ELSEVIER

Journal of Econometrics 89 (1999) 177–196

JOURNAL OF
Econometrics

Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters

Tülin Erdem[a,]*, Michael P. Keane[b], Baohong Sun[c]

[a] *Haas School of Business, University of California, Berkeley, Berkeley, CA 94720-1900, USA*
[b] *Department of Economics, University of Minnesota, Minneapolis, MN 55455, USA*
[c] *Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*

Abstract

Discrete choice models have been widely estimated on scanner panel data to study consumer choice. One challenge in scanner panel research is that only the prices of the items bought are recorded. The ad hoc models used to fill in the missing prices of non-purchased brands may create a self-selection bias in estimating consumer price sensitivities. This type of bias is also present in existing studies of coupon effects. To obtain consistent estimates of price elasticities in the presence of missing price and coupon values, we estimate a brand choice model jointly with models for the price and coupon processes. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C23; C25; M31

Keywords: Brand choice; Scanner data; Self-selection bias; Price elasticities

1. Introduction and background

Over the past decade, with the increased availability of electronic *scanner panel data* on household purchasing behavior, discrete choice models have become very important for marketing managers to diagnose the impact of

* Corresponding author. E-mail: erdem@haas.berkeley.edu

PII: S0304-4076(98)00060-8

marketing mix strategies on consumer purchasing behavior. Indeed, since Guadagni and Little's seminal work (1983), there has been a plethora of choice models estimated on scanner panel data in marketing. Most of this research has focused on the impact of firm marketing mix strategies, especially pricing and promotion strategies, on consumer brand choice.

One challenge that faces researchers in scanner panel research is that only prices of the items bought are recorded. This is a direct consequence of how the data are collected: information on purchased items is scanned into a database. Thus, we do not generally observe the prices consumers face for the alternatives they did not buy. Traditionally, the missing prices of non-purchased brands have been filled in rather arbitrarily by some ad hoc method such as forward or backward extrapolation of prices from previous or subsequent weeks. For example, A.C. Nielsen Co., who has released most of the scanner panel data sets available for academic use, never collected complete daily in-store price information for all brands. Rather, they employed a complex algorithm, which involves backward and forward extrapolation of prices, to fill in missing prices of non-purchased brands.[1] Nielsen does provide price files with their scanner data sets, and these files do contain nearly daily information on the prices of all brands in all stores. But it is important to note that most of the prices in these files are actually imputed using the above-mentioned algorithm, rather than being directly recorded.

To give an idea of the magnitude of the missing price problem, consider the Nielsen scanner data on Ketchup for Sioux Falls over the period 1986–1988. For the three major brands, five major sizes (i.e., 11 brand-size combinations since not every brand has each size), 860 days of the sample period and 19 largest stores in the city, there are a total of 179 740 prices. Of these only 35 568 were actually recorded by in store scanners at the point of purchase. Thus, 80.21% of the prices in the data set must be imputed.[2] This illustrates the practical severity of the missing price problem in scanner data.

[1] The documentation that Nielsen made available to the academic community describes in detail the procedures employed for filling in the missing prices of items never bought at a certain store on a specific day. This documentation is available from the first and second authors upon request. It should also be noted that even after the complex ad hoc procedure Nielsen employs to construct the price files, often these data files may still contain some missing data.

[2] Note that the above figures utilize all information available in the data files. Thus, if the same brand size was bought buy any consumer at the same store on the same day, we have an observed price. We also did the following exercise: if a price was observed for a brand-size at a store on any day during a Monday to Sunday week, we assumed that we then know the price for the whole week. Even then, 34.44% of the prices are missing. However, note that we do not in fact know the day of the week when the prices are changing for different items at a store. Indeed, the algorithm used by Nielsen to construct the store price files assumes that all prices change once a week on the same day for all products and involves a complex mechanism to decide which day of the week will be chosen as the 'assumed' day of a price change for all products.

The current practice in marketing is either to fill in prices for non-purchased brands using ad hoc methods like backward or forward extrapolation or to rely solely on store price files provided by Nielsen which are themselves imputed by using ad hoc methods. However, except for a very few studies (e.g., Allenby and Rossi, 1991; Erdem and Keane, 1996; Tellis, 1988), most of the published papers in marketing do not describe the method employed to fill in the missing data. This is mainly because most researchers use the Nielsen price files as they are without commenting on the ad hoc methods employed by Nielsen to fill in the missing data.

There has been no research conducted on the potential effects of these ad hoc methods on parameter estimates. However, use of ad hoc methods to deal with missing prices for non-purchased brands may create a self-selection bias (Heckman, 1974, 1985) in estimating consumer latent utility functions and consumer price response coefficients. This may lead to misleading conclusions with respect to pricing, promotion and segmentation strategies. To gain an intuitive understanding of the nature of the bias created by missing prices in scanner data, consider the following argument. If agents are price sensitive, then the expected price of a brand conditional on the fact that the brand is purchased (that is, the mean accepted price) will be less than the mean *offer* price for that brand. Similarly, the expected price of a brand conditional on the fact that the brand is not purchased will exceed the mean offer price for that brand. Thus, if one substitutes for missing prices in scanner data using the averages of accepted prices in earlier and subsequent weeks, one will tend to underestimate the prices of non-purchased brands. This, in turn, will lead to estimates of price elasticities of demand for a specific brand that are biased towards zero.

This type of bias is likely to be much more severe in existing studies of coupon effects on purchase behavior. There, it has been common to assume zero for the values of the coupons available for non-purchased brands – rather than attempting to form some proxy for these values as is commonly done with missing prices. Thus, the coupon values available for non-purchased brands may be seriously underestimated. In this case, one is overestimating the prices of non-purchased brands, leading to upward biased estimates of price elasticities.

The main aim of this research is to develop a method to assess and account for the selection bias created by missing price and coupon availability data for non-purchased items in choice models estimated on scanner panel data. Given the wide use of such choice models in understanding the impact of pricing and promotion policies on consumer purchasing behavior, it is of fundamental importance to obtain consistent estimates of price coefficients in these models.

The next section develops our proposed method to deal with missing price and coupon availability data. Section 3 describes the data we used in the empirical application and discusses the results we obtained from the empirical analysis. Section 4 concludes.

2. Correcting for the self-selection bias in choice model estimates

To obtain consistent estimates of price elasticities in the presence of missing prices and coupon values, we estimate a brand choice model jointly with models for the price and coupon processes. This is analogous to the procedure typically used in labor economics to estimate industry/occupational choice models when only accepted wages (i.e., the wage for the chosen industry/occupation) are observed (see, for example Heckman and Sedlacek, 1985; Keane et al., 1988). Assume that utility for person i at time t conditional on purchase of brand j is given by[3]

$$U_{ijt} = \alpha_j + \beta_i P_{ijt} + \gamma_i C_{ijt} + \mu_{ij} + \omega_i A_j + \varepsilon_{ijt} = U^*_{ijt} + \varepsilon_{ijt}, \quad j = 1, \dots, J, \tag{1}$$

In Eq. (1), α_j is a brand-specific intercept. P_{ijt} is the price of brand j to person i at time t (determined by the store that person i visits at t, which is assumed to be exogenous). C_{ijt} is the value of the coupon available to person i at time t for purchase of brand j (which will be zero if they have no coupon available). β_i and γ_i are price and coupon sensitivity coefficients, which are assumed to be individual-specific. ε_{ijt} is an idiosyncratic shock to the preference of i for brand j which is i.i.d over time and across brands.

Previous scanner data research has shown that it is important to incorporate unobserved attributes of alternatives in order to capture choice behavior (e.g., Allenby et al., 1997). Such unobserved attributes can be either unique or common.[4] In Eq. (1), μ_{ij} is a time-invariant random effect capturing i's preference weight for an unobserved unique attribute of brand j. A_j denotes brand j's level of an unobserved common attribute. ω_i is the utility weight that consumer i assigns to the common attribute. It should be noted that in the literature on choice models, it is fairly typical to postulate that persistence in individual choices is captured by taste heterogeneity and to assume the remaining error terms to be i.i.d. This is indeed why it is important to allow for unobserved attributes of alternatives in choice models.

We normalize $\alpha_J = A_J = 0$ for identification and $\mu_{ij} = 0$ for convenience (but without loss of generality).[5] Distributional assumptions for β_i, γ_i, μ_{ij} and ω_i are as follows:

$$\beta_i \sim \mathrm{N}(\beta, \sigma^2_\beta), \quad \gamma_i \sim \mathrm{N}(\gamma, \sigma^2_\gamma), \quad \mu_{ij} \sim \mathrm{N}(0, \sigma^2_\mu), \quad \omega_i \sim \mathrm{N}(0, 1).$$

[3] We treat the purchase occasion as the time unit and use it interchangeably with 'time'.

[4] A common attribute of ketchup brands could be thickness or richness. A unique attribute may be the image associated with a particular brand (e.g., 'Heinz has been the leading brand for these many years').

[5] This is equivalent to replacing the μ_{ij} for $j = 1, J - 1$ with $\mu'_{ij} = \mu_{ij} - \mu_{iJ}$.

The mean restrictions on μ_{ij} and ω_i are needed to identify the brand-specific intercepts (α_j). The variance restriction on ω_i sets the scale for the common attribute (A). Further, β_i and γ_i are assumed to be correlated. This correlation is denoted by λ.

Define a consumer's type as $v_i = (\mu_{i1}, \dots, \mu_{i.J-1}, \omega_i, \beta_i, \gamma_i)$. Let π_{ijt} denote the probability of consumer i purchasing brand j on purchase occasion t, conditional on v_i, and the set of coupons and prices a consumer actually faces:

$$\pi_{ijt} = Prob[\,j|\mu_{ij}, \omega_i, \beta_i, \gamma_i, (\alpha_k, A_k, P_{ikt}, C_{ikt}, k = 1, 2, \dots, J)]. \tag{2}$$

Assuming that ε_{ijt} has a Type I extreme value distribution, we have the conditional logit form for the choice probabilities (McFadden, 1974):

$$\pi_{ijt} = \frac{\exp\{U^*_{ijt}\}}{\sum_{k=1}^{J}\exp\{U^*_{ikt}\}}. \tag{3}$$

The hypothetical likelihood function contribution for person i that could be formed if all offer prices and available coupon values were observed is given by

$$L_i^h = \int_v \prod_t \left(\sum_{j=1}^{J} I_{ijt}\pi_{ijt} \right) f(v)\, \mathrm{d}v. \tag{4}$$

I_{ijt} is the indicator function such that $I_{ijt} = 1$ if consumer i purchases brand j on purchase occasion t and $I_{ijt} = 0$ otherwise, and $f(.)$ is the joint density function of the random effects vector v. Note that the dimension of this integral is $(J-1)+3$.

Of course, we cannot form the likelihood as indicated in Eq. (4). To form the likelihood function for the observed data, assuming that prices and available coupon values are only observed for the chosen brand, we must first specify the distribution of prices and coupons and then integrate out the unobserved prices and coupon values in Eq. (4).

We now describe our distributional assumptions with regard to prices and coupons. In the data set we use for estimation, observed prices tend to be bunched at a small set of values. Therefore, we assume that the (per ounce) price of brand j that consumer i faces at time t takes on one of a discrete set of values:

$$P_{ijt} \in \{P_1, P_2, \dots, P_L\}, \forall j \tag{5}$$

where L is the number of prices. Associated with the (per ounce) price P_l, $l = 1, 2, \dots, L$ is the probability ρ_{jsl} that consumers face this price for brand j size s.

Similarly, we assume that coupons take on one of a discrete set of values:

$$C_{ijt} \in \{0, C_1, C_2, \dots, C_R\}, \forall j, \tag{6}$$

where R is the number of coupon values. The probability that C_{ijt} takes on the value C_r for $r = 1, 2, \dots, R$ is denoted by δ_{jsr}. There is also a probability δ_{j0} that a consumer has no coupon available when purchasing brand j, in which case we set $C_{ijt} = 0$.

Note that if a consumer uses no coupon to purchase a brand it could be because no coupon was available for that brand. Alternatively, it could be that the consumer attaches some cost to using a coupon (e.g., the time cost of cutting it out of the newspaper) and this cost exceeds the value of the available coupon. Since scanner data contains no information on coupon *availability* and coupon *use* decisions (i.e., we never observe if a person had a coupon available and chose not to use it) it is impossible (at least in the absence of very strong assumptions) to disentangle these alternative explanations for non-use of coupons. Thus, we make no attempt to model coupon use decisions.[6] Rather, we simply assume a probability that no coupon is available, recognizing that this subsumes both the case in which no coupon was available and in which one was available but the consumer chose not to use it.[7]

The price and coupon probability parameters ρ_{jsl} and δ_{jsr} will be estimated jointly with the utility function parameters. A completely non-parametric approach would estimate these as a large set of free parameters, restricted only in that the sum of price (coupon) probabilities for each brand/size must equal one. A much more parsimonious representation is obtained by imposing some smoothness on the probabilities. We assume that the price probabilities and conditional coupon probabilities for brand j are given by the functions:

$$\rho_{jsl} = \varphi_p(P_l) = \frac{\exp\{(a_{1j} + a_s)P_l + a_2P_l^2 + a_3P_l^3 + a_4P_l^4\}}{A}, \quad l = 1, 2, \dots, L, \tag{7}$$

$$\delta_{jsr} = \varphi_r(C_r) = \frac{\exp\{(b_{1j} + b_s)C_r + b_2C_r^2 + b_3C_r^3 + b_4C_r^4\}}{B}(1 - \delta_{j0}), \quad r = 1, 2, \dots, R, \tag{8}$$

where A and B are normalizing constants equal to the sum over l and r of the exponnetial terms in the numerators in Eqs. (7) and (8), respectively. In Eqs. (7)

[6] The coupons that are available to consumers are manufacturer's and store coupons. In the empirical analysis, the coupon values we use are the sum of the manufacturer's and store coupon values.

[7] Thus, our procedure deals with the blatant endogeneity problem that arises because coupon values are only observed for purchased brands. But we do not address the more subtle endogeneity problem that may arise if tastes for brands influence coupon use and/or availability. Similarly, price itself may be endogenous if consumers wait to make purchases at dates or in stores at which the price of a favorite brand is low. Previous scanner data research has not addressed these endogeneity problems either.

and (8), the parameters a_{1j} and b_{1j} are brand-specific, whereas a_s and b_s are size-specific. For identification, we normalized the parameters a_s and b_s to be zero for one size.

By increasing the order of these polynomials and allowing for more flexible size interactions one would approach to a case in which all the ρ_{jsl} and δ_{jsr} are free parameters. However, the functions (7) and (8) are sufficiently flexible to capture complex multi-modal distributions (as we show in Section 3) and use of more flexible polynomials did not significantly improve the fit.

Integrating over the prices and coupon values for non-purchased brands, the likelihood contribution for person i is

$$L_i = \int_v \prod_t \sum_j I_{ijt}\varphi_p(P_{ijt})\varphi_c(C_{ijt}) \sum_{P^U,\, C^U} \pi_{ijt}(v, P_{ijt}, C_{ijt}, P^U, C^U)\rho^U\delta^U f(v)\, \mathrm{d}v \qquad (9)$$

where $P^U = \{P_{ikt}\}_{k \neq j}$ and $C^U = \{C_{ikt}\}_{k \neq j}$ are vectors of unobserved prices and coupon values, while $\rho^U = \prod_{k \neq j}\rho_{k,\, s,\, l_k}$ and $\delta^U = \prod_{k \neq j}\delta_{k,\, s,\, r_k}$ and are the joint probabilities of those vectors. The l_k and r_k in the ρ^U and δ^U expressions denote the indices of the price and coupon values for the kth brand.[8]

The likelihood function is formed by multiplying the individual contributions (9) for all households. We will estimate the price and coupon processes jointly with the choice model by maximizing this likelihood function with respect to the parameters of the choice model, the price process and the coupon process.

It is important to understand how joint estimation of the price process with the choice model will lead to different inferences about the price process than simply assuming unobserved prices come from the same distribution as observed prices. To the extent that agents are price sensitive, an estimated offer price distribution that implies that observed (i.e., accepted) prices are below the mean of the offer price distribution will tend to maximize the likelihood of the observed choices and prices. This upward adjustment in the mean of unobserved prices leads to consistent estimation of price elasticities (of course, as always, consistency requires that our model is ‘correct’).

Note that forming the likelihood function contribution in Eq. (9) involves a high-dimensional integration over the missing prices and coupon values and

[8] Note that the choice probability $\pi_{ijt}(v,P_{ijt},C_{ijt},P^U,C^U)$ is conditional on i’s unobserved type v and the unobserved prices P^U and coupons C^U for non-purchased brands. The corresponding unconditional choice probability is obtained by integrating over the joint distribution of these unobservables. Of course, since the distributions of P^U and C^U are discrete, the integration over the distributions of P^U and C^U means taking a weighted sum of $\pi_{ijt}(v,P_{ijt},C_{ijt},P^U,C^U)$ over all possible combinations of the unobserved price and coupon values with the weights equal to the probabilities of each unobserved price and coupon value combination. The result can then be integrated over the distribution of v, or the order of this integration and the above explained summation can be reversed to get the same result. Note that we use the price and coupon probabilities relevant for the size that the consumer bought. We do not model size choice.

over the unobserved random effects. It is not feasible to evaluate such high-dimensional intergrals using numerical methods like quadrature in the context of maximum likelihood estimation. Rather, we turn to simulation estimation techniques of the type considered by Lerman and Manski (1981), McFadden (1989), Pakes and Pollard (1989), Keane (1993, 1994), McCulloch and Rossi (1994) and Geweke et al. (1994). One approach is to use simulated maximum likelihood (SML) estimation, in which the likelihood contributions in Eq. (9) are replaced by simulated values of the form

$$L_i \approx D^{-1}\sum_{d=1}^{D}\prod_{t}\left\{\sum_{j=1}^{J} I_{ijt}\varphi(P_{ijt}, C_{ijt})\left\{M^{-1}\sum_{m=1}^{M}\pi_{ijt}(v_d, P_{ijt}, C_{ijt}, P^{\mathrm{U}}_{m}, C^{\mathrm{U}}_{m})\right\}\right\} \tag{10}$$

where $\pi_{ijt}(v_d, P_{ijt}, C_{ijt}, P^{\mathrm{U}}_{m}, C^{\mathrm{U}}_{m})$ denotes the probability that i chooses j at t given the draw v_d for the individual parameters and draw $(P^{\mathrm{U}}_{m}, C^{\mathrm{U}}_{m})$ for the unobserved prices and coupons for the non-purchased brands. D and M denote the total number of draws associated with v_d and $(P^{\mathrm{U}}_{m}, C^{\mathrm{U}}_{m})$, respectively. The SML estimator is consistent and asymptotically normal (with a limiting distribution centered around zero) in sample size N if the number of draws used to simulate the likelihood function grows with sample size at a sufficient rate so that $D/\sqrt{N}\rightarrow\infty$ and $M/\sqrt{N}\rightarrow\infty$ as $N\rightarrow\infty$. However, Monte-Carlo work by Keane (1993, 1994) and Geweke et al. (1994), among others, suggests that SML has excellent small sample properties provided reasonably accurate simulators are used.

It is important to note that there is a very interesting feature that arises in the brand choice context that has not arisen in previous applications of selection models. Specifically, prices faced by consumer i in a store at time t are in fact observed for some brands other than that which i purchased. The way that scanner data are constructed, it is often the case that we have data on one or more consumers who shopped in that same store on that same day. If one or more consumers bought a brand other than that bought by i, then we will observe in the data the prices that i faced for these other brands. Thus, the set of unobserved prices for consumer i at time t is in general a subset of the set of prices for brands not purchased by i.

In the case where prices are unobserved only for brands not purchased by *any* consumer at time t, the simulated likelihood contribution for i is

$$L_i \approx D^{-1}\sum_{d=1}^{D}\prod_{t}\left\{\sum_{j=1}^{J} I_{ijt}\varphi(P_{ijt}, C_{ijt})\left\{M^{-1}\sum_{m=1}^{M}\pi_{ijt}(v_d, P^{\mathrm{O}}_{t}, C_{ijt}, P^{\mathrm{U}}_{tm}, C^{\mathrm{U}}_{m})\right\}\right\}, \tag{11}$$

where $\pi_{ijt}(v_d, P^{\mathrm{O}}_{tm}, C_{ijt}, P^{\mathrm{U}}_{tm}, C^{\mathrm{U}}_{m})$ denotes the probability that i buys j at t given the observed *set* of prices P^{O}_{t} (which now includes both P_{ijt} and, possibly, the prices of some other brands), the observed coupon value C_{ijt}, the draw v_d for the individual parameters, draw P^{U}_{tm} for the set of unobserved prices for brands not bought by *any* consumer at t, and draw C^{U}_{m} for the coupons available for all non-purchased brands.

Suppose that the possibility that purchase decisions by other consumers affect the set of prices we observe for person i at time t is ignored. Specifically, suppose we ignore any observed prices for brands not purchased by i at t, treating the prices of all brands that i did not purchase at t as unobserved even if some other consumer did purchase one of those brands in the same store. This would lead to a consistent but inefficient estimator, since we are ignoring information but forming the correct 'limited information' (LI) likelihood. Thus, if our model assumptions are correct, the LIML estimates based on Eq. (10) should be consistent but inefficient while full information maximum likelihood (FIML) estimates based on Eq. (11) should be consistent and efficient. Thus, similarity of model estimates obtained by maximizing likelihoods based on Eq. (10) vs. Eq. (11) should provide a check on our assumptions.

3. Empirical analysis

3.1. Data

We apply our proposed method to A. C. Nielsen scanner panel data for ketchup to demonstrate its effectiveness in assessing and eliminating the self-selection bias created by missing price and coupon availability data for non-purchased items. We use data collected from the Springfield, MO test market to estimate our proposed model.

The sample consists of 344 households who made a total of 1392 purchases of ketchup. The analysis includes three brands, namely, Heinz, Hunts and store brands, which together capture 88.90% of the total market share. Each brand has package sizes of 14, 28, 32, 44 and 64 oz. We normalize price and coupon values into dollars per 32 oz. The average price and coupon value are \$1.228 and \$0.443, respectively. Note that the coupon values are the sum of the manufacturer's and store coupon values. Table 1 reports the descriptive statistics for the brands under analysis.

3.2. Results

In this section, we report the results of estimating three alternative models. In the first model (Model 1), we assumed that the price and coupon values of all alternatives not chosen by the consumer are unobservable. The likelihood for this model is simulated as in Eq. (10), where we integrate over the prices and coupon values for non-chosen alternatives. We set $M = D = 100$ and normalize $a_s = 0$ and $b_s = 0$ for the 64 oz size.

Table 1
Descriptive statistics

Brand name	Market share (%)	Mean price	% of Purchases made without coupon	Mean non-zero coupon
Heinz	55.9	\$1.377	86.5	\$0.410
Store brands	12.5	\$0.838	96.8	\$0.100
Hunts	20.5	\$1.186	84.3	\$0.477
	88.9	\$1.228		\$0.443

The second model (Model 2) treats prices as unobservable *only* for brands not purchased by any consumers in the same store on the same day. In the event that a consumer does not purchase a brand, we search over other consumers for one who bought the same brand in the same store on the same day. If such a consumer exists, we use the price paid by this consumer to substitute for the missing price. The likelihood for this model is simulated as in Eq. (11), where we are integrating over the remaining unobserved prices and coupons. In our data set, 389 missing prices out of 2784 were filled in this way. Thus, we were able to substitute for only 14% of missing prices using this method.

In the third model (Model 3), we assume all prices and coupon values are observable. Prices for non-purchased brands are filled in using a method like those traditionally used in the marketing literature. The prices for non-purchased brands are filled in using a multi-step process. First, we followed the procedure we adopted when we filled in the missing prices in Model 2. We filled in the remaining missing prices by using the average price marked for brand j during a week, averaging over days when sales were observed. If the whole week had no sales, we used the average price over non-promotion days for the whole sample period. Finally, the available coupon values for non-purchased brands are assumed to be zero (that is, it is assumed that consumers did not have coupons for the brands they did not purchase).

Table 2 reports the results for Models 1, 2 and 3. The mean price coefficients (β) are -4.18, -4.98 and -2.24, respectively. Given that the standard errors of these parameter estimates are 0.45, 0.50 and 0.65, respectively, one can argue that the mean price coefficient estimates obtained in Models 1 and 2 are reasonably close whereas Model 3 yields a much smaller price coefficient. Furthermore, the estimates for the mean coupon coefficients (γ) in Models 1 and 2 are 4.96 and 3.89, which are reasonably close as well, given the standard errors of 2.03 and 1.65. As previously discussed, the similarity of the price and coupon coefficient estimates in Models 1 and 2 provide support for the validity of our model assumptions. Finally, the mean coupon coefficient estimate in Model 3 is 1024.50. This large magnitude is not surprising because of the serious inherent

Table 2
Model estimation[a] (based on Eq. (1))

Parameters		Model 1	Model 2	Model 3
Choice model	(Heinz)	6.60(1.29[b])	4.39(1.71)	3.12(5.29)
Brand specific coefficient α	(Hunts)	6.91(1.02)	5.62(1.22)	4.66(4.10)
Mean price coefficient β		– 4.18 (0.42)	– 4.98(0.51)	– 2.24(0.65)
Standard deviation of price coefficient σ_β		1.842(0.34)	1.56(0.73)	5.67(1.52)
Mean coupon coefficient γ		4.96(1.96)	3.89(1.64)	102450(153.24)
Standard deviation of coupon coefficient: σ_γ		1.41(0.85)	1.25(0.76)	444.87(81.38)
Correlation between distributions of β and γ		– 0.48(0.24)	– 0.49(0.24)	– 0.53(0.25)
Common attribute a	(Heinz)	5.46(2.41)	4.89(2.43)	6.26(3.46)
	(Hunts)	4.35(1.63)	3.88(1.33)	5.49(1.91)
Preference heterogeneity parameter[c] σ_μ		0.62(0.33)	0.60(0.32)	0.71(0.25)
Price polynomial				
Brand-specific coefficient a_1	(Heinz)	14.93(2.50)	13.26(5.51)	
	(Store brands)	13.74(2.39)	12.35(5.90)	
	(Hunts)	13.13(2.47)	11.77(5.69)	
Size-specific coefficient a_s	(14 oz)	0.72(0.24)	0.77(0.26)	
	(28 oz)	– 1.85(0.25)	– 1.95(0.41)	
	(32 oz)	– 1.19(0.41)	– 1.42(0.56)	
	(44 oz)	– 2.00(0.35)	– 2.00(0.75)	
Quadratic price coefficient a_2		– 8.65(2.13)	– 7.31(3.64)	
Cubic price coefficient a_3		0.99(0.93)	0.60(0.86)	
Quadric price coefficient a_4		0.32(0.20)	0.37(0.15)	

Table 2 Continued

Parameters		Model 1	Model 2	Model 3
Coupon polynomial				
Probability of having no coupon δ_0	(Heinz)	0.86(0.016)	0.87(0.022)	
	(Store brands)	0.98(0.014)	0.97(0.017)	
	(Hunts)	0.88(0.061)	0.87(0.043)	
Brand specific coefficient b_1	(Heinz)	8.37(86.75)	10.68(59.10)	
	(Store brands)	7.80(77.39)	12.78(67.10)	
	(Hunts)	9.19(84.24)	10.70(58.72)	
Size specific coefficient b_s	(14 oz)	– 6.85(71.94)	– 7.13(23.34)	
	(28 oz)	– 6.96(71.60)	– 6.53(23.30)	
	(32 oz)	– 8.78(71.73)	– 11.50(23.20)	
	(44 oz)	– 8.01(71.49)	– 7.48(24.15)	
Quadratic coupon coefficient b_2		– 3.52(44.56)	– 6.38(24.55)	
Cubic coupon coefficient b_3		47.19(128.78)	62.95(87.25)	
Q coupon coefficient b_4		– 78.78(69.90)	– 74.69(85.28)	
Overall Log-likelihood Value		– 6830.9	– 6816.1	– 1081.1
(Choice model π, price/coupon polynomials φ)		(– 1467.1, – 5363.8)	(– 1460.5, – 5355.6)	(– 1081.1, 0)

[a] Number of observations = 1392. Number of households = 344.

[b] Standard error

[c] Standard deviation of the preference weights for the unique attribute.

endogeneity problem in Model 3. Since the use of a coupon perfectly predicts the brand choice (the coupon is only used when that brand is bought), the coupon coefficient grows arbitrarily large as the search algorithm proceeds.[9]

The estimates for heterogeneity parameters are all statistically significant. Thus, there is heterogeneity in preferences (σ_μ is 0.62 in Model 1, 0.60 in Model 2 and 0.71 in Model 3), price sensitivities (σ_β is 1.84, 1.56 and 5.67 in Models 1, 2 and 3, respectively) and coupon sensitivities (σ_γ is. 1.41, 1.25 and 444.87 in Models 1, 2 and 3, respectively). Again note that the estimates obtained from Models 1 and Model 2 are similar.

We now turn to the parameters that describe the offer distributions for prices and coupons. The estimated probabilities associated with having a coupon are low in this data since the estimates for the probability of having no coupon (δ_{j0}) are higher than 0.85 for all three brands. The estimates for the price and coupon polynomial parameters are difficult to interpret except for a few parameters. Therefore, rather than discussing polynomial parameter estimates, we report on the shapes of the implied offer distributions.

Table 3 reports the simulated frequencies for prices of the three brands under analysis. Since we have four sizes for each brand, we only report the results for the 32 oz size. We also report the observed price distributions (that is, frequencies observed in the data) for comparison purposes. Investigation of Table 3 reveals that the observed frequencies of lower prices are systematically higher than the offer frequencies. For example, the observed frequency of 69 cents for store brands is 46.2% whereas the actual (predicted) frequency is 41.5%. This discrepancy is indicative of the self-selection bias. When one looks only at the accepted prices, one tends to observe the lower end of the price distribution. However, note that the self-selection bias is more substantial for store brands than national brands. This suggests that consumers who consider buying store brands are more price elastic than consumers who are much less likely to buy store brands. Therefore, store brands are rarely bought at high prices which leads to more missing prices of store brands than national brands.

Finally, for coupons, comparisons of the parameter estimates for the probability of having no coupon available (Table 2) and the percentages of purchases made without a coupon in the sample (that is, sample frequencies of purchases without a coupon conditional on purchase (Table 1)) reveal that, as expected, the probability of using a coupon conditional on purchase exceeds the probability of having a coupon available. For example, 15.7% of Hunts' purchases are made with a coupon (Table 1) whereas the probability that a coupon is available

[9] Note that analytically, a coefficient of infinity will maximize the likelihood function. However, since we are using an iterative search routine on a finite computer, the likelihood ceases to improve by an amount that exceeds the convergence criterion after the coupon coefficient achieves some large finite value.

Table 3
Comparison of sample (accepted) frequency and offer[a] (actual) frequency for possible price values (32 oz)

	Heinz, 32 oz		Store brands, 32 oz		Hunts, 32 oz	
Possible values ($)	Sample freq. (%)	Offer[a] freq. (%)	Sample freq. (%)	Offer freq. (%)	Sample freq. (%)	Offer freq. (%)
0.69		0.0	46.2	42.2		0.0
0.79	2.3	2.2	19.4	18.2	5.6	5.4
0.85		0.1	1.6	1.1		0.1
0.89	3.1	3.1	8.6	8.7	20.7	19.3
0.97		0.1	1.6	1.6	0.8	0.9
0.98		0.1		0.2		0.1
0.99	27.6	26.4	2.7	3.0	16.9	16.4
1.08	1.0	1.2		0.7		0.1
1.09	1.0	1.2	5.4	5.3	2.6	2.4
1.16	0.8	0.8	1.6	1.6		0.2
1.18		0.4		0.2		0.1
1.19	31.7	29.6		0.3	13.9	13.2
1.25	0.2	0.8	5.4	8.6		0.1
1.27		0.2		0.1		0.1
1.29	1.5	1.8		0.1	7.1	6.2
1.31		0.1	1.6	1.7	0.8	0.8
1.35		0.1	2.7	3.0		0.1
1.36		0.8		0.8	4.5	4.4
1.38		0.2		0.2		0.2
1.39	15.7	14.4		0.1	6.0	6.1
1.42		0.2		0.2	8.3	8.1
1.45	12.3	11.5		0.2	4.5	4.7
1.47		0.2	1.1	1.1		0.3
1.50	0.4	0.4	1.1	1.2	2.3	2.7
1.54	1.3	1.3		0.2	3.0	3.1
1.58		0.4		0.1		0.1
1.59	1.0	1.0	0.5	0.4	0.4	0.5
1.67		0.2	0.5	0.5		0.1
1.79		0.1		0.1	2.6	2.5
1.99		0.1		0.1		0.1

[a] The columns labeled as ‘sample frequency’ report the accepted price distributions. The columns labeled as ‘offer frequency’ contain the predicted frequency conditional on purchases based on our fitted polynomial for prices.

for Hunts as predicted by Model 2 is 13% (Table 2). The discrepancy between the two probabilities is smaller for Heinz and smallest for the store brands.

Now we turn to the goodness of fit of the estimated models. We should note that goodness of fit comparisons among these models are not meaningful since different ‘data’ were used in each case (missing coupons and prices are treated differently). However, it should be mentioned that Model 3, which is severely misspecified, seems to have the best fit: the log-likelihood is -1081.6 whereas the choice part of the likelihood associated with Models 1 and 2 are -1467.9 and -1461.0, respectively.[10] Thus, not surprisingly, badly misspecified models may yield better goodness of fit statistics.

Models 1, 2 and 3 based on Eq. (1) specify separate coefficients for prices and coupons. We also estimated a constrained version of Eq. (1). Namely, we set $\beta = -(\gamma)$. Thus, we constrained the price and coupon coefficients to have the same magnitudes and opposite signs. This implies that only price net of coupons matters in purchase decisions.

Note that Model 3B is the model traditionally used in studies that incorporate coupons in marketing. In contrast to Model 3, the endogeneity problem is not as blatant in Model 3B because one no longer has an explanatory variable that can only take a non-zero values for the brands actually bought. However, the right hand side still contains the choice indicator hidden in the net price variable (that is, net price = price − (choice indicator × coupon) where the choice indicator or choice dummy equals one only for the particular brand that the consumer actually bought. Thus, Model 3B is still subject to the endogeneity problem.

Table 4 reports the results for the constrained models (Models 1B, 2B and 3B) where the price variable is net of coupons. A comparison of Table 2 with Table 4 shows that constraining the price and coupon coefficients to be the same magnitude with opposite signs causes the fit for all the models to deteriorate.[11] Thus, the results show evidence for differential response of consumers to price cuts versus coupons.

Finally, to completely eliminate the endogeneity problem inherent in Models 3 and 3B, we re-estimate the ad hoc model (Model 3) by ignoring coupons. We call this Model 3C (Thus, Model 3C suffers under omitted variables problem.). Table 5 reports the estimates and goodness of fit statistics for this model. The main result is that all three ad hoc models, that is, Models 3, 3B and 3C, underestimate the price sensitivity compared to the models that deal with the self-selection problem (Models 1, 2, 1B and 2B).

Overall, the results indicate that self-selection bias due to missing price and coupon data exists in this data set. Models that use ad hoc methods to fill in the missing prices generate smaller price coefficients than our proposed models that account for the self-selection problem. To further investigate the degree of bias, we conducted policy experiments where Hunts' cuts its price by 10 and 25%. To

[10] The likelihood for Model 3 is not comparable to these for Models 1 and 2 because Models 1 and 2 have to fit the price and coupon data while for Model 3 this data as given. Therefore,we divided the likelihoods into a choice and price/coupon polynomial part. In Table 2, the first and second entries in the parenthesis for the loglikelihood values reflect the loglikelihoods associated with the choice and price/coupon polynomials in Models 1 and 2.

[11] This result is to be expected for Model 3 because using prices net of coupons eliminates the coupon variable as a perfect predictor of choice.

Table 4
Model Estimation (based on Eq. (1) with Constraints $\beta = -\gamma$ and $\sigma_\beta = \sigma_\gamma$)

Parameters		Model 1B	Model 2B	Model 3B
Choice model	(Heinz)	6.48 (2.61)	4.84(1.27)	5.84(1.02)
Brand specific coefficient α_1	(Hunts)	5.73 (1.24)	5.58(1.71)	3.77(1.00)
Mean price coefficient β		– 4.67 (0.48)	– 4.75(0.51)	– 2.20(0.45)
Standard deviation of price coefficient σ_β		1.87 (0.35)	1.98(0.52)	1.89(1.56)
Common attribute A_1	(Heinz)	5.98(2.53)	5.18(0.49)	7.35(3.46)
	(Hunts)	4.91(1.61)	3.38(1.96)	3.68(1.93)
Preference heterogeneity parameter σ_μ		0.99(0.41)	0.92(0.41)	0.98(0.22)
Price Polynomial				
Brand-specific coefficient a_1	(Heinz)	13.87(2.51)	14.98(2.01)	
	(Store brands)	13.77(2.39)	14.08(2.11)	
	(Hunts)	13.18(2.48)	13.33(3.16)	
Size-specific coefficient a_s	(14 oz)	0.71(0.23)	0.74(0.24)	
	(28 oz)	– 1.84(0.25)	– 1.90(0.31)	
	(32 oz)	– 1.27(0.40)	– 1.37(0.43)	
	(44 oz)	– 2.03(0.35)	– 2.15(0.42)	
Quadratic price coefficient a_2		– 8.62(2.14)	– 8.36(2.56)	
Cubic price coefficient a_3		1.02(0.93)	0.69(0.86)	
Quadric price coefficient a_4		0.34(0.19)	0.38(0.57)	
Coupon Polynomial				
Probability of having no coupon δ_0	(Heinz)	0.86(0.016)	0.86(0.022)	
	(Store brands)	0.97(0.015)	0.98(0.016)	
	(Hunts)	0.87(0.061)	0.88(0.032)	
Brand specific coefficient b_1	(Heinz)	8.67(58.12)	10.68(55.12)	
	(Store brands)	7.82(58.19)	12.78(55.28)	
	(Hunts)	10.01(58.27)	10.79(55.12)	
Size-specific coefficient b_s	(14 oz)	– 6.46(58.14)	– 7.13(55.48)	
	(28 oz)	– 6.26(56.21)	– 6.59(13.32)	
	(32 oz)	– 8.68(73.22)	– 11.50(13.83)	
	(44 oz)	– 8.32(72.43)	– 7.48(15.16)	
Quadratic coupon coefficient b_2		– 3.22(44.61)	– 6.38(29.38)	
Cubic coupon coefficient b_3		63.13(44.43)	43.59(35.22)	
Q coupon coefficient b_4		– 62.17(64.12)	– 74.69(56.24)	
Overall log-likelihood value		– 6852.9	– 6844.6	– 1493.3.0
(Choice model π, price/coupon polynomials φ)		(– 1481.5, – 5371.4)	(– 1477.4, – 5367.2)	(– 1493.3, 0)

Table 5
Estimation of the ad hoc model without coupons: model 3C

Parameters		Model 3C
Brand specific coefficient α	(Heinz)	5.65(1.42)
	(Hunts)	4.51(1.32)
Mean price coefficient β		– 2.93(0.63)
Standard deviation of price coefficient σ_{β}		5.56(1.83)
Common attribute A	(Heinz)	7.21(2.47)
	(Hunts)	4.08(1.88)
Preference heterogeneity parameter σ_{μ}		1.28(0.63)
Log-likelihood value		– 1487.5

assess the impact of the policy change on market shares, one has to compare the appropriate baselines with the after-policy figures. Table 6 reports the simulation results. The results provide strong evidence of the bias created by the self-selection problem. In particular, the results show strong evidence for downward biased price effects obtained by ad hoc models. For example, a 10% cut by Hunts increases Hunts' market share by 64% in Model 1 and 72% in Model 2, whereas the predicted increases in market share in Models 3, 3B and 3C are only 46, 31 and 39%, respectively.

4. Conclusions

In this paper, we show that self-selection bias exists due to the missing prices and coupon availability data for non-purchased brands in scanner panel data on ketchup. We proposed a model to correct for this self-selection bias. Comparisons of our model with traditional models that do not account for the self-selection problem indicate that the effects of price reductions on market share are estimated to be two-thirds as great in models that have the self-selection problem. Our results also reveal that the self-selection bias will be more severe for lower-price brands such as store brands. Finally, our proposed modeling approach provides an alternative method that allows for the incorporation of coupons in the analysis without creating an endogeneity problem.

Acknowledgements

The authors thank Tom Wansbeek, Michel Wedel and two anonymous referees for their helpful comments. Support for this research was funded by NSF grant SBR-9511280.

Table 6
Simulation results

Brand		Baseline					Hunts cuts price by 10%				
	Sample	Model 1	Model 2	Model 3	Model 3B	Model 3C	Model 1	Model 2	Model 3	Model 3B	Model 3C
Heinz	64.2	62.2	62.8	63.4	60.9	61.9	52.1	50.8	56.3	57.8	56.5
Store brands	13.4	13.3	13.4	13.4	14.5	14.0	7.7	8.3	9.8	10.0	9.9
Hunts	22.4	24.5	23.8	23.2	24.6	24.1	40.2	40.9	33.9	32.2	33.6
							Hunts cuts price by 25%				
							Model 1	Model 2	Model 3	Model 3B	Model 3C
							18.9	15.9	26.8	27.4	26.1
							4.5	5.3	10.1	9.4	8.8
							76.6	78.8	63.1	63.2	65.1

Appendix A.

In this appendix we address some numerical issues. The functional forms (7) and (8) for the price and coupon probabilities allow one to estimate discrete probabilities for the observed discrete prices and coupons. Indeed, they are the underlying functional assumptions for our estimation. However, these functional forms offer numerical difficulties in the maximization of the above introduced likelihood function.

Since the likelihood function defined above cannot be maximized analytically, one has to employ some iterative numerical algorithm. Gradient-based search algorithms rely on incrementing the model parameters by small amounts and then re-computing the likelihood function so as to obtain numerical derivatives. However, if the parameters appearing in functions (7) and (8) are changed by a sufficiently small amount, then none of the discrete set of randomly drawn prices and coupons will change. Thus, the simulated likelihood is not a smooth function of the parameters appearing in Eqs. (7) and (8). To eliminate this problem, we approximate the discrete distributions defined in Eqs. (7) and (8) by continuous distributions obtained from using the simple smoothing technique described below. To describe the technique, we concentrate our attention on prices. Coupons are handled similarly.

Employing a suitable transformation we can set $\rho_{js1} = 0, j = 1, 2, \dots, J$ for any s where P_{js1} is the lowest price for brand j and size s. We can then order the prices so that $P_{j,s,q-1} < P_{jsq}, q = 2, 3, \dots, L$. With this we define $F(P^{\mathrm{c}}_{js)} = \sum_{m=1}^{q} \rho_{jsm}$, whenever $P_{j,s,q-1} < P^{\mathrm{c}}_{js} \leqslant P_{jsq}, q = 2, 3, \dots, L$, where P^{c}_{js} are the continuous prices. Further, $F(P^{c_{js}}) = \rho_{js1} = 0$ whenever $P^{\mathrm{c}}_{js} \leqslant P_{js1}$. The function $F(P^{\mathrm{c}}_{js)})$ is a step function because it has the same value for any P^{c}_{js} between $P_{j,s.q-1}$ and P_{jq}, and defines a piece-wise continuous approximation to the distribution of the discrete prices. With P_{js1} being the lowest price, the probability of P_{js1} and of prices lower than P_{js1} is zero in the distribution defined as above.

However, the function F is not one to one and, hence, it is not invertible. Smoothing F as follows solves the non-invertibility problem:

$$F^*(P^{\mathrm{c}}_{js)}) = E_q F\ (P_{j,s,q+1}) + (1 - E_q)F(P_{jsq}), P_{qjl} < P^{\mathrm{c}}_{js} < P_{j,s,q+1},$$

$$q = 1, 2, \dots, L - 1$$

where $E_q = (P^{\mathrm{c}}_{js} - P_{jsq})/(P_{j,s,q+1} - P_{jsq})$. It is clear that the image of F^* is the interval [0,1] so that F^* can be viewed as uniformly distributed in [0.1]. Now the continuous price simulators P^{c}_{js} can be obtained by drawing U^{c} from the uniform distribution over [0,1] and setting $P^{\mathrm{c}}_{j} = F^{*-1}(U^{\mathrm{c}})$.

References

Allenby, G., Rossi, P.E., 1991. Quality perceptions and asymmetric switching between brands. Marketing Science 14, 300–325.

Erdem, T., Keane, M.P., 1996. Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. Marketing Science 15 (1), 1–21.

Geweke, J., Keane, M.P., Runkle, D., 1994. Alternative computational approaches to inference in the multinomial probit model. Review of Economics and Statistics 76 (4), 609–632.

Guadagni, P.M., Little, J.D.C., 1983. A logit model of brand choice calibrated on scanner data. Marketing Science 2 (2), 203–238.

Heckman, J.J., 1974. Shadow wages, market wages and labor supply. Econometrica 42, 679–693.

Heckman, J.J., Sedlacek, G., 1985. Heterogeneity, aggregation and market wage functions: An empirical model of self-selection in the labor market. Journal of Political Economy 93, 1077–1125.

Keane, M.P., 1993. Simulation estimation for panel data models with limited dependent variables, In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics. Elsevier, Amsterdam.

Keane, M.P., 1994. A computationally practical simulation estimator for panel data. Econometrica 62 (1), 95–116.

Keane, M.P., Moffitt, R., Runkle, D., 1988. Real wages and the business cycle: estimating the impact of heterogeneity with micro data. Journal of Political Economy 96, 1077–1125.

Lerman, S., Manski, C., 1981. On the use of simulated frequencies to approximate choice probabilities. In: Manski, C., McFadden, D., (Eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA.

McCulloch, R., Rossi, P.E., 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64, 207–240.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers of Econometrics. Academic Press, New York, pp. 105–142.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimates. Econometrica 57, 1027–1058.

Tellis, G.J., 1988. Advertising exposure, loyalty and brand purchase: a two-stage model of choice. Journal of Marketing Research 30, 369–379.