Current Issues in Discrete Choice Modeling

Keane, Michael

1997

Marketing Letters 8:3 (1997): 307–322

Current Issues in Discrete Choice Modeling

MICHAEL P. KEANE
1035 Management and Economics, University of Minnesota, 271 19th Avenue South, Minneapolis, MN 55455

Abstract

Until recently, computational constraints forced researchers in the discrete choice area to limit themselves to very simple statistical models, such as the multinomial logit (MNL), in which choice probabilities could be evaluated quickly on a computer. But the MNL only makes sense as a behavioral model under very special circumstances. Recent advances in computation make it possible to estimate richer behavioral models that generate very complex choice probability expressions. This paper discusses a number of possible avenues for future research in the discrete choice area in light of these developments.

Key words: discrete choice, brand choice, choice dynamics, stated preference, contingent valuation, purchase quantity, new products

I. Introduction

Discrete choice models have been widely used in marketing, transportation, economics and many other areas to study both revealed and stated preference data. The most commonly used discrete choice model has of course been the multinomial logit (see McFadden (1974)). The appeal of the multinomial logit (MNL) arose from the fact that it is simple to estimate. Choice probabilities and their derivatives take a simple closed form, so that the likelihood for the MNL can be quickly constructed and easily maximized using modern digital computers.

The problem with MNL is that it makes very strong assumptions about consumer behavior. The assumption that has received the most attention is the independence of irrelevant alternatives (IIA) property, which says that if a new alternative is added to a choice set then choice probabilities for all existing alternatives fall proportionately. Despite the fact that this and other behavioral assumptions underlying MNL are untenable in many contexts, researchers were still willing to use MNL because attempts to relax its behavioral assumptions would typically lead to models that were computationally intractable. Specifically, such attempts would typically lead to choice models in which the choice probabilities took the form of high dimensional integrals with no closed form, as occurs, for example, in the multinomial probit model (MNP).

Today we live in a rather different world. The speed of digital computers has increased dramatically since 1974. This alone would make it feasible to estimate discrete choice models in which choice probabilities have no closed form, provided the integrals that must be evaluated numerically are below a certain dimension (say 4). But in addition, there have been important advances in estimation technology. The most important of these is the development of highly accurate smooth probability simulators (see McFadden (1989), Keane (1993, 1994), Borsch-Supan and Hajivassiliou (1993)). These allow (approximate) likelihood functions for discrete choice models to be constructed far more quickly than would be possible using numerical integration.

In my view, most researchers in marketing, transportation, economics and other areas where discrete choice models are widely used have not yet grasped the full import of these events. Most interpret them to mean that it is now possible to relax IIA and estimate MNP or other generalizations of MNL. The real meaning is far more profound. Given current technology, it is no longer necessary to adopt what might be called a "statistical" approach to empirical work in the discrete choice area—that is, the common practice of starting with a particular *statistical* model that one knows is tractable (like MNL or now MNP), and then adopting (ex-post) whatever explicit or implicit behavioral assumptions are needed to fit the research question at hand into that framework.

Given current technology, it is often feasible to adopt a "theory-based" approach to discrete choice modeling. Given a research question, one starts by writing down an appealing *theoretical* model of the choice behavior of interest. Then, given assumptions about the distributions of the stochastic terms in that model, one obtains expressions for choice probabilities—and hence a *statistical* model that may be estimated. Since one's starting point was to write down a theoretical model that captures the salient features of the particular choice behavior under study, there is no reason to expect that the statistical model so obtained will resemble any textbook model like MNL or MNP. But in many instances, it will still be possible to use modern simulation technology to evaluate the required choice probability expressions and form the appropriate likelihood.

A good example is work on welfare program participation by Keane and Moffitt (1996). We start with a utility based optimizing model of the behavior of single women with children confronted by the complex array of constraints presented by the set of programs that make up the US welfare system. The expressions for program participation probabilities derived from this theoretical model are quite complex—in fact, the limits of integration in the expressions cannot be expressed analytically. Nevertheless, it is straightforward to use simulation techniques to evaluate the choice probabilities and form the likelihood.

In my view, a major obstacle to progress in building our understanding of choice behavior is that many researchers continue to adopt a statistical modeling approach. And many of the research papers that are being published simply apply new statistical methods to discrete choice data—rather than attempting to develop and estimate new and interesting theoretical models of behavior. More careful attention must be paid to behavioral theory if we hope to make further progress. In this paper I discuss a number of issues in the discrete choice modeling area that are currently attracting attention, focussing on theory-based approaches to modeling. The topics are: joint modeling of brand choice and the timing and volume of purchases, modeling demand for new products, error structures in choice models, and contingent valuation.

II. Joint modeling of brand choice, purchase timing and purchase quantity: static case

A number of authors have estimated statistical models for both the purchase timing and brand choice processes. For example, Zufryden (1977) specifies an Erlang distribution of interpurchase times in a category and a beta distribution of conditional brand choice probabilities, while Jeuland et al (1980) alter the statistical model by adopting a Dirichlet distribution for purchase probabilities. Shoemaker et al (1977) introduce correlation between interpurchase times and brand choice probabilities. Models with a no-purchase option are also purchase timing models (e.g., Bucklin and Lattin (1991) specify a MNL for brand choice, nested in a logit for purchase in the category).

Other authors jointly estimate statistical models of brand choice and purchase quantity (e.g., Krishnamurthi and Raj (1988) estimate switching regressions for purchase quantity where a MNL for brand choice determines if the left hand side is observed). Also, Guadagni and Little (1983) and Keane (1996) estimate MNL and MNP models (respectively) with brand/size combinations as alternatives, thus incorporating an aspect of quantity choice, but ignoring purchase of multiple units.

Recently, some authors have considered all three of the timing, brand and quantity decisions. Gupta (1988) specifies a separate statistical model for each. But Chiang (1991) and Chintagunta (1993) set all three within a single utility maximization problem. They appeal to the Hanemann (1984) result that when brands are perfect substitutes and quantity is infinitely divisible, there is a separation of brand choice and quantity decisions. The timing/brand choice part reduces to a multinomial choice model with a no-purchase option, and a quantity regression can be estimated in a separate step. Achieving this separation is important because, with multiple brands, the number of feasible brand/quantity combinations quickly becomes astronomical, and it is not computationally feasible to search over all combinations to find the optimal choice.

The Chiang (1991) and Chintagunta (1993) papers are laudable in that they attempt to apply a "theory-based" approach by starting from an optimizing model of consumer behavior. However, they err in applying the Hanemann (1984) results in cases where they do not apply—the former to coffee that comes in 16 oz units and the latter to yogurt which comes in 6 and 8 oz units. Even if all brands are perfect substitutes in utility, if quantity is discrete one does not achieve a separation between timing/brand choice and quantity decisions.

To illustrate, suppose utility is linear in the quantity of two alternative brands in one category, and logarithmic in a composite other commodity. Substituting a static budget constraint into the utility function one obtains the expression $U = \alpha b_1 + \beta b_2 + \ln(Y - p_1 b_1 - p_2 b_2)$, where b_j is quantity of brand j, p_j is its price, Y is income, and α and β are parameters capturing preferences for the two brands. If quantity is infinitely divisible, the consumer will make a purchase in the category if and only if $p_1 < \alpha Y$ or $p_2 < \beta Y$. Conditional on purchase, the consumer buys only brand 1 if $\alpha/p_1 > \beta/p_2$, and only brand two otherwise (i.e., he/she buys only the brand that provides the most utils per dollar). For example, suppose that $\alpha = 0.2$, $\beta = 0.095$, $p_1 = 2$, $p_2 = 1$ and $Y = 20$. Then the consumer will choose to buy only brand one, and he/she will set $b_1 = 5$. But now suppose that the brands can only be bought in units of 2 (i.e., 2, 4, 6, etc.). Then the optimal choice is to buy *both* brands, and to set $b_1 = 4$ and $b_2 = 2$. The timing/brand choice and purchase quantity decisions do not separate.

Thus, the computational problem of dealing with discrete quantity choice problems remains open. At present, there is no alternative to exogenously imposing constraints that keep the size of the choice set tractable, such as that agents cannot chose more than one brand on a purchase occasion. Research on behaviorally appealing approaches to this problem is needed.

Another important question is whether the decision to buy in a category and the brand choice/quantity decision are best thought of as flowing from the same utility maximization problem. For example, if store visits are costly, the decision to buy in a category may be made prior to seeing prices, in-store promotions, etc., followed by a separate in-store brand/quantity choice problem. Note that brand choice models that use shopping occasions as the "time" line (by including a no-purchase option) implicitly assume that the category purchase decision, but not the store visit decision, is influenced by prices or promotions of brands in the category. In contrast, models that use purchase occasions as the "time" line (excluding the no-purchase option), implicitly assume that the store visit and category purchase decisions are not influenced by the prices or promotions of brands in the category. It would be interesting to develop and estimate theoretical models of consumer behavior that are explicit about the stage of the choice process at which prices and promotions are observed.

III. Purchase timing, brand choice and quantity decisions of forward-looking consumers

If consumers are forward-looking, and there exist intertemporal linkages in preferences or constraints, then current demand will depend not just on current prices but also on expected future prices. That is, consumers solve a dynamic rather than a static optimization problem.

Intertemporal linkages may arise even for a perishable good like walleye pike, which must be eaten soon after being caught. Suppose I go to the store and see that the price of walleye is higher today than normal, so I expect it to drop next week. If my preferences for walleye are additively separable over time, and if I cannot save, then I still face a static choice problem—only the current price matters to my decision, not the price I expect next week. But, if by passing up walleye this week I can save money for next week or raise my marginal utility of walleye consumption next week, the fact that I expect next week's price to be lower may give me an incentive to defer the purchase.

For storable consumer goods and durable goods intertemporal linkages are obviously important. In the former case an obvious source of intertemporal linkage is inventories. In the latter case it is that the good, once purchased, provides a flow of consumption over a long period of time. In either case, consumers will have an incentive to time purchases so they buy when the price is "low" relative to expected future prices. This is easy to say, and introspection suggests that we all engage in such behavior. But forward-looking behavior is very difficult to formalize in a model. The whole area of modeling expectations is extremely difficult, both conceptually and operationally.

Economists often assume that agents form expectations "rationally." Here, the technical meaning of rational is to form expectations in a manner that is consistent with the "true" model. When applied to a storable consumer good like detergent, this means consumers know the process that generates detergent price fluctuations, and form expectations of future prices optimally given that process. The computational demands of pursuing the rational expectations approach in discrete choice models are formidable. In such models, consumers behave *as if* they solve complex dynamic programming (DP) problems to determine optimal choices. Hence, the researcher must also solve a DP problem to determine consumers' choice probabilities. An example is the recent paper by Erdem and Keane (1996). The statistical model that arises from our theoretical model requires many hours of supercomputer time to estimate, despite that fact that we used simulation and interpolation techniques to approximate both the DP solution and the resultant choice probabilities.

While the "full blown optimization" approach exemplified by the Erdem and Keane paper is now feasible for many problems, many researchers will want to adopt less computationally intensive alternatives, while still allowing for effects of expected future prices on current demand. Some alternatives are: 1) assume expectations are formed adaptively, so expected future prices are a simple function of past prices, 2) attempt to measure expectations directly and include them in the model, 3) assume that agents make choices using relatively simple rules-of-thumb, 4) assume that agents have short-time horizons or only re-optimize on rare occasions (thus simplifying DP problems).

The problem with such simplifications, as well as with purely statistical models that ignore entirely issues of how agents form expectations and form decision rules, is that they can run a-foul of the "Lucas critique of econometric policy evaluation," due to Marschak (1952) and Lucas (1976). To understand the Lucas critique, consider the following simple example. Suppose after a few years of shopping I realize that my favorite supermarket always cuts the price of my favorite detergent, say Era, from $4.00 to $3.50 per 32oz during one week of every month. Most other shoppers at the store realize this too. So those who like Era adopt a decision rule that says: "Only buy Era if the price is $3.50, and then buy enough so that my inventory will last at least 7 weeks (the longest possible time until the next sale)." There are some random deviations from this rule (e.g., sometimes people may have big positive usage shocks that cause them to stock out and deviate from the rule), but it is the dominant pattern of behavior. If one were to estimate a MNL model for detergent choice in the store, the price elasticity of demand for Era would appear to be enormous.

Now, suppose the store changes its policy, and starts to charge $4.00 for Era all the time. The MNL model would predict that sales of Era would fall drastically—in fact, almost to zero. But what would really happen? The shoppers who like Era would let 7 weeks go by and then be surprised to find that the price was still $4.00 in week 8. At this point, it seems implausible that sales of Era would really drop to almost zero. Rather, some shoppers would buy Era anyway because they prefer it sufficiently to other brands.

Some would indeed switch brands. Some would check out prices in a different store, and perhaps even switch stores. All the regular Era shoppers would realize that the pricing policy had changed and start to develop a new decision rule. Clearly the MNL model estimated on the data from the old monthly price cut regime will have little hope of predicting what will happen under the new constant price regime. This inability of statistical models to predict effects of regime shifts is the basic idea of the Lucas critique.

To successfully forecast behavior after a regime change one needs to model how agents tailor their decision rules to particular regimes. One needs a so-called "structural" model whose parameters are "primitives" of agents' preferences, information processing systems, etc., which remain *fixed* across regimes. Then, given any regime, one can solve for the optimal decision rules of agents in that regime. In other words, the goal is a model that is generalizable across regimes.

Only a "full blown optimization" approach in which agents *choice of decision rule* is modeled can achieve this goal. In contrast models with fixed rule-of-thumb decision rules would break down after the regime shift in the detergent example. One such model is a "reference price" model, in which current demand depends in a *fixed* way on deviations of current prices from average or "reference" prices (see Winer (1986)). In the detergent example, if such a model was estimated on data from the monthly price cut regime, it would predict almost zero demand for Era when price equals "reference" price. It would therefore predict a drop in Era sales to almost zero in the constant price regime.

Nevertheless, simple statistical models which incorporate expectations (e.g., reference price and other rule-of-thumb type models) may be very useful for predicting responses to "marginal" changes within fixed regimes. If a statistical model is estimated on historical data in which prices are drawn from a certain stochastic process, that model could be reliably used to forecast demand tomorrow given a particular realization from the same process. It is only if we attempt to use the model to forecast the effect of a fundamental change in the stochastic process for price that we run a-foul of the Lucas critique. Despite these concerns, some authors have recently applied control theory to derive optimal pricing or advertising rules for retailers while taking statistical representations of consumers' decision rules as invariant to pricing or advertising policy (e.g., Chintagunta and Vilcassim (1992), Kopalle and Winer (1996)). Such attempts ignore the basic message of the Lucas critique that decision rules will change when policy rules change.

It is worth stressing that problem raised by the Lucas critique is really an identification problem: If we observed data from many regimes, we could fit a purely statistical model to the data, perhaps using regime indicators as explanatory variables, and successfully use the model to predict changes in behavior across different regimes. So one response to the critique is to collect more data.

Finally, note that the rational expectations assumption is not *de rigueur* in structural modeling. Given costs of computation, it may well be optimal for agents to form expectations in simpler ways. The optimal expectation formation mechanism may depend both on the regime and more fundamental primitives that characterize the consumer's information processing costs and capabilities. Obviously, these are difficult issues that lie at the cutting-edge of research in the social and behavioral sciences.

IV. Modeling demand for new products

In thinking about demand for new products, it is useful to differentiate them along two dimensions. First, is the new product only marginally different from existing products, or does it represent a fundamental change? Second, are there "network externalities" associated with the adoption of the new product?

An example of a new product that is only marginally different from existing products is a new HMO that has a different monthly premium and covers different fractions of prescription expenses than existing HMOs. In this case, it would be possible to estimate a choice model on existing market or "revealed preference" (RP) data, which exhibits variation in the attributes of interest, and use the model to predict demand for the new product.

Data from stated preference (SP) choice experiments can also be used to estimate such models (see Louviere (1988) or Hensher (1994)). In an SP experiment, respondents are presented with a series of artificial choice sets. A set of attributes of the alternatives is defined and varied across the sets. Thus, in contrast to market data, the researcher controls the attribute variation in the data (Bunch et al (1996) provide evidence on the statistical efficiency of various experimental designs). As pointed out by Adamowicz et al (1994), there are often cases where attributes are collinear in market data, making it difficult or impossible to predict the effect of independent variation in an attribute. In such cases SP data has a key advantage over market data.

However, there are important potential problems with the use of SP data. Most obviously, respondents have no incentive to make choices in an SP experiment in the same way they would in the market. Furthermore, even if people do respond as if they applied their true utility weights to the attributes presented in the experiment, the SP choice situation is typically somewhat different from a market choice situation. Some aspects of the market choice context, such as search costs, are absent in the SP experiment. And in the experiment alternatives are defined solely by the attributes presented, while in market data there may be attributes observed (or perceived) by the consumer but unobserved by the researcher, and these will be pushed into the choice model error terms.

Despite these potential problems, a number of recent papers have obtained a rather striking result that lends credence to the use of SP experiments. In a large number of cases in which both RP and SP data have been used to fit MNL models for the same good, researchers have not rejected the hypothesis that the estimated utility weights are proportional in the RP and SP choice models. Examples are Adamowicz et al (1994), Ben-Akiva and Morikawa (1990), and Hensher and Bradley (1993). These researchers generally find that the only significant difference between the MNL models estimated on RP and SP data is the variance (or "scale") of the error term (see Swait and Louviere (1993) for a discussion of tests of this hypothesis).

A difficult question is why the process generating responses to SP choice experiments should resemble that generating market choice behavior. In the papers mentioned above, the choice tasks are all of a simple variety that is familiar to consumers (e.g., choose among several varieties of detergent based on price and other attributes). Respondents may get utility from demonstrating they can make rational price/quality tradeoffs in such situations. Or, the researchers may have successfully "framed" the choice tasks so that respondents believe it is in their interest to reveal accurately how they value attributes (because, say, that information will be used to design new products). An interesting avenue for future research is to conduct SP experiments of varying designs so as to discover conditions under which response patterns do and do not resemble those in market data.

If the scale of the error term does indeed differ between RP and SP choice models it creates a problem when applying the SP model to predict market share for a new product. To deal with this, utility weights estimated from the SP data should be scaled proportionately so as to set predicted market shares as close as possible to those in the historical market data. This scaled model can then be used to forecast market shares after the new product introduction.

Another issue is how to optimally combine RP and SP data to more efficiently estimate choice model parameters. Given the processes assumed to be generating the RP and SP data (e.g., the MNL in the papers cited above), one can form the correct likelihood for the combined data. Each choice observation, whether from the market or SP data, contributes one term to the likelihood. There is no reason to apply weights to the observations. However, if consumers' respond as if their preference weights are the same in the RP and SP data, consumer heterogeneity will generate dependence across responses that must be accounted for in forming the likelihood. There has also been work on how to incorporate SP data on stated attribute importance with market choice data in order to obtain more efficient estimates of choice model parameters (see McFadden (1986) and Harris and Keane (1996)).

Now consider the problem of predicting market share for a fundamentally new product. Examples of products that were fundamentally different when first introduced are lap-top computers and cellular phones. Prior to their introduction all computers and phones had the attribute "not mobile," and these products introduced variability on that dimension for the first time.

Such "fundamental change" may create an identification problem. Market data cannot be used to estimate the effect of varying an attribute that is invariant in the market, unless we have a model that specifies how that attribute maps into some more fundamental determinant of utility. In a structural or utility based approach this is often the case. For example, the welfare program participation model of Keane and Moffitt (1996) can be used to predict how changes in aspects of program rules that are invariant in the data would affect participation decisions, provided the rule changes affect agents' income conditional on participation. But if the model only specifies that certain attributes affect utility, without specifying any more fundamental mechanism through which they do so, then those attributes must vary in the data in order to identify their effect on choice.

Because of the identification problem in market data, SP experiments may often be the only means to predict demand for a fundamentally different product. But fundamental changes also create problems for SP experiments, because of the difficultly of conveying to participants the meaning of a change along a dimension that is invariant in their experience. One approach is to use computer simulation or virtual reality techniques that help consumers understand the innovation (see Weinberg (1992) and Urban et al (1996)).

There is considerable scope for experimentation in this area. Another problem is that the SP model will again need to be scaled to historical market choice data. And we cannot test whether the utility weight for the fundamentally changed attribute "scales."

Finally I turn to the issue of forecasting demand for a new product when there are network externalities. A network externality arises when the utility I get from adopting a product depends on how many other people adopt it. An example is the decision to adopt a new word processing package from a startup software company (as opposed to an update from a major vendor). If I am alone in adopting the new package, I face several problems. For example, I will not be able to send documents electronically to others to print, since they will lack the necessary software. And I will not be able to obtain support for the package locally, since I will be the only person using it.

Thus, only certain segments of consumers will adopt the new package initially: those who do not need to send documents electronically to others, those who are good at deciphering software manuals, those with great confidence that many other people will soon adopt the new package, etc. Later, after some consumers have already adopted, a wider group of consumers may be willing to adopt, etc. This suggests using a sequence of SP choice experiments to predict demand for the new product at different stages of its life cycle. The state of adoption could be part of the framing of the choice task. That is: 1) no one has yet adopted the product, no support is available, etc., 2) a certain fraction of the population has adopted and a certain level of support is available, etc. At some point one finds the level of prior adoption such that the same fraction of consumers would choose to adopt, and this is the steady state adoption level.

Given this series of conditional choice models, one must translate the predicted growth in market shares across models into real time. This is a difficult problem. One possibility is to find a previously introduced product that is in some sense similar (including similar initial and steady state adoption levels), and assume the diffusion time path will be the same. Also, one could introduce questions about timing of adoption into the SP choice task, so as to obtain the information needed to predict, given an initial market share, how much time would pass before a certain fraction of additional people adopt. In the Urban et al (1996) study on the electric car, "managerial judgement" was used to predict the number of recharging stations available at different points in time. But this seems to skirt the issue, since the state of infrastructure by any point in time will depend on cumulative adoption decisions up to that point, which is what the choice model is meant to predict.

In principle, market data could also be used to forecast demand for a new product that entails a network externality, provided the product is only marginally different from other such products introduced in the past (e.g. an improved word processing package that is not fully compatible with older versions). But, while historical data on consumers' past adoption decisions for similar products could be collected, such data appears to be rare at the individual level. Further, the type of panel data needed to model the dynamics of adoption as described above appears to be nonexistent. Thus, at present SP choice experiments appear to be the only practical option in this context.

Finally, note that demand for new products may depend crucially on expectations. Prices of many new products drop systematically with time, so expected future prices will affect current demand. And in cases of network externalities, expected future infrastructure will matter too.

V. Error structures in choice models

The "error term" in discrete choice models has often been treated as noise or intrinsic randomness in behavior arising from unknown and unspecified sources. This is misguided. It is important to interpret the error term in the light of behavioral models. To illustrate the kind of misunderstanding that can arise from failure to do so, consider a simple random utility model, in which there are heterogeneous preferences for observed and unobserved brand attributes:

$$U_{ijt} = \alpha_{ij} + P_{ijt}\gamma_i + X_{ijt}\beta_i + \epsilon_{ijt} \tag{1}$$

Here, U_{ijt} is the utility that i receives given choice of alternative j on occasion t. In market data, t could index time, store visits, or purchase occasions. In an SP experiment, t would index choice tasks. P_{ijt} denotes price, and X_{ijt} denotes an observed attribute of j (which for complete generality I allow to vary over people and choice occasions). The α_{ij} denotes the person specific intercept for alternative j, which can be interpreted as arising from i's preferences for unobserved attributes of j. The γ_i and β_i are individual specific utility parameters which are intrinsic to the consumer and hence invariant over choice occasions. The ϵ_{ijt} can be interpreted as occasion specific shocks to i's tastes, which for convenience I assume to be independent over choice occasions, alternatives and consumers.

Suppose I estimate a multinomial choice model, falsely assuming that the intercept and slope parameters are *homogeneous* in the population. The error term in this model will be

$$w_{ijt} = \hat{\alpha}_i + P_{ijt}\hat{\gamma}_i + X_{ijt}\hat{\beta}_i + \epsilon_{ijt} \tag{2}$$

where ˆ denotes the individual specific deviation from the population mean. Observe that (from the analyst's perspective) the variance of this error term for person i on choice occasion t is

$$\mathrm{Var}(w_{ijt}) = \sigma_\alpha^2 + P_{ijt}^2\,\sigma_\gamma^2 + X_{ijt}^2\,\sigma_\beta^2 + \sigma_\epsilon^2 \tag{3}$$

and the covariance between choice occasions t and $t-1$ is

$$\mathrm{Cov}(w_{ijt}, w_{ij,t-1}) = \sigma_\alpha^2 + P_{ijt}P_{ij,t-1}\,\sigma_\gamma^2 + X_{ijt}X_{ij,t-1}\,\sigma_\beta^2 \tag{4}$$

Equations (3) and (4) reveal two interesting consequences of ignoring heterogeneity in preferences. First, the error variance will differ across choice occasions as the price P and attribute X are varied. If one estimates a MNL model with a constant error variance, this will show up as variation in the intercept and slope parameters across choice occasions. In an SP experiment context, this could lead to a false conclusion that there are order effects in the process generating responses.

Second, equation (4) shows how preference heterogeneity leads to serially correlated errors. That heterogeneity is a special type of serial correlation is apparently not well understood in the marketing and transportation literatures. To obtain efficient estimates of choice model parameters one should include a specification of the heterogeneity structure in the model. But more importantly, if preference heterogeneity is present it is not merely a statistical nuisance requiring correction. Rather, one must model the heterogeneity in order to obtain accurate choice model predictions, because the presence of heterogeneity will alter cross-price elasticities, lead to IIA violations, etc. This is just one example of how paying attention to the behavioral source of the error terms in a choice model leads to new insights into how the model should be estimated, interpreted and applied.

VI. Contingent valuation

Contingent valuation (CV) is a specific SP data technique in which the researcher attempts to uncover the demand function for a particular good. This is typically done either by asking a sample of consumers how much they are willing to pay for the good (the "open-ended" approach), or by presenting consumers with various different prices, and asking if they would buy the good at those prices (the "referendum" approach). The fraction of consumers willing to buy at each price traces out the demand curve. Referendum CV is a special case of an SP choice experiment in which price is the only attribute varied, so only the complete attribute bundle the good represents can be valued.

As with SP choice experiments, a key advantage of CV over studies based on market data is that it can be used to predict demand for goods that are fundamentally different from any that are currently being traded in markets. As a result, CV has been widely used to predict demand for public goods. An example of such a good would be the Porcupine Caribou herd, whose existence provides utility (i.e., "existence value") to many people.

Recently, CV has received a great deal of attention because of its use in litigation (such as the highly publicized Exxon Valdez case) to assess existence values of public goods damaged in environmental accidents. As a result, the National Oceanic and Atmospheric Administration (NOAA) recently formed a distinguished panel of economists to evaluate the use of CV in determining existence values. Arrow et al (1993) concluded that if a number of stringent design criteria were met, then "CV studies convey useful information," although the report also suggests that few if any existing CV studies meet all these criteria. One key criterion which the panel emphasized was that "the payment scenario should be convincingly described, preferably in a referendum context, because most respondents will have had experience with referendum ballots."

The NOAA panel recommendation of the referendum format has raised controversy, with researchers taking different positions on whether that format is more incentive compatible and/or psychometrically reliable than the open-ended format. Consider first the issue of incentives to misstate preferences. Suppose a CV study deals with a wolf conservation program. In an open-ended format, respondents are asked how much they would be willing to pay in taxes to implement the program. Suppose a respondent wants to see the conservation program implemented, and believes that the higher the mean response to the CV question the more likely is implementation. However, he/she does not believe that the mean response will affect either the form of the program finally implemented or the cost in taxes (i.e., the respondent has a set belief about what the program will really cost). In that case, the respondent has an incentive to report the highest possible willingness to pay, subject to the constraint that the response look reasonable so it is not discarded as an outlier. Analogously, in a referendum format, such a respondent would have an incentive to answer "yes, I would vote for the referendum" no matter how high a tax was quoted in the question.

As Green et al (1995) note, if the above scenario is changed only in that the respondent believes the probability of program implementation depends not on the mean response to the CV question, but rather on the fraction of respondents who indicate support for the program (either by stating a willingness to pay that exceeds cost, or answering yes in the referendum format when the stated cost is greater than or equal to true cost), then the respondent has an incentive to be truthful in *either* format. For example, suppose the respondent is truly willing to pay $100 for the program. It is then optimal for him/her to express a willingness to pay of exactly $100, regardless of format, because this insures that he/she is voting "Yes" contingent on the true program cost being less than or equal to $100, and "No" otherwise. (An interesting question is how the respondent should respond in the referendum format if expressing a willingness to pay of *exactly* $100 is not an option).

Thus, Green et al (1995) argue that the referendum format should not be preferred to the open-ended format on a priori incentive compatibility grounds. However, if one wants an incentive compatible design the real issue is which format makes it easiest to frame the question in an incentive compatible way. If one wants respondents to believe that the probability of implementation depends on the fraction of respondents who vote yes, it seems easiest to do this in the referendum format.

It is worth noting, however, that the NOAA panel did not prefer referendum CV because of incentive compatibility concerns (that was a minor point in the report) but rather stressed the fact that "most respondents will have had experience with referendum ballots," making it a task with which respondents are familiar. We do not yet understand why and under what conditions choice behavior in SP choice tasks resembles revealed preference behavior in the "scaling" sense described in section IV, but at this point it appears likely that familiarity of the task is important—since agents have little or no economic incentive to work hard on the task, the cognitive burden had better be small!

Green et al (1995) also argue that the emphasis on incentive compatibility in CV surveys is misplaced, since few respondents will believe they can affect either the probability of program implementation or cost. They argue that avoidance of various psychometric biases is likely to be the more important consideration in choosing between referendum and open-ended formats. For instance, they present evidence that in referendum CV the quoted program costs act as "anchors," such that respondent's valuations move toward the anchor with which they are presented.

A potential explanation for anchoring is that it is a learning phenomenon. If we ask people what they are "willing to pay" for a government program, those who support it might respond by forming a guess of what it *should* cost, if implemented cost effectively. And costs quoted in a referendum CV survey may be used as indicators of true program cost. As Hanemann (1994) notes, a related response behavior is observed in experiments where open-ended designs have been applied to market goods, in which "people are more likely to tell you what the good costs than what it is worth to them." This is not surprising. Why should I be "willing to pay" more for a market good than the going market price, even if at a higher market price I would still buy the good?

A key issue, in all SP task design is that the attributes and prices presented be believable. In a CV task respondents may not believe the government can provide a certain environmental amenity at a certain price, and in an SP choice task they may not believe a manufacturer could provide a certain product attribute. In an SP task where the choice is among software packages, I would ignore an indication that one had an easy to read manual (since that's impossible), and in a CV task I would not believe the state government could restore the Minnesota river to pristine condition with a \$10 increase in state taxes. It is not obvious how people might respond in such contexts.

A key criticism of CV is that it is subject to "embedding" effects, whereby people exhibit implausibly low marginal willingness to pay for increments to environmental amenities (see McFadden (1994)). Such behavior may stem from scenario implausibility. For example, suppose a person says that are willing to pay \$100 to clean up the Minnesota river, but he/she indicates they are unwilling to pay much more for a program to clean up all the rivers in the state. The respondent may simply be skeptical of the government's ability to successfully implement the larger project.

CV has also been criticized on the grounds that respondents pay insufficient attention to opportunity costs. But this is also true in SP choice experiments. For example, in Urban et al (1996), the percentage of people indicating they would purchase a GM electric car dropped substantially when other alternatives were presented explicitly. With market goods it is often possible to deal with such problems by "scaling" an SP choice model so its predictions align with observed market shares for previously existing products. But as the NOAA panel noted, a fundamental problem in the case of public goods is that there may be no market data available to make such adjustments.

But it is sometimes possible to compare CV estimates of demand for existence of public goods to revealed preference data. For example, there have been actual referenda on proposals to conserve forests. And some people are willing to pay a premium to help preserve existence of natural resources by buying products like tuna caught in dolphin friendly nets or recycled paper. A major research program is needed to compare behavior in SP and CV choice tasks with behavior in market contexts. Much work along these lines has already been done (e.g., Carson et al (1996), Cameron (1992), Bishop and Heberlein (1979), Davis (1963) and the SP studies cited in section IV). But we are a long way from having general rules for adjusting SP and CV choice models so they can reliably predict behavior in market contexts—even when no market data exist to aid in calibration.

Finally, the referendum CV format posses a number of interesting statistical problems, which are discussed at length in Hanemann and Kanninen (1996). An important theme of their work is that simple logit and probit models of the type usually estimated in the CV literature rarely satisfy the restrictions of economic theory. This harkens back to my point in section I that particular statistical models should not be the starting point for analysis of choice behavior. Rather, one should start from sensible behavioral models and then work out the statistical models that emerge.

VII. Conclusion

Recent advances in computation make it possible to estimate behavioral models that generate very complex choice probability expressions. I argue that this should lead to a fundamental change in mind-set among researchers in the discrete choice area. Given a research question, one should start by developing an appealing theoretical model of the behavior of interest, and then derive the implied statistical model. Given modern technology, it will often be possible to press ahead with estimation even if the statistical model so obtained is quite complex. This contrasts with the traditional approach (often necessitated by computational limitations) of trying to find a way to fit the research question into a logit or other simple statistical framework, regardless of whether the behavioral assumptions in the framework are sensible or appealing in the context under study.

In this paper I have considered several possible avenues for future research opened up by these advances. These include: 1) joint modeling of the process by which consumers decide on store visits, purchase in a category and brand choice, so as to shed light on the role played by price and promotion activity at each stage of the process, 2) modeling consumers' choice of purchase decision rules in alternative price setting environments, 3) combining stated and revealed preference data to estimate richer models of the demand for new and existing products, 4) estimation of models which rich structures of preference heterogeneity across consumers, and 5) estimation of models of the behavioral processes underlying responses in SP and CV choice tasks. It is very exciting to be working in the discrete choice area at a time when so many possibilities are opening up.

Acknowledgments

This paper was prepared in collaboration with the participants in the workshop "Current Issues in Discrete Modeling" at the Third International Symposium on Choice Behavior held in Harrison, NY on June 14–17, 1996. Participants in the workshop were Jordan Louviere (co-chair), Wiktor Adamowicz, David Bunch, Richard Carson, Michael Hanemann, David Hensher, Bernard Pailthorpe and Joffre Swait. It should be stressed however, that the views expressed here are my own and not necessarily those of the workshop participants. Helpful comments by the editor and three anonymous referees are appreciated.

References

Adamowicz, W., Louviere, J., and M. Williams (1994), “Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities,” *Journal of Environmental Economics and Management*, 26, 271–292.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R. and H. Schuman (1993), “Report of the NOAA Panel on Contingent Valuation,” *Federal Register*, 58:10, 4601–4614.

Ben-Akiva, M. and T. Morikawa (1990), “Estimation of Switching Models from Revealed Preferences and Stated Intentions,” *Transportation Research A*, 24A:6, 485–495.

Bishop, R. and T. Heberlein (1979), “Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?,” *American Journal of Agricultural Economics*, 61:5, 926–930.

Borsch-Supan, A. and V. Hajivassiliou (1993), “Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models,” *Journal of Econometrics*, 58, 347–368.

Bucklin, R. and J. Lattin (1991), “Purchase Incidence and Brand Choice,” *Marketing Science*.

Bunch, D., Louviere, J. and D. Anderson (1996), “A Comparison of Experimental Design Strategies for Choice-Based Conjoint Analysis with Generic-Attribute Multinomial Logit Models,” manuscript, Graduate School of Management, University of California at Davis.

Cameron, T. (1992), “Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods,” *Land Economics*, 68:3, 302–317.

Carson, R., Flores, N., Martin, K., and J. Wright (1996), “Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods,” *Land Economics*, 72, 80–89.

Chiang, J. (1991), “A Simultaneous Approach to the Whether, What and How Much to Buy Questions,” *Marketing Science*, 10:4, 297–315.

Chintagunta, P. (1993), “Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households,” *Marketing Science*, 12:2, 184–208.

Chintagunta, P. and N. Vilcassim (1992), “An Empirical Investigation of Advertising Strategies in a Dynamic Duopoly,” *Management Science*, 38:9, 1230–1244.

Davis, R. (1963), “Recreation Planning as an Economic Problem,” *Natural Resources Journal*, 3, 239–249.

Erdem, T. and M. Keane (1996), “Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets,” *Marketing Science*, 15, 1–20.

Green, D., Jacowitz, K., Kahneman, D., and D. McFadden (1995), “Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods,” Working Paper, University of California–Berkeley.

Guadagni, P. and J. Little (1983), “A Logit Model of Brand Choice Calibrated on Scanner Data,” *Marketing Science*, 2, 203–238.

Gupta, S. (1988), “Impact of Sales Promotions on When, What and How Much to Buy,” *Journal of Marketing Research*, 25, 342–355.

Hanemann, W. M. (1984), “Discrete/Continuous Models of Consumer Demand,” *Econometrica*, 52, 541–561.

Hanemann, W. M. and B. Kanninen (1996), “The Statistical Analysis of Discrete Response CV Data.” In I. Bateman and K. Willis (eds.) *Valuing Environmental Preferences*: *Theory and Practice of the Contingent Valuation Method in the US*, *EC and Developing Countries*, Oxford: Oxford University Press.

Harris, K. and M. Keane (1996), “A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Stated and Revealed Preference Data,” *Journal of Econometrics*, forthcoming.

Hensher, D. (1994), “Stated Preference Analysis of Travel Choices: The State of the Practice,” *Transportation*, 21.

Hensher, D. and M. Bradley (1993), “Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models,” *Marketing Letters*, 4:2, 139–151.

Jueland, A., Bass, F. and G. Wright (1980), “A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes,” *Operations Research*, 28:2, 255–277.

Keane, M. (1993). Simulation Estimation for Panel Data Models with Limited Dependent Variables. *The Handbook of Statistics*, G. S. Maddala, C. R. Rao and H. D. Vinod editors, North Holland publisher.

Keane, M. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 62:1, 95–116.

Keane, M. (1996), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics*, forthcoming.

Keane, M. and R. Moffitt (1996), "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review*, forthcoming.

Kopalle, P. and R. Winer (1996), "A Dynamic Model of Reference Price and Expected Quality," *Marketing Letters*, 7:1, 41–52.

Krishnamurthi, L. and S. P. Raj (1988), "A Model of Brand Choice and Purchase Quantity Price Sensitivities," *Marketing Science*, 7:1, 1–17.

Louviere, J. (1988), "Analyzing Decision Making: Metric Conjoint Analysis." Beverly Hills: Sage.

Lucas, R. (1976), "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy*, 1, pp. 19–46.

Marschak, J. (1952), "Economic Measurements for Policy and Prediction." In *Studies in Econometric Method*, W. C. Hood and T. J. Koopmans (eds.), New York: Wiley, pp. 1–26.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic Press, 105–142.

McFadden, D. (1989), "A Method of Simulated Moments for the Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57:5, 995–1026.

McFadden, D. (1989), "The Choice Theory Approach to Market Research," *Marketing Science*, 5, 275–297.

McFadden, D. (1994), "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*, 76, 689–708.

Shoemaker, R., Staelin R., Kadane J., and F. Shoaf (1977), "Relation of Brand Choice and Purchase Timing Behavior," *Journal of Marketing Research*, 14, 458–468.

Swait, J. and J. Louviere (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, 30, 305–314.

Urban, G., Weinberg, B. and J. Hauser (1996), "Premarket Forecasting of Really-New Products," *Journal of Marketing*, 60, 47–60.

Weinberg, B. (1996), "An Information-Acceleration-Based Methodology for Developing Preproduction Forecasts for Durable Goods: Design, Development and Initial Validation," Ph.D. dissertation, Sloan School of Management, MIT.

Winer, R. (1986) "A Reference Price Model of Brand Choice for Frequently Purchased Products," *Journal of Consumer Research*, 13, 250–256.

Zufryden, F. (1977), "A Composite Heterogeneous Model of Brand Choice and Purchase Timing Behavior," *Management Science*, 24:2, 121–136.