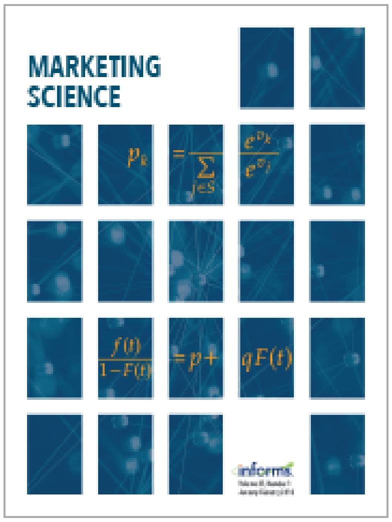


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Enriching Scanner Panel Models with Choice Experiments

Joffre Swait, Rick L. Andrews,

Enriching Scanner Panel Models with Choice Experiments

Joffre Swait • Rick L. Andrews
Advanis Inc., 12 W. University Ave. #205, Gainesville, Florida 32601, and University of Alberta
Ourso College of Business Adminstration, Louisiana State University, Baton Rouge, Louisiana 70803
Joffre_Swait@Advanis.ca • randrews@lsu.edu

This research examines the methods, viability, and benefits of pooling scanner panel choice data with compatible preference data from designed choice experiments. The fact that different choice data sources have diverse strengths and weaknesses suggests it might be possible to pool multiple sources to achieve improved models, due to offsetting advantages and disadvantages. For example, new attributes and attribute levels not included in the scanner panel data can be introduced via the choice experiment, while the scanner panel data captures preference dynamics, which is, at best, difficult with experimental data. Our application, involving liquid laundry detergent, establishes the feasibility and desirability of doing such augmentations of scanner panel data: The joint scanner panel/choice experiment model has significantly better prediction performance on a holdout data set than does a pure scanner panel model. Thus, we extend the concept of choice *data enrichment* into another domain and demonstrate that data enrichment can add significantly to one's understanding of preferences reflected in scanner panel data.
(*Data Enrichment; Choice Models; Scanner Panel; Choice Experiment*)

Introduction

The advent of scanner panel data in the 1980s constituted a major milestone for consumer packaged goods manufacturers, retailers, and marketing academics, because the data affords such deep insights into longitudinal consumer behavior. Focusing attention on academic research, the availability of this data source has permitted detailed study of consumer brand and volume choice in any number of product classes, with particular emphasis on the inclusion of marketing mix variables, pricing, consumer dynamics, and taste heterogeneity. To cite only a few of the most significant papers in this research stream, we mention Guadagni and Little (1983), Gupta (1988), Allenby (1990), Bucklin and Gupta (1992), Rossi and Allenby (1993), Fader and Hardie (1996), Ainslie and Rossi (1998), and Bucklin et al. (1998).

Among the most noteworthy strengths of scanner panel data are (1) the fact that the choices represent actual market transactions (denoted "revealed preferences" (RP) in the economics literature), (2) that individual households can be studied, (3) that the data capture the behavioral dynamics (e.g., learning, variety-seeking, taste evolution) inherent in sequences of choices over time, and (4) that context (e.g., advertising, in-store promotional mix, competitive sets) changes over time, permitting insights into its impact on brand and volume-buying behaviors. Despite these advantages, experience and theoretical considerations have given rise to a number of concerns about scanner panel data: (1) the difficulty, even impossibility, of analyzing "what-if" situations involving new product introductions and deletions; (2) problems with the quality of the environmental variables, such as correlation among marketing mix variables and incomplete information on prices and coupon holdings; (3) panel representativeness issues (Gupta et al. 1996); and (4) the lack of process measures, such

0732-2399/03/2204/0442
1526-548X electronic ISSN

as attitudes, to help untangle alternative explanations for behaviors, making it difficult to extend theoretical insights from consumer behavior and psychometrics into scanner panel analyses.

The intriguing possibility of combining multiple choice data sources to alleviate such data limitations was first recognized by Morikawa (1989), who used a travel mode selection choice experiment (called "stated preference" (SP) data) to augment/complement RP data collected in a survey. Since then, the combination of different types of choice data has continued to be of interest in transportation research (e.g., Swait et al. 1994, Brownstone et al. 2000) and econometrics (e.g., Hensher et al. 1999). In marketing, the discipline has been alerted to the possibility of data enrichment (e.g., Swait and Louviere 1993, Louviere et al. 2000), but to our knowledge, the only studies to combine panel data with choice experiments appeared in the environmental economics literature (e.g., Blamey et al. 2001, Blamey et al. 2001). Fader and Hardie (1996), who developed a feature-based conjoint-like model for scanner panel data, mentioned the possibility of combining supermarket scanner data with conjoint data from a choice experiment:

> [Routine grocery purchasing] is more natural and less obtrusive than that of any laboratory-based conjoint experiment. This touch of reality makes our model an appealing complement to the standard conjoint methodology. There are interesting possibilities of merging the two methods, especially because conjoint can incorporate new attributes (and levels) that do not currently exist in the market. The development of such a combined model would be a useful extension to this research.

Following this suggested direction, the objective of the present research is to investigate the feasibility, advantages, and disadvantages of combining household scanner panel data and choice experiments (Louviere and Woodworth 1983, Louviere et al. 2000), which use carefully configured experimental designs to determine the effects of variables on choice behavior. The structure of the remainder of this paper is as follows. First, we review in more detail the strengths and weaknesses of both scanner panel and experimental choice data. Second, we outline the statistical method underlying the combination of the two data sources in question. Third, we describe the scanner panel data we use and the choice experiment we have conducted, followed by the estimation results. Fourth, we discuss some of the insights gained from the data combination exercise and conduct holdout prediction comparisons. Finally, we close the paper with a discussion of the advantages and disadvantages of the proposed data combination method and point to a number of research extensions we believe are warranted.

Literature Review

The emergence of scanner panel data has revolutionized market response research. The availability of actual market transactions by individual households has the potential to allow research rich in implications for managerial decision makers (Bucklin and Gupta 1999). Academics and industry consultants have used scanner panel data to investigate pricing, trade promotion, consumer promotion, advertising, product policy and strategy, and distribution and retail management. Experimental choice data are generally collected in surveys administered in a variety of ways, including self-administration, personal interviews, and web surveys. Because experimental choice data have a different set of strengths and weaknesses than scanner panel data, the combination of the two data sources has the potential to mitigate the shortcomings of each source taken separately.

Unlike scanner panel data, experimental choice data do not represent actual market choices. As such, stated choice intentions are likely to be subject to an upward bias since choices are less likely to consider constraints (e.g., financial) that operate in real markets (see Morwitz and Schmittlein 1992, Morwitz et al. 1999). This characteristic of SP data often raises qualms about their use in such disciplines as economics. Marketing, more commonly faced with the need of planning for market circumstances that fall outside the scope of historic market conditions, has developed a more balanced approach, even a certain affinity, to the use of hypothetical choice data. Nonetheless, the discipline recognizes that SP data can overstate the demand for goods, given the hypothetical nature of the market context, and thus, scanner panel data would contribute invaluable realism if combined with choice experiment data.

However, since the choices in scanner panel data occurred in an uncontrolled environment, we cannot analyze "what-if" situations involving new product/attribute introductions and deletions without making burdensome and simplifying assumptions (Fader and Hardie 1996). In contrast, choice experiments can easily extend ranges of attributes beyond what exists in historical data, and it can also straightforwardly manipulate the presence/absence of brands and/or products (Lazari and Anderson 1994).

Another problem is that price and promotional data are often highly correlated in scanner panel data, making identification of these effects more questionable. Experimental choice data are less likely to face parameter identification problems related to poor conditioning of the matrix of explanatory variables, given the orthogonality or near-orthogonality of the experimental design, increasing the reliability and precision of statistical estimates.

There are other limitations in the shopping environment information in scanner panel data, as well. Only the prices of the items bought are recorded in some scanner panel data sets, and ad hoc methods must be used to fill in the missing prices of nonpurchased brands (Erdem et al. 1999). No such problem exists with choice experiments because the prices of alternatives are determined by the experimental design. While data are usually collected on coupon redemption of scanner panelists (i.e., we know which panelists redeemed a coupon), there is usually no information on when a panelist had a coupon available but decided not to use it (Bucklin and Gupta 1999). In choice experiments, the availability and value of coupons can be changed according to an experimental design that will permit estimation of their impact on behavior. Note, however, that the manner in which this type of context manipulation is done is critical to the realism of the choice task. For example, if coupons must be clipped from a newspaper and brought to the store with the shopper, manipulating their value in the experiment would assume that responsive consumers are actually coupon-clippers. Today, however, it is common to see coupons being distributed in-store via flyers, electronic dispensers, and at checkout registers. Thus, a choice experiment can more realistically manipulate the presence and value of such in-store coupons. Currently, it is feasible to implement choice experiments in virtual environments that would allow detailed study of context variables heretofore manipulated only with great difficulty (if at all) in traditional choice experiments: Examples might be horizontal and vertical positioning of products on shelves, different product size arrangements than are normally used, different packaging designs, availability of end-of-aisle displays, and so forth (see, e.g., Burke et al. 1992).

In addition, other, more ancillary, advantages arise from the fact that the data collection methods used to administer choice experiments would usually differ from those used to collect scanner panel data. The sample selection process for scanner panel data is typically not under the control of the researcher designing the choice experiment, but is instead designed and executed by an independent organization such as ACNielsen or Information Resources, Inc. As a result, the sampling errors and biases present in experimental data may be different from those present in scanner panel data. As shown by Gupta et al. (1996), scanner panel households' purchase behaviors may not be representative of the population of all households shopping at the stores due to (1) nonprobability (judgment) sampling of households, (2) refusal by selected households to participate in the panel, (3) attrition in the panel because of households dropping out, and/or (4) panelists' incomplete use of identification cards to record purchases in the store. In contrast, for choice experiments, the selection methods used to sample respondents and the tasks required of respondents will likely result in *different* errors and biases than those present in scanner panel data, resulting in a possible overall improvement in sample quality and, therefore, parameter estimates when the two data sources are combined. Kim and Rossi (1994) report that consumers with high purchase frequency or volume are more price sensitive than consumers with low frequency or low volume of purchase, so sample representativeness *is* a relevant issue.

Finally, there are issues with the value of the data available for describing panelists. The company collecting the scanner panel purchase data (e.g., ACNielsen) also administers surveys to the panelists to collect the descriptive data, so this process is not under the control of the researcher administering the choice experiment. Previous studies using scanner panel data (e.g., Bucklin and Gupta 1992, Gupta and Chintagunta 1994) have found only very weak relationships between household demographics and marketing mix sensitivities using scanner panel data (though, Ainslie and Rossi 1998, found somewhat better relationships by pooling data from multiple product categories to improve the signal-to-noise ratio). In addition, there are no process measures, such as attitudes, nor consumption measures, to help one untangle alternative explanations for behaviors due to the fact that only purchase behavior is recorded (Winer 1999). In contrast, the surveys used for choice experiments can gather detailed information on expectations, beliefs, attitudes, perceptions, and psychographic and sociodemographic characteristics that can be used to improve the specification and identification of demand models. This is particularly important given the recent trend in choice modeling of exploring the behavioral mechanisms that underlie consumer choice processes, especially consumer dynamics (e.g., Erdem and Keane 1996, Gönül and Srinivasan 1993, Keane 1997). (We thank an anonymous reviewer for suggesting this point.)

The fact that different choice data sources have diverse advantages and disadvantages, and the subsequent insight that it might be possible to pool multiple sources to achieve improved predictive models (due to offsetting advantages and disadvantages), has led to a stream of research in transportation and environmental economics aimed at developing the methods for and testing the validity of such *data enrichment* efforts. This research stream has demonstrated that data enrichment is feasible, valid, and useful from both statistical and substantive perspectives. Louviere et al. (2000, Chapter 13) present supportive evidence that demonstrates that the basic underlying theoretical insight of data enrichment, namely, that response coefficients estimated from choice data (and more generally, preference data) arising from multiple elicitation procedures are often equivalent, up to a scale factor proportional to the relative variance ratio between data sources (see also, Swait and Louviere 1993). That is to say, *if two different elicitation procedures (say, RP and SP) are used to estimate utility partworths for attributes, then if the two sources are indeed capturing the same utilities, their partworths will be equal up to a scale constant that is related to the relative precision of measurement inherent in each source*. We elaborate, subsequently, on this theoretical and empirical insight of the "data enrichment" research stream.

While in this paper we shall be talking about scanner panel and experimental choice data, it must be noted that the data enrichment insight is more general and applies to any number of choice data sources that share some set of common attributes (see also, Swait and Louviere 1993, Swait and Bernardino 2000). For example, Louviere et al. (2000, Chapter 13) use the analysis by Deighton et al. (1994), which examined the impact of advertising on switching and repeat purchase behavior for scanner panel purchases of ketchup, liquid detergent, and powder detergent, and provide strong evidence that the impact of advertising on preferences is equal across the three product categories, once scale differences are accounted for. In support of the generalizability of this insight, we mention, also, an independent study by Andrews and Currim (2002), who pool scanner data from three product categories (liquid laundry detergent, paper towels, and margarine), controlling for scale differences between categories. Their results show that 32% of sample panelists have the same responses to price, store feature advertising, aisle display, and state dependence, up to a scale factor proportionality, across at least two of the three product categories.

Data Enrichment

To set the stage for subsequent model development, it is useful at this point to summarize the method of data enrichment (see also Louviere et al. 2000, Chapter 8). Suppose there are two data sources, arbitrarily called RP and SP (which represent scanner data and choice experiment data, respectively, in the current context). In the former data, choices are observed from choice set C^{RP}, and alternatives $i \in C^{RP}$ are described by attribute column vectors X_i^{RP} and Z_i. Each alternative has random utility $U_i^{RP} =$

$\beta^{RP}X_i^{RP}+\gamma Z_i+\varepsilon_i^{RP}$, where $i\in C^{RP}$, β^{RP}, and γ are conformable response parameter row vectors, and ε_i^{RP} is an error term. Assuming the ε's are IID Gumbel with scale factor $\mu^{RP}=\pi/\sqrt{6}\sigma^{RP}$, where σ^{RP} is the standard deviation of the common error term distribution, the choice probabilities in this data source are given by the following MNL model (see Ben-Akiva and Lerman 1985):

$$P_i^{RP}=\frac{\exp[\mu^{RP}(\beta^{RP}X_i^{RP}+\gamma Z_i)]}{\sum_{j\in C^{RP}}\exp[\mu^{RP}(\beta^{RP}X_j^{RP}+\gamma Z_j)]}. \qquad (1)$$

The inverse relationship between the scale factor μ^{RP} and the standard deviation σ^{RP} means that as the standard deviation approaches infinity (i.e., choice is essentially random), scale approaches zero; conversely, as the standard deviation approaches zero (i.e., choice is essentially deterministic), scale approaches infinity. When only the RP data source is being analyzed, the scale factor μ^{RP} is not identifiable and is implicitly assumed to have a value of one.

In the SP data source, the set of alternatives among which choice is exercised may be different from that of the RP data set, and so we denominate it C^{SP}. Also, the characterization of alternatives is accomplished with two vectors of attributes, namely, X_i^{SP} and W_i. Thus, SP alternatives have in common with RP alternatives the vector of attributes X, but the data sources may differ in describing their respective alternatives through attributes Z and W. In the SP data source, the sensitivity parameters are β^{SP} and θ, respectively, corresponding to X^{SP} and W. Assuming again that the SP error terms in a random utility specification are IID Gumbel, but with scale factor $\mu^{SP}=\pi/\sqrt{6}\sigma^{SP}$, the choice probabilities are

$$P_i^{SP}=\frac{\exp[\mu^{SP}(\beta^{SP}X_i^{SP}+\theta W_i)]}{\sum_{j\in C^{SP}}\exp[\mu^{SP}(\beta^{SP}X_j^{SP}+\theta W_j)]}. \qquad (2)$$

As was the case with the RP data, when only the SP data source is being analyzed, the scale factor μ^{SP} is not identifiable and is implicitly assumed to have a value of one.

The essential underpinning of data enrichment is to note that the two data sources have in common the attribute vector X, for which it will be assumed during pooling that the response parameters are equal (i.e., $\beta^{RP}\equiv\beta^{SP}\equiv\beta$). However, it will be noted that in each data source, these common parameters are multiplicatively impacted by their respective scale factors (see Expressions 1 and 2). Hence, to avoid confounding the scale factor μ and tastes β, it is necessary to control for differences in scale between the data sources when imposing the condition of equality of β for attributes X. *Thus, the basic data enrichment exercise imposes equality of common response coefficients β, while allowing scale/variance to differ between sources*. Because the scale factor of a single data source is not identifiable, what is actually estimated in a data enrichment exercise is a relative scale factor between data sources. For example, in the RP-SP data enrichment we have been discussing, we estimate the ratio μ^{SP}/μ^{RP} by normalizing μ^{RP} to unity to allow identification of μ^{SP}. Previously, we had referred to the combination of two data sources as involving their partworths being equal up to a scale constant that is related to the relative precision of measurement inherent in each source: This relative precision is given by the ratio μ^{SP}/μ^{RP}.

This stylized model of the data fusion process is sufficiently flexible for the purposes of this research. It recognizes that the basis for the pooling of various choice data sources is a (sub)set of common attributes or variables (X's) for which we desire improved statistical estimates of partworths through the pooling process. It also permits the consideration of source-specific factors during pooling, such as differing choice sets (C's) and attributes (W's and Z's). As we shall see below, in the pooling of experimental choice and scanner panel data, there are a significant number of variables that are source-specific. We should also note here that the use of the MNL model, above, is merely for illustrative purposes; analogous scale/variance ratio relationships would exist were Multinomial Probit or any form of random parameter/finite mixture models to be used instead.

Empirical Application

In this section, we present the results of a data enrichment exercise using liquid laundry detergent scanner panel data and experimental choices elicited through a mail survey. We first present details of the two data sources, then discuss the general specification of the source-specific and pooled models. Following, we present source-specific models, as well as data enrichment results, and a discussion of estimation results.

Data Description

Scanner Panel Data. Information Resources, Inc. (IRI) panelists, located in a Chicago suburban area, are tracked over the 112-week period from September, 1995 to November, 1997. The first 34 weeks of purchases are used to initialize loyalty variables, with the remaining 78 weeks used for model estimation and validation. Five hundred twenty-six total panelists were randomly divided into two groups: 400 panelists (or about 3/4 of the total), making 2,546 liquid laundry detergent purchases, comprise the data used for model estimation; another 126 panelists, totaling 725 purchases, are used for model validation. All panelists purchased from among 84 UPCs of liquid laundry detergent.

Liquid laundry detergent is characterized by a reasonable number of attributes, though some of these have a large number of levels; this has a significant impact on the experimental choice data source. These attributes (see Table 1) are brand, concentration, formula, biodegradability, scent, type of deal, discount, size, number of loads, regular price, feature ad, and aisle display. A visit to a supermarket indicated that, at the time of this writing, and for some time before, most liquid laundry detergent packages contained information on the number of wash loads that each UPC is likely to yield. Our analysis of current offerings in the market shows that the number of loads is simply calculated as 30% of the size (in ounces) of the package, to a very close approximation, so in the scanner panel data this attribute is basically confounded with size. In addition, and quite interestingly, it should be noted that there is no difference between the number of loads ratings for Regular and Ultra concentration.

In the scanner panel, data source information on the attribute levels of different UPCs is quite sparse. In addition to the examples already cited, information on biodegradability is missing for most UPCs, and type of deal and amount of promotion is missing for all UPCs except the one purchased by the panelist.

Table 1 Liquid Laundry Detergent Scanner Panel and SP Attributes and Levels

Attribute	Scanner Panel Levels	SP Levels
Brand	Ajax	Ajax
	All	All
	Arm & Hammer	Arm & Hammer
	Cheer	Cheer
	Era	Era
	Fab	Fab
	Gain	Gain
	—	Purex
	—	Store brand
	Surf	Surf
	Tide	Tide
	Wisk	Wisk
	XTR11	XTR11
	Yes	Yes
Concentration	Regular	Regular
	Ultra	Ultra
Formula	Regular	Regular
	Color safe bleach	Color safe bleach
	No dye or perfume	No dye or perfume
	Bleach alternative	Bleach alternative
	Stain remover	Stain remover
	Fabric softener	Fabric softener
	—	Wrinkle reducer
	—	Allergen treatment
	—	Chlorine neutralizer
Biodegradable	Information not available for all UPCs	Yes
		No
Scent	Regular	If formula = "No" dye or perfume, Unscented
	Unscented	Else, Unscented
		Regular
		Herbal
		Mountain Spring
		Fresh Rain
		Floral
Type of dispenser	Regular cap	Regular cap
		Dispenser pump
Type of deal	Information on deal origin available only for purchased UPC	If brand = store brand, none (50% of items shown)
		Store promotion
		Else, None (50% of items shown)

Table 1 (cont'd.)

Attribute	Scanner Panel Levels	SP Levels
Type of deal		Store promotion Coupon available in-store Requires preclipped coupon
Discount	Information on size of discount available only for purchased UPC	If type of deal = none, 0% off regular price Else, 5% off regular price 10% off regular price 25% off regular price (Calculated to nearest rounded nickel.)
Size (oz.)	50 oz. 64 oz. 90 oz. 100 oz. 128 oz. 200 oz.	If top shelf, 50 oz. 64 oz. 90 oz. Else If middle shelf, 90 oz. 100 oz. 128 oz. Else bottom shelf, 128 oz. 200 oz. 300 oz.
Regular price ($)	Regular price of UPC during shopping trip in question	$= \text{Size} * (\text{Global price for scenario} + \text{price deviation by brand}) * (1 - \text{discount})$
Global price for scenario ($/oz.)		\$0.03 \$0.06 \$0.09
Price deviation by brand ($/oz.)		−\$0.015 \$0.00 \$0.015
Number of loads	$= 0.30 * \text{size}$, so confounded with size	$= 0.3 * \text{size}\ (1 + \text{number of loads perturbation factor})$
Number of loads perturbation factor		−0.05 0.00 0.05
Store feature advertisement	None "A" feature "B" feature "C" feature "Super A" feature Collapsed to 0/1 (No/Yes) due to sparse data	
Aisle display	None Lobby display Front end-aisle display Mid-aisle display Back end-aisle display Specialty display Shipper display Promotional display Collapsed to 0/1 (No/Yes) due to sparse data	

In the case of scent, the RP data permits distinguishing regular scent from nonregular scent. In addition, though UPCs of the Purex brand are present in the stores during the period of the scanner panel, none of the panelists chose any of these UPCs; nor was there any liquid detergent store brand at all in the data. Due to the fact that three and one half years passed between the period covered by the scanner panel and the time of collecting the experimental choice data (see below), certain formula changes have become common in the market: It is now common to see liquid laundry detergents featuring wrinkle reducers, allergen treatment, and chlorine neutralizer. The scanner panel data has no UPCs with these formulas.

Thus, the scanner panel choice data has certain incomplete or missing information (brand, formula, scent, biodegradability, type of deal, and discount), some that is outdated (lack of certain formulas, lack of store brands, lack of 300 oz. size), and some that is unique (feature ad and aisle display).

Experimental Choice Data. To complement the scanner panel data just described, an experimental choice task was incorporated in a mail survey that was sent in October, 2000, to a random sample of households in the Chicago metropolitan area, largely consistent with the geographic scope of the scanner panel. The choice task, an example of which (with survey instructions) is shown in Figure 1, was designed to simulate the usual shelf arrangement found in supermarkets: UPCs are arranged by size, with smallest at the top and largest at the bottom. Each task contained thirteen UPCs: twelve experimentally designed profiles, described by the attributes in Table 1, arranged in

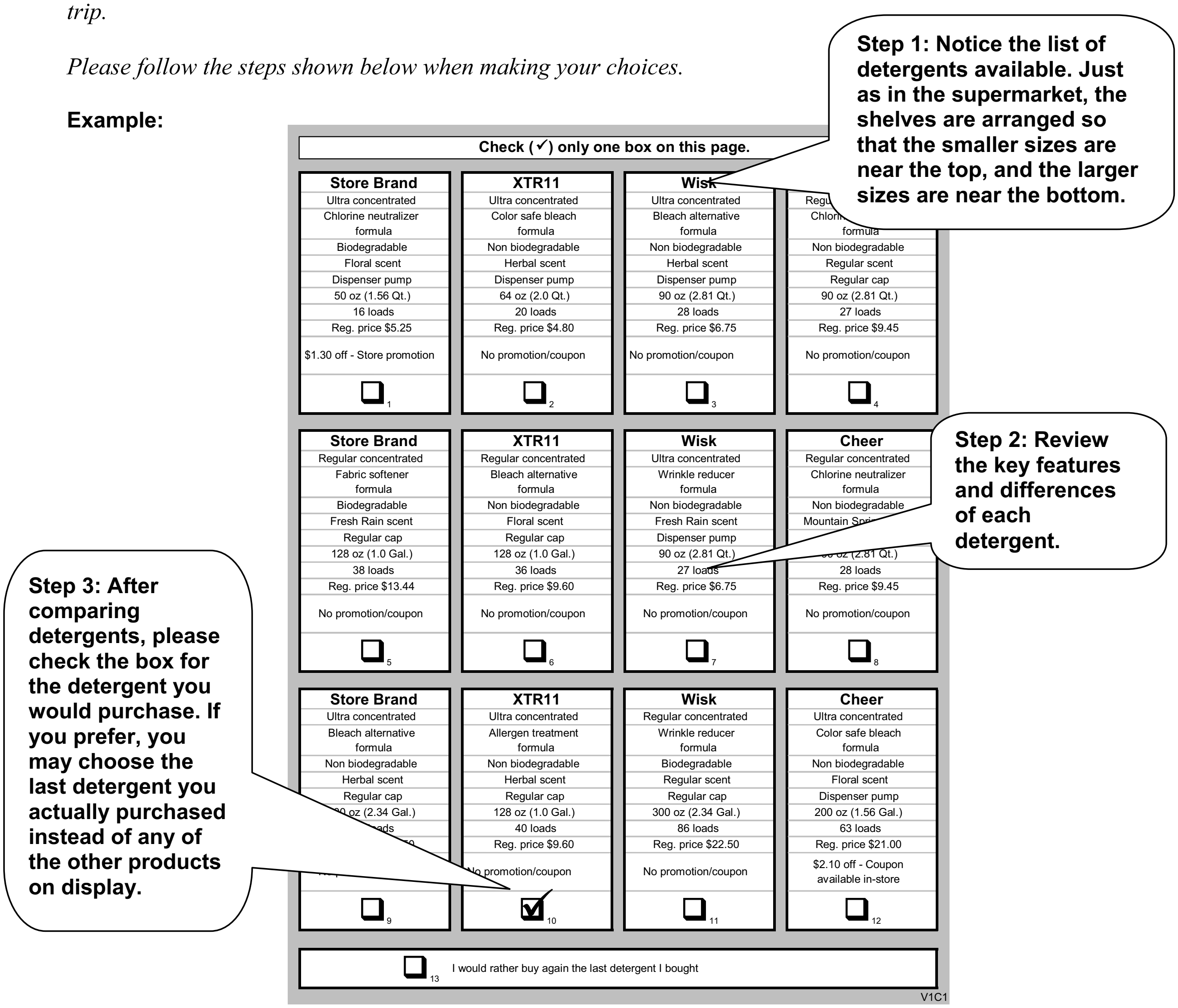
Figure 1 Laundry Detergent Choice Task Layout

The next 8 pages contain descriptions of different detergents (all liquid) like the ones you might find on your next trip to the supermarket. Suppose that on your next trip to buy detergents, you find the products shown (and only these) available at the store. (If you would prefer to buy a powder detergent, we would ***still*** *like you to tell us which of the products described you would most prefer if these were the only options available.)*

After comparing the various features of each detergent, please indicate which product you would choose. ***A description of these features is given on the facing page.*** *You have the option to choose any* ***one*** *of the detergents or, if you prefer, the last detergent you purchased and which you described in the previous pages. Please treat each of the 8 choice situations independently, as if each one were a different shopping trip.*

Please follow the steps shown below when making your choices.

Example:



Remember, you should choose only **one** product on each page.

the 3×4 grid shown in Figure 1; the thirteenth product, the base, was the person-specific UPC last purchased by the individual. A description of this last UPC, in terms of the attributes of Table 1, was obtained via the survey. Note that all three UPCs shown in a given column of a choice set are of the same brand, so that three different brands (plus, perhaps, a fourth brand in the thirteenth UPC) are present in each choice scenario.

To overcome some of the data deficiencies of the scanner panel data source, the choice experiment used more complete lists of attribute levels than are present in the scanner panel data. Specifically,

(1) We include Purex and a store brand among the brands manipulated;

(2) We add wrinkle reducer, allergen treatment, and chlorine neutralizer to the formula attribute;

(3) We specify six scents to permit differentiation among UPCs, which the scanner panel data does not do;

(4) We introduce a new attribute, type of dispenser, which in the scanner panel data has only the regular cap level, but in the experiment, has the additional level of a dispenser pump, such as can be commonly found on hand soap dispensers;

(5) We specify the type of deal that an UPC has ongoing, being careful to control that promotions do not occur too frequently across choice scenarios. The type of promotion can be store or manufacturer initiated, and the source of the corresponding coupons is specified (preclipped from home, or in-store). In association, we also control, independently, the amount of discount that is offered;

(6) We introduce the 300 oz. size, present today in supermarkets, but nonexistent in the scanner panel data;

(7) Much greater price variation was introduced both within and across choice scenarios, as compared to the scanner panel data, which creates decreased price tracking between brands, and allows improved statistical efficiency in price sensitivity estimation. Table 1 shows how UPC prices are calculated by scenario, brand, and size;

(8) Finally, the confound between number of loads and size was eliminated by the introduction of a designed perturbation factor.

Thus, the choice experiment allows the introduction and identification of a significant number of effects omitted from or nonexistent in the scanner panel data. The large number of levels of certain attributes (e.g., brand, scent) and the nestings (or interdependencies) between attributes present in this design (see Table 1, again) make an orthogonal design strategy somewhat onerous in this study. (This is not to say that it could not have been done, of course.) We opted instead for a compromise design created by applying an interchange heuristic[1] to an initial random design using a pool of candidate choice sets, until a final design with reasonably low intercolumn correlations was generated using 480 choice sets of the type shown in Figure 1. These were then blocked into 60 versions of eight choice sets. Respondents were randomly assigned to each survey version.

The survey contained 18 ($8.5'' \times 11''$) pages in the form of a booklet. Its sections were "General Laundry Habits," "General Detergent Usage," "Most Recent Purchase," "Choosing Liquid Detergent," and "Household Characteristics." While these titles are generally indicative of the type of information elicited in each section, we would like to highlight some particular pieces of information that we will employ in the modeling of the SP data. Specifically, the "Most Recent Purchase" section was needed to obtain a complete characterization (in terms of the attributes in Table 1) of the last laundry detergent purchased by the respondent. This last purchase could be a powder detergent, which was our motivation for inclusion of this UPC as the thirteenth alternative in all choice sets (see Figure 1); thus, all respondents, even powder-loyalists, could perform the choice tasks presented in the survey. This data also allowed us to include in the SP model some control for consumer state dependence with respect to brand, size, scent, and biodegradability via the definition of appropriate dummy variables related to the last purchase. This same section elicited additional information related to

[1] See Sándor and Wedel (2002) for recent developments on the creation of choice experiments that are optimal for mixed logit or other random effects models, and Toubia et al. (2003) for a new adaptive question design algorithm permitting adaptation within respondents, based on the respondent's answers to previous questions.

certain attribute cutoffs to implement the cutoff utility estimation method of Swait (2001); the cutoffs elicited were for brand, biodegradability, size, scent, and maximum price. For example, to provide information for the maximum price cutoff, respondents were asked, "What is the most that you would be willing to pay for a container of detergent, no matter what the form or size?" Such cutoffs, straightforwardly elicited from respondents in the survey, are not available in scanner panel data; as we shall subsequently see, they are *very* important in explaining choice behavior.

The survey booklet was mailed out, with a recruitment letter on a university's letterhead, to a random sample of 2,000 Chicago area residences. The recruitment letter asked the person responsible for laundry decisions in the household to take 15 to 20 minutes to complete the survey and emphasized the scientific importance of their participation; a toll-free number was provided in case of any questions concerning survey legitimacy or about clarification of instructions. No financial or other types of incentives were offered. Approximately ten days after sending the initial survey, a postcard reminder was sent to respondents whose surveys had not yet been received back; the postcard urged completion, and offered to mail out a second survey document if the first had been misplaced. No other follow-up was performed. A total of 407 surveys were processed by the end of the cutoff date, and of these 384 were usable for the modeling of SP choice behavior. This difference arises mostly because a number of individuals did not complete their choice tasks. The usable sample created a total of 3,026 SP choices.

It is apparent that the experimental choice exercise is able to (1) easily manipulate certain market variables that would be impossible or prohibitively costly to do in real markets (e.g., availability of brands, presence of store brand, new attributes such as type of dispenser, wider price ranges); and (2) introduce richer respondent characterization than is currently available from scanner panel data (e.g., cutoffs, brand perceptions). Nonetheless, the choice experiment is not a particularly good vehicle for manipulating the effects of advertisement (i.e., feature ad) and in-store merchandising (i.e., aisle display) (at least with the delivery mechanism used for this survey, though one could envision far more expensive data collection options that would allow realistic simulation of advertising and display effects in choice tasks). For these effects, we must strictly rely on the RP data source. In addition, the scanner panel data is able to yield time-dependent loyalty measures based on households' choice patterns; in a survey context, this would be difficult to do, particularly due to poor respondent memory for purchases of this type.

Thus, each data source has its strengths and weaknesses. The next section discusses the data enrichment model specification and estimation.

Table 2 Specification of Detergent Scanner Panel and SP Utility Functions

	Scanner Panel Utility	SP Utility
Functional Attributes		
Brand	✓ (subject to limitations)	✓
Concentration	✓	✓
Formula	✓ (subject to limitations)	✓
Biodegradable	×	✓
Scent	✓ (subject to limitations)	✓
Type of dispenser	×	✓
Type of deal	×	✓
Discount	×	✓
Size	✓ (subject to limitations)	✓
Number of loads	×	✓
Marketing/Merchandising Mix		
Regular price	✓	✓
Store feature Ad	✓	×
Aisle display	✓	×
Household Heterogeneity		
Attribute cutoffs	×	✓
Random coefficients	✓	✓
State Dependence		
Last brand	×	✓
Last size	×	✓
Last scent	×	✓
Last formula	×	✓
Last biodegradability	×	✓
Brand loyalty (G&L 1983)	✓	×
Size loyalty (G&L 1983)	✓	×
Scent loyalty (G&L 1983)	✓	×
Formula loyalty (G&L 1983)	✓	×

✓ = included. × = not present, not estimable.

Model Specification, Estimation, and Results

Specification of Scanner Panel and Experimental Choice Utility Functions. The previous section pointed out that certain effects are estimable only from one or the other of our data sources, while some are estimable from both. Table 2 summarizes the specification of the utility functions for the scanner panel and SP data.

The SP utility for a UPC can be seen to include a fuller representation of functional attributes (i.e., through additional levels and attributes). In terms of the marketing and merchandising mix, only Regular Price is included in this data source. Finally, a different type of household heterogeneity, captured through attribute cutoffs, is available in the SP data, in addition to a representation of consumer state dependence with respect to brand, size, scent, formula, and biodegradability via dummy variables that capture whether an UPC has the same level of a given characteristic as the most recent purchase made by the SP respondent. The scanner panel data has a partial representation of the functional attributes and a full representation of the traditional marketing and merchandising variables (besides price, also feature ad and aisle display status); with respect to household state dependence, the scanner panel model has Guadagni and Little (1983) loyalty variables for brand, size, scent and formula, which generalize the last purchase dummies described before for the SP data.

Parameter Estimation Method. In both data sources, Random Coefficient MNL (or Mixed MNL; see McFadden and Train 2000) models are estimated via simulated maximum likelihood. We associated independent normal distributions with (1) brand-specific constants (BSCs), to capture brand-related taste heterogeneity; (2) size; (3) regular price; and (4) aisle display status (feature ad was also tested, but never found significant). The log likelihood function for the RP source is

$$L(\psi^{RP}) = \sum_n \ln\left(\int \left(\prod_{r=1}^{R_n} P_{ni_r^*}(\psi_n^{RP})\right) f(\psi_n^{RP} \mid \psi^{RP})\, d\psi_n^{RP}\right), \tag{3}$$

where i_r^* refers to the chosen alternative at the rth purchase occasion, ψ_n^{RP} is a vector of parameters for consumer n, R_n is the number of choices, f is the multivariate density of the taste parameters (assumed multivariate normal with diagonal covariance matrix and having mean taste parameters ψ^{RP}), and $P_{ni_r^*}$ is the MNL choice probability of the rth chosen alternative of consumer n evaluated at ψ_n^{RP} (cf. Equation 1):

$$P_{ni_r^*}(\psi_n^{RP}) = \frac{\exp(\beta_n X_i^{RP} + \gamma_n Z_i)}{\sum_{j\in C^{RP}} \exp(\beta_n X_j^{RP} + \gamma_n Z_j)}. \tag{4}$$

The log likelihood for the SP source is defined analogously to (3), except that i_r^* refers to the chosen alternative in the rth choice scenario and (cf. Equation 2):

$$P_{ni_r^*}(\psi_n^{SP}) = \frac{\exp[\mu^{SP}(\beta_n X_i^{SP} + \theta_n W_i)]}{\sum_{j\in C^{SP}} \exp[\mu^{SP}(\beta_n X_j^{SP} + \theta_n W_j)]}. \tag{5}$$

The log likelihood of the joint model is simply

$$L(\psi) = L(\psi^{RP}) + L(\psi^{SP}), \tag{6}$$

because the data sources are independent.

During pooling, equality restrictions of the *means* of taste parameters β across the two data sources are imposed across common variables (see prior discussion centered around Expressions 1 and 2). The issue arises as to whether one should equate parameters of the density functions other than their means. The data enrichment process has, to this point, addressed reasons that the *means* of parameter distributions might be equal, up to scale. The same type of reasoning does not apply, for example, to the dispersion (i.e., variance) of the parameters within a data source. In addition, it behooves one to consider that parameter distributions from SP tasks may be reflective of attribute range effects (see the work of Ohler et al. 2000).

As indicated in Table 2, a number of variables are unique to each data source, and so are not restricted during pooling. The BSCs, in particular, are not restricted across data sources because such constants are strongly associated with the given data source. In fact, they partially reflect the relative opportunity that a brand had to be chosen (see Ben-Akiva and Lerman 1985, Chapter 5). In the scanner panel, the mechanisms that control the availability of a brand for choice have to do with the processes of physical distribution, shelf space allocation, store staffing, store inventory policy, and in-store merchandising activities. In the choice experiment, on the other hand, the availability of the brand for choice is completely controlled by task layout and the experimental design. Thus, in our opinion, it is highly inappropriate to restrict brand constants to be equal across such different data sources as we have in this research.

Model Estimation Results. Table 3 presents the estimation results for a number of models, which we discuss subsequently. For each data source, a Mixed MNL model is presented, as is a standard MNL model for simple comparison purposes. The latter shall not be explicitly discussed, other than to note that in both the scanner panel and SP data sources, the addition of taste heterogeneity to the statistical specification is most helpful to improving goodness-of-fit. The data-source specific Mixed MNL models (Model 2 for the scanner panel data and Model 4 for the experimental choice data) and the joint model for the combined data (Model 5) have all significant parameters of the expected signs.

The log likelihood of the joint model (−10,335.1) and the sum of the log likelihoods for the separate models (−10,242.7) are quite close, considering the large number of observations in the combined data (5,572). Given that constraints are imposed on the joint model, it is, of course, impossible for the log likelihood of the joint model to be better than the sum of the separate RP and SP likelihoods. The values of $\bar{\rho}^2$ and the Percent Right Predictions suggest (informally) that there might be some improvement in parameter-adjusted fit and prediction accuracy from estimating a joint model with pooled data. For example, if we create a weighted average of the RP and SP $\bar{\rho}^2$ values, where the weights are sample sizes in the two sources, we obtain a value of 0.4436, which is smaller than the joint model value (0.4529). Likewise, forming the same weighted average for the Percent Right Predictions, yields a value of 26%, which is slightly smaller than the joint model value (26.8%).

This gives rise to the "acid test" for the joint model: prediction to a holdout set of scanner panel choices. For this purpose, as previously mentioned, we reserved 126 households with 725 choices from the scanner panel data set. Table 4, Part a, reports several prediction measures calculated from this holdout data set for Models 2 (RP-only model), 4 (SP-only model), and 5 (joint-RP/SP model). Model 2 is a straightforward application of the estimated RP model to a validation sample. Model 4 uses the brand names, concentrations, formulas, and scents, and the size and regular price coefficients estimated from the SP data, to predict holdout scanner data choices; those variables used in the estimation of the SP model, but not present in the scanner data, are assumed to have zero values. Note that applying the SP model (which was calibrated using choices generated by the scenarios/choice sets presented in the experimental design) to the RP data assumes that its BSCs reflect the overall distribution of choices in the scanner panel data, which is extremely unlikely. This lack of adjustment to a base RP situation will most likely mean that the SP coefficients will predict very poorly to RP data because the brand-specific constants are at incorrect average levels. In addition, the SP-only model does not include any G&L state dependence effects, aisle display, or store feature ad variables, which we know from the RP models to be quite significant explanators of choice behavior. Thus, while Model 4's poorer performance is expected a priori, we believe it instructive to see how these SP coefficients will perform relatively on the holdout data. The main comparison of interest is between Models 2 and 5.

Model 5, the joint model, is applied to the holdout data by assuming zero values for all SP-specific variables (including the last purchase state dependence variables). Thus, the only coefficients affecting the fit of the joint model to the holdout data are the RP-specific coefficients (e.g., aisle display) and the coefficients shared by the RP and SP specifications (e.g., regular price). The joint model has significantly better prediction measures than the RP-only model, supporting our contention that the data enrichment exercise is quite useful in improving the quality of the parameter estimates of the resulting model. Specifically, the ρ^2 measure obtained by applying Model 5 to the holdout data is significantly higher than that of Model 2; this implies that the likelihood of the observed holdout choices is significantly higher using Model 5 than Model 2, which is confirmed in the table. The Pearson's chi-squared ratio, χ_p^2, should

Table 3 Model Estimation Results

Utility Parameter Distributions	Scanner Panel (RP) (Asymptotic *t*-Stats) MNL ❶	RCMNL ❷	Experimental Choice (SP) (Asymptotic *t*-Stats) MNL ❸	RCMNL ❹	Joint (Asymptotic *t*-Stats) RCMNL ❺
Means					
Brand					
Ajax	−0.190 (−0.74)	−0.555 (−1.99)	−0.804 (−5.90)	−0.841 (−2.69)	−4.141 (−1.89)/ − 0.638 (−2.48)
All	0.690 (3.73)	0.365 (1.47)	−0.150 (−1.60)	−0.394 (−2.16)	−0.730 (−0.48)/0.314 (1.42)
Arm & Hammer	0.553 (3.08)	0.076 (0.24)	−0.279 (−2.93)	−0.628 (−2.59)	−2.368 (−1.75)/0.205 (0.95)
Cheer	1.495 (10.58)	0.576 (0.83)	−0.021 (−0.24)	0.024 (0.11)	0.209 (0.15)/1.351 (3.39)
Era	1.528 (8.30)	1.862 (8.40)	−0.034 (−0.33)	−0.001 (−0.01)	2.190 (1.81)/1.678 (8.21)
Fab	0.044 (0.06)	0.604 (0.75)	−0.528 (−3.51)	−0.244 (−1.25)	0.356 (0.27)/0.505 (0.60)
Gain	0.137 (0.41)	0.568 (1.54)	−0.567 (−4.26)	−0.266 (−1.42)	0.281 (0.21)/0.464 (1.28)
Purex	0 (—)	0 (—)	−0.413 (−4.47)	−0.776 (−3.72)	−2.883 (−2.20)/0 (—)
Store brand	0 (—)	0 (—)	−0.652 (−6.07)	−0.869 (−4.07)	−6.627 (−3.39)/0 (—)
Surf	1.956 (13.46)	2.172 (7.59)	−0.276 (−2.36)	−0.25 (−1.34)	0.781 (0.62)/2.440 (11.89)
Tide	2.092 (14.95)	3.134 (16.67)	0.352 (4.62)	0.307 (1.76)	3.537 (2.97)/2.923 (15.97)
Wisk	1.968 (13.81)	2.876 (15.54)	0.090 (0.93)	0.187 (1.20)	2.311 (1.92)/2.660 (14.51)
XTR11	0.443 (2.15)	−0.470 (−1.86)	−0.727 (−5.23)	−0.493 (−2.46)	−1.589 (−1.12)/ − 0.513 (−2.27)
Yes	0 (—)	0 (—)	−0.510 (−3.54)	−0.244 (−1.28)	0.631 (0.48)/0 (—)
Concentration (1 = Ultra, 0 = Other, −1 = Regular)	0.220 (3.64)	0.175 (2.13)	0.024 (1.34)	0.010 (0.38)	0.146 (2.15)
Formula (Base = Regular)					
Color safe bleach	0.012 (0.18)	0.012 (0.16)	0.144 (1.23)	0.188 (1.42)	0.017 (0.24)
No dye or perfume	0.616 (2.86)	0.466 (1.79)	0.221 (1.99)	0.236 (1.86)	0.412 (1.74)
Bleach alternative	0.228 (4.58)	0.179 (3.26)	−0.005 (−0.04)	−0.009 (−0.07)	0.202 (3.75)
Stain remover	−0.295 (−1.54)	−0.153 (−0.66)	0.106 (0.92)	0.103 (0.79)	−0.049 (−0.24)
Fabric softener	0.149 (1.16)	0.260 (1.85)	−0.057 (−0.47)	−0.071 (−0.53)	0.172 (1.27)
Wrinkle reducer	0 (—)	0 (—)	−0.024 (−0.20)	−0.037 (−0.28)	−0.723 (−1.01)
Allergen treatment	0 (—)	0 (—)	0.044 (0.39)	0.019 (0.16)	−0.274 (−0.45)
Chlorine neutralizer	0 (—)	0 (—)	−0.091 (−0.78)	−0.125 (−0.98)	−1.199 (−1.78)
Biodegradable (1 = Yes, 0 = No)	0 (—)	0 (—)	0.297 (7.15)	0.370 (6.8)	2.471 (5.79)
Scent (Base = Unscented)					
Regular	0.052 (0.29)	−0.080 (−0.35)	−0.136 (−3.1)	−0.172 (−2.39)	−0.193 (−0.92)
Herbal	0 (—)	0 (—)	0.048 (0.52)	0.087 (0.83)	0.630 (0.91)
Mountain Spring	0 (—)	0 (—)	−0.144 (−2.43)	−0.163 (−1.81)	−0.862 (−1.52)
Fresh Rain	0 (—)	0 (—)	0.049 (0.59)	0.047 (0.46)	0.327 (0.52)
Floral	0 (—)	0 (—)	−0.347 (−3.9)	−0.380 (−3.71)	−2.858 (−3.96)
Type of dispenser (1 = Hand pump, 0 = Regular cap)	0 (—)	0 (—)	−0.195 (−3.89)	−0.236 (−4.43)	−1.498 (−3.71)
LN(size(oz.))	1.843 (14.94)	3.187 (16.82)	0.707 (8.24)	1.571 (10.7)	3.011 (16.84)
Number of loads	0 (—)	0 (—)	0.006 (3.02)	−0.004 (−1.14)	0.134 (8.60)
Regular price ($)	−0.608 (−17.83)	−1.162 (−16.63)	−0.083 (−16.12)	−0.158 (−12.65)	−1.101 (−17.20)
Discount ($)	0 (—)	0 (—)	0.135 (10.06)	0.210 (7.84)	1.380 (6.63)

Table 3 (cont'd.)

Utility Parameter Distributions	Scanner Panel (RP) (Asymptotic t-Stats) MNL ❶	RCMNL ❷	Experimental Choice (SP) (Asymptotic t-Stats) MNL ❸	RCMNL ❹	Joint (Asymptotic t-Stats) RCMNL ❺
Type of deal					
Store promotion (1): None (−1)	0 (—)	0 (—)	−0.063 (−1.18)	−0.095 (−1.40)	−0.493 (−1.04)
Coupon in store (1): None (−1)	0 (—)	0 (—)	0.145 (2.30)	0.153 (1.90)	1.043 (1.91)
Coupon preclipped (1): None (−1)	0 (—)	0 (—)	−0.060 (−0.96)	−0.059 (−0.77)	−0.360 (−0.66)
Attribute cutoffs					
Brand	0 (—)	0 (—)	−1.430 (−25.9)	−1.551 (−15.26)	−10.648 (−8.85)
Biodegradable	0 (—)	0 (—)	−1.045 (−12.29)	−1.265 (−12.11)	−8.726 (−8.20)
Size	0 (—)	0 (—)	−0.365 (−7.49)	−0.540 (−6.52)	−4.129 (−6.17)
Scent	0 (—)	0 (—)	−0.988 (−12.52)	−1.123 (−11.89)	−7.816 (−8.22)
(Regular price − discount)	0 (—)	0 (—)	−0.111 (−8.69)	−0.139 (−6.33)	−0.920 (−4.72)
In-store merchandising					
Store feature ad	0.957 (10.78)	0.806 (7.42)	0 (—)	0 (—)	0.812 (7.95)
Aisle display	2.119 (36.43)	2.223 (20.86)	0 (—)	0 (—)	2.163 (26.85)
State dependence—last purchase (SP)					
Last brand	0 (—)	0 (—)	1.390 (19.49)	2.162 (19.73)	16.030 (9.33)
Last size	0 (—)	0 (—)	0.550 (9.35)	0.610 (8.26)	4.257 (7.06)
Last scent	0 (—)	0 (—)	0.324 (4.95)	0.406 (5.14)	2.479 (4.33)
Last biodegradable	0 (—)	0 (—)	0.341 (6.29)	0.155 (1.92)	1.034 (1.88)
Last purchase	0 (—)	0 (—)	1.455 (14.47)	1.843 (14.3)	12.717 (9.46)
State Dependence—G&L 1983 (RP)					
Brand loyalty	3.194 (32.34)	2.939 (26.78)	0 (—)	0 (—)	3.074 (28.79)
Size loyalty	2.154 (26.61)	2.064 (23.23)	0 (—)	0 (—)	2.039 (21.75)
Scent loyalty	1.283 (9.56)	1.251 (8.68)	0 (—)	0 (—)	1.311 (9.29)
Form loyalty	1.704 (23.27)	1.758 (23.01)	0 (—)	0 (—)	1.750 (23.07)
Variances					
σ^2(Ajax)	0 (—)	0 (—)	0 (—)	1.08 (1.57)	38.001 (1.46)/0 (—)
σ^2(All)	0 (—)	0.808 (2.00)	0 (—)	1.941 (4.15)	100.54 (3.02)/0.685 (2.62)
σ^2(Arm & Hammer)	0 (—)	0.958 (1.44)	0 (—)	2.427 (3.52)	94.507 (3.09)/0.443 (1.68)
σ^2(Cheer)	0 (—)	3.388 (2.13)	0 (—)	3.696 (5.19)	201.976 (3.75)/1.218 (1.80)
σ^2(Era)	0 (—)	0 (—)	0 (—)	0.888 (2.36)	31.775 (2.35)/0 (—)
σ^2(Fab)	0 (—)	0 (—)	0 (—)	0 (—)	0 (—)/0 (—)
σ^2(Gain)	0 (—)	0 (—)	0 (—)	0 (—)	0 (—)/0 (—)
σ^2(Purex)	0 (—)	0 (—)	0 (—)	3.093 (4.48)	143.906 (3.87)/0 (—)
σ^2(Store brand)	0 (—)	0 (—)	0 (—)	1.097 (2.58)	105.691 (2.64)/0 (—)
σ^2(Surf)	0 (—)	1.129 (2.38)	0 (—)	0.668 (2.39)	19.484 (1.87)/0.159 (1.62)
σ^2(Tide)	0 (—)	0 (—)	0 (—)	6.61 (7.03)	281.591 (4.61)/0 (—)
σ^2(Wisk)	0 (—)	0 (—)	0 (—)	0.790 (3.74)	91.811 (3.8)/0 (—)
σ^2(XTR11)	0 (—)	0 (—)	0 (—)	0 (—)	0 (—)/0 (—)
σ^2(LN(size (oz.)))	0 (—)	0 (—)	0 (—)	1.480 (6.31)	31.179 (3.43)/0 (—)
σ^2(Regular price ($))	0 (—)	0.176 (6.22)	0 (—)	0.017 (5.95)	0.540 (4.52)/0.164 (6.2)
σ^2(Aisle display)	0 (—)	1.849 (5.37)	0 (—)	0 (—)	0 (—)/0.757 (4.99)
Scale Factor					
ln(SP scale)			0 (—)	0 (—)	−1.943 (−20.37)
ln(RP scale)	0 (—)	0 (—)			0 (—)

Table 3 **(cont'd.)**

Utility Parameter Distributions	Scanner Panel (RP) (Asymptotic t-Stats) MNL ❶	RCMNL ❷	Experimental Choice (SP) (Asymptotic t-Stats) MNL ❸	RCMNL ❹	Joint (Asymptotic t-Stats) RCMNL ❺
Goodness-of-Fit					
Log likelihood	−5,812.25	−5,626.22	−5,137.76	−4,616.48	−10,335.1
Number of parameters	26	32	47	59	83
$\bar{\rho}^2$ (Akaike)	0.4825	0.4984	0.3320	0.3976	0.4529
% right(estimation)	24.9	23.1	30.7	28.4	26.8
Pearson's chi-squared ratio	1.53	1.14	1.06	0.99	1.23
Number of alternatives	84	84	13	13	97
Number of HH's or respondents	400	400	384	384	784

Notes. (1) BSCs and variances of parameter distributions in Joint Models are given in SP/RP order. (2) One hundred Halton quasi-random numbers used for simulation estimation.

asymptotically converge to unity for a well-specified model and sufficiently large sample size; the ratio for Model 5 is much closer to unity than that of Model 2. This ratio is akin to a mean squared normalized residual error of prediction (see formula in table notes), so practically speaking, Model 5 has much smaller error of prediction than Model 2. This is also reflected in the final statistic, percent right predictions, which is higher for Model 5 than for Model 2.

In Table 4, Part b, we assess the prediction performance of alternative specifications of the models. Specifically, for the RP utility, we replace the Guadagni-Little specification of the state dependence variables with the last purchase specification used in the SP models. The advantage of this specification is that the RP and SP models will then have similar specifications for handling state dependence, but the disadvantage is that the RP (and, therefore, joint) models will not predict as well because the Guadagni-Little specification of the state dependence variables can reflect higher-order state dependence, whereas the last purchase variables can reflect only a first-order effect. We estimated two joint models with the last purchase effects, one in which the RP and SP last purchase effects are free, and another in which the two are constrained to be equal.

We see in Table 4, Part b, that the validation sample log likelihoods for the joint models are again higher than that of the pure scanner panel model, and by larger margins than in the original specifications in Part a. The joint model has a better log likelihood regardless of whether the state dependence effects are constrained to be equal in the RP and SP models. Likewise, ρ^2 is significantly higher for the joint models compared to the pure RP model, and the χ_p^2 values are superior, as well. However, the Percent Right Predictions is better for the joint model only when the state dependence effects are constrained to be equal across data sources, and even then, by only a small margin.

These findings demonstrate that the superior prediction accuracy of the joint RP/SP model is robust to different specifications of state dependence, though we prefer the usage of the Guadagni-Little loyalty variables in the RP utility specification (Table 4, Part a, Model 5) due to their superior explanatory power. In addition, there is evidence that these loyalty variables explain state dependence without confounding state dependence and preference heterogeneity (Abramson et al. 2000).

Discussion of Results

Model 5 implements the data enrichment hypothesis by restricting common parameters to be equal across data sources and estimating the scale factor of the SP data source relative to a normalized unit scale for the RP data source. Note that our estimate of the scale factor in Table 3 is actually the logarithm

Table 4 Prediction Results on Holdout Scanner Panel Data

(a) Models Estimated in Table 3

Model	Log Likelihood	ρ^2 *	χ_p^2: Pearson's Chi-Squared Ratio**	% Right Predictions
Scanner panel				
RCMNL (❷)	−1,749.50	0.4554	1.64	18.1
SP				
RCMNL (❹)	−2,684.41	0.1643	1.14	1.4
Joint				
RCMNL (❺)	−1,680.98	0.4767	1.25	21.6

(b) Alternative Specification of State Dependence—Last Purchase Dummies Rather Than Guadagni and Little 1983 Loyalty Variables Used in RP Utility Specifications

Model	Log Likelihood	ρ^2 *	χ_p^2: Pearson's Chi-Squared Ratio**	% Right Predictions
Scanner panel				
RCMNL	−2,124.47	0.339	7.28	14.6
Joint				
RCMNL (state dependence coefficients different by source)	−1,860.01	0.421	0.82	14.6
Joint				
RCMNL (state dependence coefficients equal across sources)	−1,855.01	0.423	0.80	14.9

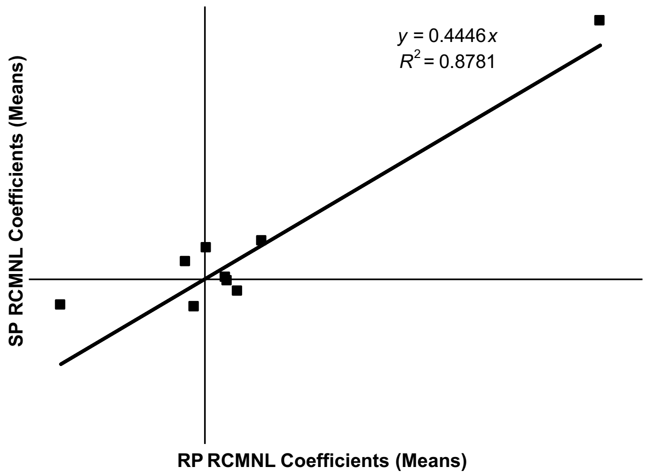
*$\rho^2 = 1 - \text{LL(model)}/\text{LL(random)}$, where LL(model) is the log likelihood of the holdout sample evaluated with the coefficients of the model in question, and LL(random) is the log likelihood assuming equal probability (random choice) for all alternatives in a choice set.

**$\chi_p^2 = (1/DF)\sum_{h,t,k}(\delta_{htk} - \hat{P}_{htk})^2/(\hat{P}_{htk}(1-\hat{P}_{htk}))$, where DF is equal to number of choices for prediction to holdout, $\delta_{htk} = 1$ if household h chooses UPC k on purchase occasion t, 0 o.w., $\hat{P}_{htk}$ is the corresponding estimated choice probability using the appropriate model.

of the scale factor (to preserve nonnegativity of the estimate); hence, the estimated scale for the SP data set, relative to the unit normalization for the scanner panel data, is $\exp(-1.943) \approx 0.143$.

Though we have already demonstrated in Table 4 that the data enrichment exercise results in improved prediction accuracy, one can also test the hypothesis of scalability using a likelihood ratio test, as shown by Swait and Louviere (1993). In our case, Model 5 is a restricted model to be compared to the combined independent use of Models 2 and 4, hence, the chi-squared statistic is $-2[-10{,}335.1 - (-5{,}626.22 - 4{,}616.48)] = 184.8$ with eight degrees of freedom; the critical value is just 15.5 at the 95% confidence level.

Figure 2 Scanner Panel vs. SP Coefficients



Note. SP coefficients are concentration, color safe bleach, no dye or perfume, bleach alternative, stain remover, fabric softener, regular scent, size, and regular price.

While this rigorous test rejects the hypothesis of parameter equality up to a scale difference between these two data sources, it is useful to consider just how surprisingly close to scaling this result actually is. Figure 2 plots the nine coefficients from the RP and SP models that are restricted to be equal up to a scale factor in the joint model (concentration, color safe bleach, no dye or perfume, bleach alternative, stain remover, fabric softener, regular scent, size, regular price). The predicted linear relationship between parameter vectors is quite strong between this scanner panel data set and the choice experiment. Louviere et al. (2000, Chapter 13), present a sizeable body of evidence concerning inter-data-source scalability. They show that, at an empirical level, this simple hypothesis seems to systematically account for a very large percentage (generally, more than 70%) of the variation of estimated partworths originating from multiple data sources. In our data, some 88% of the variability in partworths common to both data sources is accounted for by the scalability hypothesis.

These findings make an important contribution to the literature on data enrichment, namely, that pooling multiple data sources can result in improved prediction accuracy, even when the scalability hypothesis is formally rejected via a likelihood ratio test. Thus, the formal test may be too rigorous, especially when improved prediction accuracy is a major priority. Future research should investigate the effects of data characteristics (e.g., relative sample sizes of sources, total sample size, number of common and unique parameters, correlation of common parameter vectors across sources, etc.) on (1) the probability of rejecting the scalability hypothesis and (2) validation sample prediction accuracy. We discuss other avenues for future research and conclude in the next section.

Conclusion and Future Research

This research has applied the data enrichment concept and technique to scanner panel data, augmenting liquid laundry detergent panel data with corresponding data from a choice experiment (Louviere and Woodworth 1983, Louviere et al. 2000). In this particular exercise, supplementing the scanner panel with a choice experiment (or SP data) is able to bring about a number of benefits: (1) New attributes and attribute levels are introduced via the choice experiment; (2) brands not represented in the scanner panel data are included in the joint model estimation; (3) a richer description of household heterogeneity is included via the SP data, specifically through the inclusion of respondent-provided attribute and brand cutoffs (Swait 2001); (4) different types of deals are introduced, including in-store and preclipped coupons; and last, but certainly not least, (5) the SP design matrix is used to improve conditioning of the joint model, yielding more robust partworth estimates. The scanner panel data, of course, contributes a number of strengths that are difficult, at best, to achieve with SP data: (1) It captures the dynamics of preferences, including higher-order state dependence; (2) in-store aisle display and feature ad effects are present, depicting a store environment that would be difficult to replicate in a choice experiment; and (3) its choices represent actual market transactions, implying that actual money was traded for goods, which is not the case in the SP data.

Empirically, we find that the joint model calibrated using pooled data predicts (significantly) better to a holdout scanner panel data set than the corresponding scanner panel model. We believe that this improved prediction accuracy results from improved parameter estimates due, at least in part, to the advantages of the data enrichment concept described in the previous paragraph. As mentioned, future research should investigate the effects of specific data characteristics on both validation sample prediction accuracy and the probability of rejecting the scalability hypothesis.

We note that our research maintained the MNL model (in a random coefficients variation) as the working specification. This is not particularly limiting, since McFadden and Train (2000) have proven that the Mixed MNL model is capable of simulating non-IIA behaviors through the random parameter distributions. We alert researchers employing other choice model specifications, both within and without the GEV family of models, that the hidden impact of source-specific scale factors is still there: This impact is common to all discrete outcome latent variable models (see Louviere et al. 2000, Chapter 13). Another cautionary note is that random coefficients models assume that individuals have the same random component variances (Swait and Louviere 1993). If individuals have different random component variances, estimates of means and variances of utility parameters are confounded with distributions of random component variances, and it is not possible to interpret the results as estimates of the utilities. (We thank an anonymous reviewer for making this point.) Further, if individuals have their own utility scales and units of measurement for each attribute, then sorting out these effects (which are aggregated in such a complex econometric model) would be difficult at best. Identifying ways to separate these effects is a topic for future research.

Given the continuing importance of the scanner panel data source, future research should investigate more thoroughly the factors that might contribute to the successful enrichment of multiple scanner panel data sets (e.g., across product categories—crackers and yogurt, or within category—just crackers in a before-and-after comparison study, or crackers in separate markets), as well as the enrichment of scanner panel with experimental choice and other conjoint data. Such factors as task layout, task complexity (e.g., number of alternatives, attributes), number of

SP choice replications, survey length, and so forth, are likely to influence the success of a data enrichment exercise (see Swait and Adamowicz 2001a, b). Prescriptions on good SP task design practice will be most helpful to scanner panel practitioners in implementing data enrichment.

Other preference elicitation methods besides choice experiments exist (Green and Srinivasan 1978, 1990), and are also amenable to the type of data fusion done here. The details of which models and how common taste parameters are equated across data sources will certainly be different, but conceptually the same end goal can be attempted. Research into the use of other preference elicitation mechanisms can give researchers and practitioners more flexibility in implementing data enrichment.

References

Abramson, Charles, Rick L. Andrews, Imran S. Currim, Morgan Jones. 2000. Parameter bias from unobserved effects in the multinomial logit model of consumer choice. *J. Marketing Res.* **37**(November) 410–426.

Ainslie, Andrew, Peter E. Rossi. 1998. Similarities in choice behavior across product categories. *Marketing Sci.* **17**(2) 91–106.

Allenby, Greg M. 1990. Hypothesis testing with scanner data: The advantage of Bayesian methods. *J. Marketing Res.* **27**(November) 379–389.

Andrews, Rick L., Imran S. Currim. 2002. Identifying segments with identical choice behaviors across product categories: An intercategory logit mixture model. *Internat. J. Res. Marketing* **19**(March) 65–79.

Ben-Akiva, Moshe, Steve Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. MIT Press, Cambridge, MA.

Blamey, R. K., J. W. Bennett, M. D. Morrison, J. J. Louviere. 2001. A choice experiment of green product choice. J. W. Bennett, R. K. Blamey, eds. *Environmental Choice Modelling*. Edward Elgar, Northampton, MA, 115–132.

——, ——, J. J. Louviere, M. D. Morrison. 2001. Yea-saying and validation of a choice model of green product choice. J. W. Bennett, R. K. Blamey, eds. *Environmental Choice Modelling*. Edward Elgar, Northampton, MA, 178–201.

Brownstone, David, David Bunch, Kenneth Train. 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Res. Part B* **34** 315–338.

Bucklin, Randolph E., Sunil Gupta. 1992. Brand choice, purchase incidence, and segmentation: An integrated approach. *J. Marketing Res.* **29**(May) 201–215.

——, ——. 1999. Commercial use of UPC scanner data: Industry and academic perspectives. *Marketing Sci.* **18**(3) 247–273.

——, ——, S. Siddarth. 1998. Determining segmentation in sales response across consumer purchase behaviors. *J. Marketing Res.* **34**(May) 189–197.

Burke, Raymond R., Bari A. Harlam, Barbara E. Kahn, Leonard M. Lodish. 1992. Comparing dynamic consumer choice in real and computer-simulated environments. *J. Consumer Res.* **19**(June) 71–82.

Deighton, J., C. Henderson, Scott Neslin. 1994. The effects of advertising on brand switching and repeat purchasing. *J. Marketing Res.* **31**(February) 28–43.

Erdem, Tülin, Michael P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15**(1) 1–20.

——, ——, Baohong Sun. 1999. Missing price and coupon availability in scanner panels: Correcting for the self-selection bias in choice model parameters. *J. Econometrics* **89**(1) 177–196.

Fader, Peter, Bruce Hardie. 1996. Modeling consumer choice among UPCs. *J. Marketing Res.* **33**(November) 442–452.

Gönül, Füsun, Kannan Srinivasan. 1993. Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues. *Marketing Sci.* **12**(Summer) 213–229.

Green, Paul, V. Srinivasan. 1978. Conjoint analysis in consumer research: Issues and outlook. *J. Consumer Res.* **1** 61–68.

——, ——. 1990. Conjoint analysis in marketing research: New developments and directions. *J. Marketing* **54**(4) 3–19.

Guadagni, P., John Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2**(3) 203–238.

Gupta, Sachin, Pradeep K. Chintagunta. 1994. On using demographic variables to determine segment membership in logit mixture models. *J. Marketing Res.* **31**(February) 128–136.

——, ——, Anil Kaul, Dick R. Wittink. 1996. Do household scanner data provide representative inferences from brand choices: A comparison with store data. *J. Marketing Res.* **33**(November) 383–398.

Gupta, Sunil. 1988. Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* **25**(November) 342–355.

Hensher, David, Jordan Louviere, Joffre Swait. 1999. Combining sources of preference data. *J. Econometrics* **89**(1–2) 197–221.

Keane, Michael P. 1997. Current issues in discrete choice modeling. *Marketing Lett.* **8**(3) 307–322.

Kim, Byong-Do, Peter E. Rossi. 1994. Purchase frequency, sample selection, and price sensitivity: The heavy user bias. *Marketing Lett.* **5**(1) 57–67.

Lazari, Andreas G., Donald A. Anderson. 1994. Designs of discrete choice set experiments for estimating both attribute and availability cross effects. *J. Marketing Res.* **31**(August) 375–383.

Louviere, Jordan, David Hensher, Joffre Swait. 2000. *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*. Cambridge University Press, Cambridge, U.K.

——, George Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *J. Marketing Res.* **20**(November) 350–367.

McFadden, Daniel, Kenneth Train. 2000. Mixed MNL models for discrete response. *J. Appl. Econometrics* **15**(5) 447–470.

Morikawa, Takayuki. 1989. *Incorporating Stated Preference Data in Travel Demand Analysis*. Ph.D. dissertation. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Morwitz, Vicki, David Schmittlein. 1992. Using segmentation to improve sales forecasts based on purchase intent: Which intenders actually buy? *J. Marketing Res.* **29**(November) 391–405.

——, Joel Steckel, Alok Gupta. 1999. When do purchase intentions predict sales? Working paper. Stern School of Business, New York University, New York.

Ohler, Tobias, Aihong Le, Jordan Louviere, Joffre Swait. 2000. Attribute range effects in binary response tasks. *Marketing Lett.* **11**(3) 249–260.

Rossi, Peter E., Greg M. Allenby. 1993. A Bayesian approach to estimating household parameters. *J. Marketing Res.* **30**(May) 171–182.

Sándor, Zsolt, Michel Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Sci.* **21**(4) 455–475.

Swait, Joffre. 2001. A non-compensatory choice model incorporating attribute cutoffs. *Transportation Res. Part B* **35** 903–928.

——, Wiktor Adamowicz. 2001a. Choice complexity and decision strategy selection. *J. Consumer Res.* **28**(June) 135–148.

——, ——. 2001b. Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organ. Behavior Human Decision Processes* **86**(2) 141–167.

——, Adriana Bernardino. 2000. Distinguishing taste variation from error structure in discrete choice data. *Transportation Res. Part B* **34**(1) 1–15.

——, Jordan Louviere. 1993. The role of the scale parameter in the estimation and use of multinomial logit models. *J. Marketing Res.* **30**(August) 305–314.

——, ——, Michael Williams. 1994. A sequential approach to exploiting the combined strengths of SP and RP data: Application to freight shipper choice. *Transportation* **21**(2) 135–152.

Toubia, Olivier, Duncan I. Simester, John R. Hauser, Ely Dahan. 2003. Fast polyhedral adaptive conjoint estimation. *Marketing Sci.* **22**(3) 273–303.

Winer, Russell S. 1999. Experimentation in the 21st century: The importance of external validity. *J. Acad. Marketing Sci.* **27**(3) 349–358.

This paper was received September 9, 2002, and was with the authors 2 months for 2 revisions.