MARKETING SCIENCE
Vol. 23, No. 1, Winter 2004, pp. 1–3
ISSN 0732-2399 | EISSN 1526-548X | 04 | 2301 | 0001

DOI 10.1287/mksc.1040.0060

Editorial

Endogeneity in Marketing Decision Models

Steven M. Shugan*
University of Florida, Gainesville, Florida 32611, marketingscience@cba.ufl.edu

There are many critical concerns (including the accounting for endogeneity) when one is properly estimating response functions. However, it is sometimes (certainly not always) better to leave some variables exogenous when building mathematical models intended to help decision makers. The exogenous variables allow the decision maker to better adapt the mathematical model to different situations and to incorporate myriad variables and constraints outside of the model.

Key words: endogeneity; decision models; game theory; constrained models; marketing theory

Endogeneity

When reviewing academic manuscripts, it is natural to develop a fitness checklist of requirements (Griffin 2001, Lazear 2000). Some checklists include the consideration of competition, heterogeneity, pricing, quality, channels, product-line concerns, equilibrium behavior, rationality, managerial usefulness, validity, and so on. Of course, *Marketing Science* entails a high level of rigor that usually entails at least awareness of most of the items on our slightly longer-than-average list.

As reviewers, when using our list, we must remember that the relative importance of each item on the list varies across research objectives (Shugan 2003b). Moreover, emphasizing individual items on the list might either assist or hinder the research objective. We must, therefore, customize our fitness list to reflect the objective of the research effort. We must avoid mechanically using the same list to evaluate every research effort without consideration of the research objective. We certainly want to avoid the exclusion of some types of research merely because those types of research lack one particular item on our paragon checklist.

Recently, we have had a surge of reviewers who have added "endogeneity" to their checklist, and *Marketing Science* is beginning to reject papers because some variables are not endogenous. The objective of the following discussion is to remind everyone that making a variable endogenous has costs as well as benefits. There is no intention here to diminish in any way the critical importance of considering endogeneity in a wide variety of vital applications (e.g., see Hamermesh 1993, Chintagunta 2001). Instead, the intent is to note that in some situations, making variables endogenous detracts from the research objective and diminishes the value of the corresponding research.

As we know, there are probably numerous articles and books dealing with endogeneity. Most often, endogeneity creates a defect in certain types of data (Shugan 2003a). The problem is much more prevalent in naturally occurring data than in traditional market research data (i.e., questionnaires, test markets, telephone surveys) and is less of a problem for experimental data (e.g., Anderson and Simester 2004), although any adaptive question-design method is potentially subject to endogeneity bias (Toubia et al. 2003).

With naturally occurring data, our dependent variable might influence our predictor variables (e.g., see Theil 1971). For example, we estimate the demand for corn as a function of its price, but we forget that the amount of corn grown (i.e., supply) depends on the market price for corn (Working 1927). Ignoring that endogeneity, of course, creates a bias in the estimated price elasticity of demand. Similarly, in marketing, when we estimate the sales for a product as a function of advertising expenditures to determine the impact of advertising on sales (e.g., Vakratsas et al. 2004), we assume that sales do not influence advertising expenditures. That assumption might be inappropriate when marketing managers set advertising expenditures as a percent of sales, and so our data are defective (Blattberg and Neslin 1990).

Adjusting for endogeneity can reveal interesting new findings. For example, Elberse and Eliashberg (2003), by adjusting for endogeneity, find that variables such as movie attributes and advertising expenditures, which are usually assumed to influence audiences directly, mostly influence revenues indirectly, namely, through their impact on exhibitors' screen allocations. Similarly, Boulding and Christen (2003) find that market entry order for new products

*Steven M. Shugan is the Russell Berrie Foundation Eminent Scholar in Marketing.

is endogenous. There are a multitude of other excellent examples.

Remedies

As with any defect in the data, there are at least two solutions. The first solution is the best. We should strive to collect better data. For example, Swait and Andrews (2003) enhance scanner panel choice data with compatible preference data from designed-choice experiments. If we are advising a firm about how to advertise, we should ask the firm to experiment at fixed levels of advertising (Little 1966).

The second solution is to make an assumption about the nature of the endogeneity (i.e., use a strong theory) and directly incorporate that relationship into the estimation (see, e.g., Aaker and Bagozzi 1979). For example, we might assume that advertising levels are set according to the industry average method, the percent-of-sales method, equilibrium profit maximization, according to gross margins, or according to some other rule, and incorporate that relationship into our estimation. We might also assume that the rule is endogenous, or we might only include a new exogenous variable that is only correlated with advertising expenditures (i.e., an instrumental variable). Likelihood-based Bayesian approaches might help compute the joint distribution of price, sales, and possible exogenous variables (Rossi and Allenby 2003).

Of course, when the strong assumption we make to remedy the defect in the data is a good approximation of reality, we can fix the formidable endogeneity problem. If the strong assumption is a poor approximation, in contrast, we could make the pesky problem worse. The same is true of other defects as well. Sometimes, naturally occurring data might just be insufficient. Fortunately, in many cases, firms are willing to experiment on a small-scale purchase environment or in a simulated purchase environment with different marketing strategies to help determine true responses.

Whether illuminating experimentation is possible or not, one could argue that when the objective is measuring a relationship, we need to at least consider all possible defects and model violations in our data (e.g., endogeneity, multicollinearities, nonlinearities, nonnormalities, outliers, heteroscedasticity, self-selection, biased measurement, autocorrelation, etc.) and how each might impact our results. We must also weigh the advantages and disadvantages of available remedies for those defects.

Models for Decision Making

Now, suppose our objective is advising a decision maker (Little 1979). In that case, making variables endogenous might detract from the usefulness of our model.

For example, a decision maker might face exogenous constraints that are unrelated to the context of the model (Lodish 1980). Exogenous factors might impose these constraints. For example, the decision making might be unable to change price because of a promise made to a distributor. A motion picture studio might fund an unprofitable motion picture as part of a previous deal with a director to produce another film. No matter how grand the model, there will usually be many remaining unmodeled variables (Wierenga 2002). Often, it might be best to model those factors as exogenous constraints rather than trying to model them as endogenous.

We could easily find situations in which, for example, an ice cream retailer wants to determine the optimal price for an ice cream line wherein all flavors are constrained to the same price, despite the fact that most pricing models would prescribe different prices for different flavors. In this case, although the model would dictate that the flavors should have different prices, we should not make these prices endogenous. Instead, we might impose a constraint on the model that should not exist within the world of the model.

Although the number of banners on a Web page could certainly be endogenous, given a constraint on the number of banners, it is legitimate to only model their placement (e.g., see Chatterjee et al. 2003).

We can also impose exogenous relationships found in empirical research. For example, Danaher et al. (2003) find that buyers of Internet retailers are more brand loyal to large-market share brands than small-market share brands. That relationship might be used in a model focusing on new-product strategy (in contrast to endogenizing buyer behavior). A model to help design an optimal customer relationship program might employ prior research by assuming an inverted-U effect of preference for sure-small over large-uncertain rewards (e.g., see Kivetz 2003), even though the model itself contains no uncertainty or other behavioral assumptions to justify that exogenous constraint.

In general, a model with exogenous parameters will fit more situations because the decision maker can substitute the appropriate values for these parameters, given knowledge of many factors outside of the model. Were we to make all of these variables endogenous, we would need to make the heroic assumption that the model contained all relevant factors. Hence, models must have boundaries, and these boundaries might exist in the form of constraints on the decision. Moreover, although these constraints might make some models appear incomplete, these models can sometimes be more useful for decision makers.

For example, consider setting the price for a capacity-constrained service. Reviewers often ask authors to make capacity endogenous. The result is an optimal level of capacity within the world of the model (i.e., the variables that the model considers and the model's implicit assumptions regarding the relationships between those variables). With endogenous capacity, the model still prescribes the best price but only given that so-called optimal capacity. However, when we treat capacity as an exogenous variable, the model might prescribe the best price when capacity levels are set based on institutional or environment constraints (e.g., zoning laws prevent a larger facility) outside of the model. Moreover, the modeler can now focus on adding more complexity where needed, rather than having the burden of adding variables to model capacity. For many, but not all, research objectives, allowing exogenous institutional constraints appears perfectly appropriate.

We do recognize that all constraints are merely large costs and, hence, that all constraints could be made endogenous. For example, a restaurant that faces a binding seating constraint could, at some cost, immediately increase capacity by adding more tables, renting tables from a neighboring restaurant, paying some patrons to leave, or simply expanding service to a remote location. Nevertheless, it is still useful to model the restaurant's problem given a binding capacity constraint.

Of course, authors continue to have the burden of showing that their models do adequately fit real-world decisions, at least in some institutions and some industries. Institutional or qualitative facts should justify exogenous constraints or parameter values. Hopefully, by applying different criteria to different types of research, we can further advance the goal of diversity in *Marketing Science*.

In sum, we should understand that there are many factors that contribute to a decision and that our models will seldom replace decision makers. Decision makers must consider multiple constraints not captured by the model (i.e., there is always a bigger picture). By allowing different types of exogenous constraints, we might make our models far more applicable to realistic settings than if we seek to make all variables endogenous.

References

Aaker, David A., Richard P. Bagozzi. 1979. Unobservable variables in structural equation models with an application in industrial selling. *J. Marketing Res.* **16**(2) 147–158.

Anderson, Eric, Duncan Simester. 2004. Long-run effects of promotion depth on new versus established customers. *Marketing Sci.* **23**(1) 4–20.

Blattberg, Robert C., Scott Neslin. 1990. Sales promotion: Concepts, methods and strategies. Prentice-Hall, Englewood Cliffs, NJ.

Boulding, William, Markus Christen. 2003. Sustainable pioneering advantage? Profit implications of market entry order. *Marketing Sci.* **22**(3) 371–392.

Chatterjee, Patrali, Donna L. Hoffman, Thomas P. Novak. 2003. Modeling the clickstream: Implications for Web-based advertising efforts. *Marketing Sci.* **22**(4) 520–541.

Chintagunta, Pradeep K. 2001. Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Sci.* **20**(4) 442–456.

Danaher, Peter J., Isaac W. Wilson, Robert Davis. 2003. A comparison of online and offline consumer brand loyalty. *Marketing Sci.* **22**(4) 461–476.

Elberse, Anita, Jehoshua Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Sci.* **22**(3) 329–354.

Griffin, Abbie. 2001. What constitutes good reviewing. *J. Product Innovation Management* **18**(1) 1–2.

Hamermesh, Daniel. 1993. *Labor Demand*. Princeton University Press, Princeton, NJ.

Kivetz, Ran. 2003. The effects of effort and intrinsic motivation on risky choice. *Marketing Sci.* **22**(4) 477–502.

Lazear, Edward P. 2000. Economic imperialism. *Quart. J. Econom.* **115**(1) 99–136.

Little, John D. C. 1966. A model of adaptive control of promotional spending. *Oper. Res.* **14**(6) 1075–1097.

Little, John D. C. 1979. Decision support systems for marketing managers. *J. Marketing* **43**(3) 9–26.

Lodish, Leonard M. 1980. A user-oriented model for sales force size, product, and market allocation decisions. *J. Marketing* **44**(3) 70–78.

Rossi, Peter E., Greg M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Sci.* **22**(3) 304–328.

Shugan, Steven M. 2002a. In search of data: An editorial. *Marketing Sci.* **21**(4) 369–377.

Shugan, Steven M. 2003b. Defining interesting research problems. *Marketing Sci.* **22**(1) 1–15.

Swait, Joffre, Rick L. Andrews. 2003. Enriching scanner panel models with choice experiments. *Marketing Sci.* **22**(4) 442–460.

Theil, Henri. 1971. *Principles of Econometrics*. Wiley, New York.

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

Vakratsas, Demetrios, Fred M. Feinberg, Frank M. Bass, Gurumurthy Kalyanaram. 2004. The shape of advertising response functions revisited: A model of dynamic probabilistic thresholds. *Marketing Sci.* **23**(1) 109–119.

Wierenga, Berend. 2002. On academic marketing knowledge and marketing knowledge that marketing managers use for decision-making. *Marketing Theory* **2**(4) 355–362.

Working, Elmer J. 1927. What do statistical "demand curves" show? *Quart. J. Econom.* **41**(1) 212–235.