Cadbury</td>
<td>Block</td>
<td>113</td>
<td>86%</td>
<td>0.36</td>
<td>4.6%</td>
</tr>
<tr>
<td>Flake</td>
<td>Cadbury</td>
<td>Indulgence</td>
<td>113</td>
<td>85%</td>
<td>0.34</td>
<td>3.7%</td>
</tr>
<tr>
<td>Snowflake</td>
<td>Cadbury</td>
<td>Indulgence</td>
<td>58</td>
<td>53%</td>
<td>0.37</td>
<td>1.5%</td>
</tr>
<tr>
<td>Dream</td>
<td>Cadbury</td>
<td>Block</td>
<td>109</td>
<td>29%</td>
<td>0.66</td>
<td>0.7%</td>
</tr>
<tr>
<td>Wispa</td>
<td>Cadbury</td>
<td>Indulgence</td>
<td>113</td>
<td>65%</td>
<td>0.33</td>
<td>2.1%</td>
</tr>
<tr>
<td>Galaxy</td>
<td>Masterfoods</td>
<td>Block</td>
<td>113</td>
<td>87%</td>
<td>0.47</td>
<td>5.1%</td>
</tr>
<tr>
<td>Galaxy Caramel</td>
<td>Masterfoods</td>
<td>Block</td>
<td>113</td>
<td>66%</td>
<td>0.34</td>
<td>2.1%</td>
</tr>
<tr>
<td>Galaxy Ripple</td>
<td>Masterfoods</td>
<td>Indulgence</td>
<td>113</td>
<td>70%</td>
<td>0.32</td>
<td>2.7%</td>
</tr>
<tr>
<td>M&amp;Ms</td>
<td>Masterfoods</td>
<td>Bite Sized</td>
<td>113</td>
<td>80%</td>
<td>0.47</td>
<td>3.6%</td>
</tr>
<tr>
<td>Maltesers</td>
<td>Masterfoods</td>
<td>Bite Sized</td>
<td>113</td>
<td>95%</td>
<td>0.54</td>
<td>8.0%</td>
</tr>
<tr>
<td>Mars</td>
<td>Masterfoods</td>
<td>Filler</td>
<td>113</td>
<td>99%</td>
<td>0.34</td>
<td>15.7%</td>
</tr>
<tr>
<td>Mars 5 Little Ones</td>
<td>Masterfoods</td>
<td>Indulgence</td>
<td>53</td>
<td>14%</td>
<td>0.35</td>
<td>0.3%</td>
</tr>
<tr>
<td>Kit Kat</td>
<td>Nestle</td>
<td>Indulgence</td>
<td>113</td>
<td>96%</td>
<td>0.34</td>
<td>11.8%</td>
</tr>
<tr>
<td>Kit Kat Chunky</td>
<td>Nestle</td>
<td>Filler</td>
<td>76</td>
<td>89%</td>
<td>0.35</td>
<td>6.7%</td>
</tr>
<tr>
<td>Kit Kat Kubes</td>
<td>Nestle</td>
<td>Bite Sized</td>
<td>18</td>
<td>63%</td>
<td>0.50</td>
<td>1.6%</td>
</tr>
<tr>
<td>Milky Bar</td>
<td>Nestle</td>
<td>Block</td>
<td>63</td>
<td>69%</td>
<td>0.25</td>
<td>4.0%</td>
</tr>
<tr>
<td>Milky Bar Munchies</td>
<td>Nestle</td>
<td>Bite Sized</td>
<td>38</td>
<td>40%</td>
<td>0.46</td>
<td>0.7%</td>
</tr>
<tr>
<td>Smarties</td>
<td>Nestle</td>
<td>Bite Sized</td>
<td>113</td>
<td>81%</td>
<td>0.41</td>
<td>3.7%</td>
</tr>
<tr>
<td>Double Cream</td>
<td>Nestle</td>
<td>Block</td>
<td>33</td>
<td>65%</td>
<td>0.49</td>
<td>1.4%</td>
</tr>
<tr>
<td>Snickers</td>
<td>Masterfoods</td>
<td>Filler</td>
<td>113</td>
<td>97%</td>
<td>0.35</td>
<td>10.8%</td>
</tr>
<tr>
<td>Snickers Cruncher</td>
<td>Masterfoods</td>
<td>Filler</td>
<td>45</td>
<td>38%</td>
<td>0.32</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

Table 3: Data summary

In table 3, we provide a summary of the dataset and we can see that most of the products were available for the whole period of observation. Products that were not available during the whole period, were launched after June 1996. No product in the dataset disappeared during the period of observation. The confectionery market is composed of thousands of different products with many different variants and SKUs. The dataset we are using aggregates across different SKUs and overall covers 70% of the market. Because of the highly fragmented nature of the category, small market shares (of the order of 1%) do not necessarily represent low levels of sales. Finally, notice that some products have significant market shares in spite of having low
It should be noted from the table 3 that some products have high levels of availability, although they are not in the market in the early weeks (because they were launched after June 1996). Other products, though, are present in the market for the whole period observation but show low levels of availability.

**Instruments**

It is well accepted that the unobserved term $\xi_{jt}$ is likely to be correlated with the price of the product. Because we are using product dummies, $\xi_{jt}$ would measure period-specific demand shocks that are not captured by the observed characteristics (e.g. advertising in period $t$, marketing activity of substitutable products included or not in the data). Though we are not directly modelling prices, the assumption we make is that prices are the sum of a term that depends on manufacturers marginal costs and retail costs and a term that depends on the unobserved demand shock $\xi_{jt}$. Six instruments are used:

**Input Prices** A price index for labor costs and spot prices for cocoa. These variables are common across brands.

**Price of products from own manufacturer** In order to capture marginal cost changes that may affect price, we average the prices for all the products from the product’s own manufacturer.

**Price of product of the same format** We expect that prices from other products within the same format will capture both, changes in marginal costs for products in the category and changes in retail costs that may affect the final consumer prices.

These variables were selected after carefully studying their explanatory power on price using hedonic regressions. Together, they explain around 80% of the observed variance in prices.

**Results**

Following Nevo (2001), we used product dummies and prices and allowed for heterogeneity in the taste for formats. To obtain the mean values of the taste for formats, we regressed the product
intercepts on the product characteristics to obtain the taste for characteristics. Estimation results for the product intercepts using the random-coefficient logit and the random-coefficient logit corrected for availability are shown in table 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>RC Logit (std. err)</th>
<th>Sales Information</th>
<th>RC Logit w/ availab.</th>
<th>Average availability</th>
<th>Average Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cadbury's Dairy Milk</td>
<td>-1.047 (0.606)</td>
<td></td>
<td>-1.130 (1.021)</td>
<td>95%</td>
<td>9.0%</td>
</tr>
<tr>
<td>Cadbury's Whole Nut</td>
<td>-1.874 (0.604)</td>
<td></td>
<td>-1.932 (0.177)</td>
<td>86%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Cadbury's Fruit &amp; Nut</td>
<td>-1.726 (0.595)</td>
<td></td>
<td>-1.794 (0.264)</td>
<td>90%</td>
<td>4.1%</td>
</tr>
<tr>
<td>Cadbury's Caramel</td>
<td>-2.488 (0.413)</td>
<td></td>
<td>-2.477 (0.189)</td>
<td>86%</td>
<td>4.6%</td>
</tr>
<tr>
<td>Flake</td>
<td>-2.863 (0.383)</td>
<td></td>
<td>-3.140 (0.375)</td>
<td>85%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Snowflake</td>
<td>-3.626 (0.657)</td>
<td></td>
<td>-3.664 (1.021)</td>
<td>53%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Dream</td>
<td>-4.226 (1.210)</td>
<td></td>
<td>-3.770 (0.177)</td>
<td>29%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Wispa</td>
<td>-3.224 (0.471)</td>
<td></td>
<td>-3.455 (0.264)</td>
<td>65%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Galaxy</td>
<td>-1.848 (0.533)</td>
<td></td>
<td>-1.853 (0.189)</td>
<td>87%</td>
<td>5.1%</td>
</tr>
<tr>
<td>Galaxy Caramel</td>
<td>-3.485 (0.444)</td>
<td></td>
<td>-3.194 (0.375)</td>
<td>66%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Galaxy Ripple</td>
<td>-3.193 (0.360)</td>
<td></td>
<td>-3.273 (1.021)</td>
<td>70%</td>
<td>2.7%</td>
</tr>
<tr>
<td>M&amp;Ms</td>
<td>-4.716 (0.586)</td>
<td></td>
<td>-4.994 (0.177)</td>
<td>80%</td>
<td>3.6%</td>
</tr>
<tr>
<td>Maltesers</td>
<td>-3.572 (0.632)</td>
<td></td>
<td>-3.974 (0.264)</td>
<td>95%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Mars</td>
<td>-1.632 (0.296)</td>
<td></td>
<td>-2.212 (0.189)</td>
<td>99%</td>
<td>15.7%</td>
</tr>
<tr>
<td>Mars 5 Little Ones</td>
<td>-9.403 (4.587)</td>
<td></td>
<td>-7.924 (0.375)</td>
<td>14%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Kit Kat</td>
<td>-1.565 (0.426)</td>
<td></td>
<td>-1.843 (1.021)</td>
<td>96%</td>
<td>11.8%</td>
</tr>
<tr>
<td>Kit Kat Chunky</td>
<td>-2.507 (0.554)</td>
<td></td>
<td>-3.054 (0.177)</td>
<td>89%</td>
<td>6.7%</td>
</tr>
<tr>
<td>Kit Kat Kubes</td>
<td>-5.322 (1.095)</td>
<td></td>
<td>-6.018 (0.264)</td>
<td>63%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Milky Bar</td>
<td>-2.780 (0.634)</td>
<td></td>
<td>-2.828 (0.189)</td>
<td>69%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Milky Bar Munchies</td>
<td>-6.581 (0.892)</td>
<td></td>
<td>-6.486 (0.375)</td>
<td>40%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Smarties</td>
<td>-5.922 (0.516)</td>
<td></td>
<td>-5.479 (1.021)</td>
<td>81%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Double Cream</td>
<td>-3.097 (0.731)</td>
<td></td>
<td>-3.000 (0.177)</td>
<td>65%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Snickers</td>
<td>-1.987 (0.413)</td>
<td></td>
<td>-2.579 (0.264)</td>
<td>97%</td>
<td>10.8%</td>
</tr>
<tr>
<td>Snickers Cruncher</td>
<td>-5.185 (1.548)</td>
<td></td>
<td>-5.353 (0.189)</td>
<td>38%</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

Table 4: Estimates for the random coefficients multinomial logit for each product

It can be seen from the results in table 4, that when availability is corrected for using the proposed methodology, two thirds of the product intercept estimates are higher. As we pointed out in the introduction, not accounting for availability “penalizes” products with low distribution coverage due, for instance, to sluggish retailer adoption. In contrast, well established products with very high distribution and market shares (Mars, Snickers, Kit Kat and popular plain chocolates) have lower product intercept estimates. The overall relationship between the coefficients and the distribution intensity can be seen in figure 2, where the differences between the coefficients obtained by the method correcting for availability and the standard method have been plotted against average availability after launch.

These patterns also translate to the product characteristics, although the format coefficients are harder to interpret because they depend on many products. Nonetheless, it is important to see that controlling for availability produces different estimates(table 5), particularly in the price coefficient, which is smaller in absolute value (i.e. less price sensitivity).

In many cases, the best selling brands in a given category have high levels of availability. If
Table 5: Estimates for the price coefficient and product format taste coefficient from three different models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Format</th>
<th>Estimate</th>
<th>Estimates (std. errors)</th>
<th>RC Logit</th>
<th>RC Logit availab.</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>mean</td>
<td>-5.414 (1.121)</td>
<td>-6.061 (0.375)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td>1.212 (1.185)</td>
<td>2.365 (0.399)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>block</td>
<td>mean</td>
<td>-2.509 (0.569)</td>
<td>-2.399 (0.447)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td>0.054 (1.376)</td>
<td>0.279 (1.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>indulgence</td>
<td>mean</td>
<td>-2.828 (0.853)</td>
<td>-3.299 (0.671)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td>1.063 (0.137)</td>
<td>1.897 (0.177)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>filler</td>
<td>mean</td>
<td>-3.979 (0.697)</td>
<td>-3.883 (0.548)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td>0.233 (0.526)</td>
<td>1.122 (0.264)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bite</td>
<td>mean</td>
<td>-5.042 (0.763)</td>
<td>-5.390 (0.600)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td>2.459 (0.240)</td>
<td>3.031 (0.189)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

the estimation were restricted to the top few brands in the category, the results from the two methods would not differ much from each other. However, in the category we are analyzing, the top brands account for only a fraction of the total category sales. So estimating the model using only the top brands (a practice that could be sensible when analyzing other categories such as ketchup) would leave a large proportion of the market out of the estimation. In general, we would expect this to be the case for situations in which a large number of products have to be included in the estimation (e.g. Nevo (2001) analysis of the cereal product category.)

One of the main uses of structural models of demand is the computation of price elasticities and optimal prices. To this end, it is important to have unbiased price coefficients. If we compare the estimates of the mean of the price coefficient for the two random coefficient logit models,
we see that correcting for availability produces a higher price sensitivity (larger coefficient, in absolute value). Sales variations due to changes in distribution are captured by the price coefficient (or other environmental factors in more general models) and this may result in the price coefficient being contaminated by non-price factors. It should be pointed out, that there is no general rule as to how the price coefficient may change and it will depend on each particular market circumstance.

<table>
<thead>
<tr>
<th>Product</th>
<th>RC Logit</th>
<th>RC Logit availab.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dairy Milk</td>
<td>-2.99</td>
<td>0.07 0.08 0.06</td>
</tr>
<tr>
<td>Whole Nut</td>
<td>0.16</td>
<td>-3.09 0.08 0.06</td>
</tr>
<tr>
<td>Fruit &amp; Nut</td>
<td>0.16</td>
<td>0.07 -3.07 0.06</td>
</tr>
<tr>
<td>Caramel</td>
<td>0.18</td>
<td>0.07 0.09 -2.17</td>
</tr>
</tbody>
</table>

Table 6: Own and cross price elasticities for the Cadbury block chocolate products.

For convenience, we do not reproduce the complete $24 \times 24$ elasticity matrices here, but they are available from the authors. As an example though, consider the Cadbury Block chocolates in table 6. Own elasticities are smaller in absolute value when we correct for availability. This seems to contradict the shift in the price coefficient, but we remind the reader that the computation involves the preference distributions as well. It was shown earlier that the bias on the product intercept is stronger for products with low average availability. The relationship between elasticities and availability seems to be more complex. Cadbury block chocolates, for example, are widely distributed (see table 3) but the elasticities computed vary for each method. Correcting for the availability of other products affects the product-type effect, which in turns affect the elasticity even for fully distributed products.

Elasticities are sometimes used to analyze the competitive landscape by computing summary measures. Usually, elasticities are used to compute the competitive clout and the vulnerability (Kamakura and Russell 1989). In figure 3 a few selected products are positioned in the clout-vulnerability plane. For some products the relative location in the clout-vulnerability space is unchanged: Mars seems to be in an advantageous competitive position compared to Snickers, the other popular filler bar. Interestingly, some products do shift in position when availability

---

$^5$Competitive clout is a measure of how other products sales are affected by a change in price of the focal brand, while vulnerability measures the change in sales associated to changes in prices of competing brands. It is a summation of the squares of the cross elasticities: $C_i = \sum_{j \neq i} \varepsilon_{ji}^2$ and $V_i = \sum_{j \neq i} \varepsilon_{ij}^2$
is corrected for. Kit Kat seems to have moderate vulnerability if the results of the standard random coefficients logit model are used, but seems much less vulnerable to competitors price cuts when availability is corrected for. Some new products shown in the lower left corner of the lower panel of figure 3 are inferred to have higher vulnerability in the standard logit model that does not account for the lower levels of distribution.

![Graph showing clout and vulnerability](image)

Figure 3: Clout and vulnerability

Finally, we examine the effect that correcting for availability has on optimal prices. In order to illustrate this, we compute the optimal markup (price minus marginal costs) of a monopolist retailer selling all the products. A retailer selling all products will maximize profits by setting prices $p - mc = (p_1 - mc_1, \ldots, p_J - mc_J)$ equal to

$$p - mc = -\Omega^{-1}S.$$

The matrix $\Omega$ is a $J \times J$ matrix containing $\frac{\partial S_i}{\partial p_j}$ in position $(i, j)$ and $S$ is a vector of market shares(see Besanko et al. 1997, for example).
We compare the optimal prices under the two alternative estimation procedure. In order to compute the prices from the markups we assume a margin of 30%. This number is an approximation that we obtained from conversations with managers and research on wholesale prices from trade journals and trade websites. In figure 4 the prices are plotted for all 113 weeks.

Figure 4: Actual prices for Mars bars and optimal prices computed using the two models.

We note from figure 4 that the optimal price obtained from the corrected model is higher than the optimal price obtained from the standard model. This is in agreement with the median price elasticities for the Mars bar, which are -1.48 and -1.66 for the corrected and standard model, respectively. We should point out, however, that the interpretation of this figure should be taken with care. Our model does not model availability and prices in an equilibrium framework and therefore produces optimal prices given availability.

Summary and Further Research

We have shown a method to incorporate varying levels of availability in discrete-choice based demand estimation with aggregate data. In the current study, however, we have not modeled availability explicitly. For instance, we have not modeled the retailer’s decision to stock or delist a product or modeled the probability of stock-outs. We have taken availability as given and used it to correct the observed shares used in the estimation. However, we have also shown that even if availability is determined endogenously, our estimation method can recover the mean utility.
We recognize that in order to make further statements about the role of availability and competition in the vertical channel, the availability levels should be modeled explicitly. For example, a model of vertical interaction could be set up in which manufacturers set wholesale prices and retailers maximize category profits over retail prices and assortment levels. In order to establish the equilibria in this case, we would need to make assumptions about the channel behaviour (e.g. Vertical Nash) and about the product availability behaviour (e.g. retailers have convex shelf costs). Estimating the model parameters can then be carried out.

Although the proposed methodology recovers the parameter estimates under non-complete availability, we have not addressed the issue of correlated levels of availability. In the development of the model we took independent draws $\tilde{a}^i = (\tilde{a}^i_1, \ldots, \tilde{a}^i_J)$ to obtain the expectation over all the possible availability vectors. In other words, if $\Sigma_a$ is the covariance matrix of the draws, then we assumed that $\Sigma_a$ had diagonal form with Bernoulli variances in the diagonal.

The assumption of independence may be limiting. For example, retail stores with low degree of assortment would have many products simultaneously unavailable while bigger retailers would probably have most products in stock. In addition, some products are produced, distributed or both by the same company. Therefore, we would expect that products from the same manufactures can show correlated levels of availability across stores. In those cases, it is possible to define the correlation matrix in terms of a set of parameters, take draws from the correlated multivariate Bernoulli distribution and include these parameters in the moment conditions.

We have confined our analysis to a diagonal matrix for simplicity and because we do not have enough variation in the dataset to obtain significant non-diagonal covariance matrices. Furthermore, multivariate Bernoulli covariance matrices are not as flexible to work with as multivariate normal covariance matrices, i.e. the mean already bounds the values the covariance can take. It remains the object of future research to characterize what correlation structures are most suitable for this type of research and what are the identification issues that arise.

To summarize, in this study we have recognized that varying levels of product availability can lead to misleading inferences if unaccounted for. Further, we proposed a methodology that extends the traditional random coefficients logit model to correct for product availability. Finally we have applied the proposed framework to the UK confectionery market and found that correcting for product availability shifts the utility parameter estimates and affects the
computation of price elasticities and the competitive analysis.
A Appendix

Proposition 1

We will show that (8) can be defined by a function that falls under the assumptions of the theorem stated in Appendix 1 of Berry et al. (1995) and therefore is a contraction with modulus less than one.

First, notice that the share function in (2),

\[ S_j = \sum_{a} S_j(a) \pi(a) . \]

is continuous and differentiable. Moreover \( \frac{\partial S_j}{\partial \delta_j} > 0 \) and \( \frac{\partial S_j}{\partial \delta_k} < 0 \). Let \( A^0 \) be the set of vectors \( a \) that have \( a_j = 0 \). Then

\[
S_j = \sum_{a \notin A^0} S_j(a) \pi(a) + \sum_{a \in A^0} S_j(a) \pi(a)
\]

\[ = \sum_{a \notin A^0} S_j(a) \pi(a) , \]

because \( a_j = 0 \implies S_j(a) = 0 \).

Define the function \( f(\delta) = \delta + \ln(S_{obs}) - \ln(S(\delta)) \). To be under the assumptions of Theorem 1 in Berry et al. (1995) we need

1. \( \frac{\partial f_j}{\partial \delta_j} > 0 \), \( \frac{\partial f_k}{\partial \delta_k} > 0 \).

The own partial derivative equals

\[
\frac{\partial f_j}{\partial \delta_j} = 1 - \frac{1}{S_j} \sum_{a \notin A^0} \frac{\partial S_j(a)}{\partial \delta_j} \pi(a)
\]

\[ = 1 - \frac{1}{S_j} \sum_{a \notin A^0} S_j(a) [1 - S_j(a)] \pi(a)
\]

\[ = \frac{1}{S_j} \sum_{a \notin A^0} \pi(a) , \]

and is clearly positive because both the share and the probability are greater than zero. So
is the cross derivative because

$$\frac{\partial f_j}{\partial \delta_k} = \frac{1}{S_j} \sum_{a \in A^0} \frac{\partial S_j (a)}{\partial \delta_k} \pi (a)$$

$$= \frac{1}{S_j} \sum_{a \notin A^0} S_j (a) S_k (a) \pi (a).$$

2. $\sum_{k=1}^{J} \frac{\partial f_j}{\partial \delta_k} < 1$

The sum of partial derivatives equals

$$\sum_{k=1}^{J} \frac{\partial f_j}{\partial \delta_k} = 1 - \frac{1}{S_j} \sum_{a \notin A^0} S_j (a) [1 - S_j (a)] \pi (a) + \frac{1}{S_j} \sum_{a \notin A^0, k \neq j} S_j (a) S_k (a) \pi (a)$$

$$= 1 - \frac{1}{S_j} \sum_{a \notin A^0} S_j (a) \pi (a) + \frac{1}{S_j} \sum_{a \notin A^0} \sum_{k=1}^{J} S_j (a) S_k (a) \pi (a)$$

$$= \frac{1}{S_j} \sum_{a \notin A^0} \sum_{k=1}^{J} S_j (a) S_k (a) \pi (a) = \frac{1}{S_j} \sum_{a \notin A^0} S_j (a) \left[ \sum_{k=1}^{J} S_k (a) \right] \pi (a).$$

The sum of market shares (in square brackets in the final expression above) is smaller than one, since we are not including the outside good and thus

$$\frac{1}{S_j} \sum_{a \notin A^0} S_j (a) \left[ \sum_{k=1}^{J} S_k (a) \right] \pi (a) < \frac{1}{S_j} \sum_{a \notin A^0} S_j (a) \pi (a) = 1,$$

as required.

3. The function $f$ has a finite lower bound.

Proceeding as in Berry et al(1995) we can write market share as

$$S_j = \sum_{a \in A^0} \int \frac{e^{\delta_j + \mu_j} dG(\mu)}{1 + \sum_{k=1}^{J} e^{\delta_k + \mu_k}} dG(\mu) \pi (a)$$

$$= e^{\delta_j} \times \sum_{a \in A^0} \int \frac{e^{\mu_j} dG(\mu)}{1 + \sum_{k=1}^{J} e^{\delta_k + \mu_k}} dG(\mu) \pi (a)$$

and the function $f$ as
\[ f(\delta) = \ln(S^{\text{obs}}) - \ln \left[ \sum_{a \in A^0} \int \frac{e^{\mu_j}}{1 + \sum_{k=1}^J e^{\delta_k + \mu_k}} dG(\mu) \pi (a) \right] \]

If the \( \delta_k \) go to \(-\infty\) the second term converges to

\[ -\ln \left[ \sum_{a \in A^0} \int e^{\mu_j} dG(\mu) \pi (a) \right] \]

4. There is a value \( \bar{\delta} \) such that if \( \delta_j \geq \bar{\delta} \) then for some \( k \), \( f_k(\delta) < \delta_k \).

By inspecting the signs of the partial derivatives, we verify that we are under the assumptions of the Lemma in the appendix of Berry (1994) that ensures the existence of \( \bar{\delta} \).
References


