המרכז למחקר בכלכלה חקלאית
THE CENTER FOR AGRICULTURAL ECONOMIC RESEARCH

Working Paper No. 20114

Test Drives: Learning, Sorting, and Buying

by

Amir Heiman, Bruce McWilliams,
Edward Z. Shen, and David Zilberman

THE CENTER FOR AGRICULTURAL ECONOMIC RESEARCH

P.O Box 12, Rehovot, Israel

TEST DRIVES: LEARNING, SORTING, AND BUYING

Amir Heiman, Bruce McWilliams, Edward Z. Shen, and David Zilberman*

January 2001

*Amir Heiman is Lecturer, Department of Agricultural Economics and Management, Hebrew University; Rehovot, Israel; Bruce McWilliams is Visiting Research Economist, Department of Agricultural and Resource Economics, University of California at Berkeley; Edward Shen is a Ph.D. Candidate, Department of Agricultural and Resource Economics, University of California at Berkeley; and David Zilberman is Professor, Department of Agricultural and Resource Economics, and Member, Giannini Foundation of Agricultural Economics, University of California at Berkeley.

The authors would like to acknowledge support from BARD for this research.

TEST DRIVES: LEARNING, SORTING, AND BUYING[1]

Introduction

Customers often engage in search activities to learn about products and brands before making a purchase. This search may be about a product's price, quality, personal fit, or performance, and the process may involve either direct or indirect learning. Indirect learning arises from information that originates from the firm or other people, and includes exposure to advertisement or interactions with previous adopters (word of mouth). Direct learning is any activity that provides the consumer with first-hand experience with the product, and includes free product samples, product demonstrations (e.g., test drives or trying on clothes before purchase), and previous purchases of consumables.

While there is good empirical documentation of the effect of indirect information, and in particular of advertisement, on goodwill and purchase, not much have been done with regards to empirically analyzing the effect of direct information, particularly for durable goods, on belief updating and purchasing decisions. In this paper, we empirically analyze the pre-purchase learning process using data on car test drives. We show how test drives change consumer evaluation of the product and estimate the effect of this updated evaluation on the purchasing decision. In addition, we examine the effect of the length of the test drive on customer evaluation as well as on the final purchasing decision. We find that although the length of the test drive does not affect customer evaluation, it does affect the purchasing decision.

---

[1] We gratefully acknowledge the helpful comments of Arnold Zellner, V. Kerry Smith, and participants of a seminar at North Carolina State.

We model the entire process involved by first identifying the factors that affect the duration of test drive. We then estimate the learning effects of the test drive using a Markov transition matrix. Finally, we show that post-test drive assessments are powerful predictors of product purchase, while initial assessments are weak predictors. This paper verifies empirically that customers indeed update their product evaluations after test drives and that these updated evaluations are critical in the purchasing decision. Test drives act as a mechanism to sort out buyers and non-buyers.

The paper also makes a methodological contribution by using maximum entropy to estimate the parameters of the Markov process. Maximum entropy can be effective in estimating a large number of parameters even when data is limited.

Literature Review

Information and its updating is one of the most basic and important concepts employed in marketing. Diffusion models are based on the transfer of information from either sellers or adopters to potential customers, either implicitly (e.g., Bass, 1969; Norton and Bass, 1987; Dockner and Jorgensen, 1988) or explicitly (e.g., Hagerty and Aaker, 1984; Meyer and Sathi, 1985; Kalish, 1985; Roberts and Urban, 1988). The diffusion process in the explicit learning models is derived by customers updating their purchasing decisions over time as they take previous and new information into consideration. For example, dynamic word of mouth serves as a source of new information that is integrated with prior beliefs and reduces the uncertainty related to purchasing durable products (Kalish, 1985).

There is a great deal of evidence on the impact of indirect information, especially advertisements, on attitudes and purchases (see surveys by Little, 1979;

Feichtinger, Hartl and Sethi, 1994). Less is known on the impact of direct information. Direct information is preferred over indirect information (Smith and Swinyard, 1983; Fazio and Zanna, 1978), but incurs higher search costs and thus will be used when uncertainty is great and when there is high consumer involvement in the search process. Roberts and Urban (1988) analyze the effect of order of direct (automobiles test drives) and indirect information (video and salesperson) on attitude, Heiman and Muller (1996) document the effect of product demonstrations on sales, and Erdem and Keane (1996) analyze the effect of previous purchases on consumer beliefs in non-durables.

Test drives are a form of demonstration that enable prospective consumers to experience and learn about a particular car model prior to making a purchase without obliging the customer to buy the product. The experience may begin with the customer viewing the car in the showroom, getting inside to get a feel for it, and then taking the car for a drive.

One view of demonstrations is that they are primarily a persuasive tool that makes consumers commit to the product. For example, some studies suggest that demonstrations have a promotional effect since they often make consumers feel obligated to buy the product after they have a test drive (Freedman and Frazer, 1966; Scott, 1976). Aaker (1994) appears to support this perspective when he writes "Automobile customers are used to high-pressure salesmen pouncing on them as they arrive at the showroom and immediately pressuring them into a test drive and a purchase decision."

An alternative view of demonstrations is that they are a valuable learning tool whose effectiveness depends on the characteristics of the product and consumer, as well as on the design of the demonstration. The seller's decision to offer demonstrations and the consumer's willingness to try them are discussed extensively in Heiman, McWilliams, and Zilberman (2001). Heiman and Muller (1996) find that that the length of the demonstration matters. Short demonstrations may provide too little information and thus mislead the prospective consumer, while long demonstrations may be inefficient.

Following Roberts and Urban (1988) and Heiman and Muller (1986), we assume that demonstrations, in the form of test drives, serve primarily as a learning tool. The focus in this paper is on establishing the role of test drives on the learning process, a link that has not been analyzed empirically. We estimate the transition from prior to posterior assessments that results from test driving a particular car model and examine the impact of these assessments on the purchasing decision. We argue that test drives function as a mechanism for sorting customers into buyers and non-buyers by allowing potential customers to identify and realize their true assessment of the product. By providing this learning mechanism, sellers allow customers to move from a position of relative uncertainty to one of greater certainty, thereby increasing customer confidence in the purchasing decision.

Early studies of consumer purchases of automobiles suggest that an individual's stated likelihood of purchasing cars is not a good indication of his actual purchasing decision (Juster, 1966; Ferber and Piskie, 1965). These findings are similar to the finding in this paper that initial assessments of customers that enter a dealer are not a good indicator of their purchasing decision. Urban, Hauser and Roberts (1990) discuss how preference for automobile brands can depend on promotional activities and word of mouth, and observe that test drives positively affect consumer perception of the brands being tested. We will examine whether this test drive effect is uniformly positive for our sample. Unlike these previous studies, our sample consists of people that show up at a dealer and therefore are actually considering buying a car. This allows us to analyze the impact of test drives on both product assessment and purchase.

In this paper, we use the Markov chain as the statistical framework for modeling the learning process. An individual initially rates the product and then, after learning more about the product through a test drive, modifies his rating. We demonstrate that learning, represented by Markovian updating, can explain observed empirical data.

Some empirical studies use the Markov chain to estimate brand choice of non-durables when learning is provided by repeat purchasing opportunities (Blattberg and Sen, 1975; Givon, 1984; Zufryden, 1986), and where the dynamics may be affected by advertising (Erdem and Keane, 1996). We have not found any empirical studies that model the pre-purchase learning process and estimate the transition in perceptions for durable products.

The Markov chain model has been estimated using various classical statistical methods such as least chi-square, maximum likelihood, and Bayesian methods (Lee, Judge and Zellner, 1970). Here we use maximum entropy to estimate the Markov chain since it is an effective tool for estimating a large number of parameters with limited data (Golan, Judge and Miller, 1996). Golan, Judge and Perloff (1996) showed that the generalized maximum entropy approach of the multinomial logistic model yields better estimates of parameters in terms of MSE compared with the classic maximum likelihood approach, especially when the sample size is small.

Maximum entropy (ME) is based on information theory. It was the innovation of Jaynes (1957a, 1957b) and Shannon (1948) who defined entropy as a measure of the level of information about a phenomenon (see Appendix 1 for a brief introduction to maximum entropy). While ME models were introduced in the 1950s, their use in economics and marketing is a recent phenomenon.[2]

Survey Data

The data for the empirical model were collected in 1991-92 from a sole distributor of a new model of a high-end European sports car in Israel. Customers were screened by salespeople to ensure they were seriously interested in purchasing the car and had sufficient income to pay for it (the car cost about $35,000). The surveys were administered by an employee of the sales department, but the salesmen were not given the responses of the clients to the various questions. The data collection had four stages:

Stage 1: Customers were asked to fill out a questionnaire about their perception of the quality of the car, rating it on a scale of one to ten. Customers were also asked whether they had previous experience with the car brand or model.

Stage 2: The customer arranged the scheduling and duration of the test drive with the salesman, and the agreed upon length of the test drive was recorded by the survey administrator before the customer took the car. The scheduling and duration of the test drive depended on the availability of the car as well as on the needs and constraints of the customer.

Stage 3: When customers returned the car they were asked to rate it again using the same 10-point scale.

---

[2] Golan, Judge, and Miller (1996) developed ME into a classical linear regression model. Stutzer (1996) uses the maximum entropy framework to model option pricing in finance market. The recent developments of ME include Golan, Judge, and Perloff's (1996) multinomial logistic regression, Golan, Perloff, and Shen's (1999) simultaneous equations with a non-negativity constraint, Golan, Judge, and Karp's (1996) dynamic Euler equations, and Golan, Judge and Robinson's (1994) input-output matrices.

Stage 4: Customers negotiated with the salesperson and decided whether the purchase the car or not. The salesman reported whether the car was purchased.

The above process indicates that the duration, posterior assessment and decision to purchase were made and reported sequentially. The data collection procedure was designed to generate a recursive causal chain and thus minimize endogeneity and simultaneity problems (Wold, 1953; Green, 1993; Judge et. al, 1988). When customers give their ex-post assessment of the car, the duration of the test drive is predetermined. The duration of the test drive and ex-post assessment are lagged variables when the purchasing decision is made.

To allow customers the opportunity to evaluate this new car model, which the dealer believed had substantially superior reliability than previous models, the company decided to allow potential consumers to use the car unsupervised for up to 72 hours, and in many cases longer[3], rather than the 10-20 minutes with supervision typically allotted to customers. The mean test drive duration was 18.8 hours. In fact, the duration of test drives varied considerably, with 29% of the customers test driving the automobile for one hour or less and 39% keeping the car for 24 hours or longer.

There were 215 usable observations.[4] Of these, 43 (20%) had previous experience with the car and 94 (43.7%) purchased the car. To reduce the number of estimated parameters and improve the significance and interpretability of the results (particularly for the Markov transition matrix) given the size of our sample, the 10-point scale is reduced to a four-point one by grouping the evaluations as follows: 1-4

---

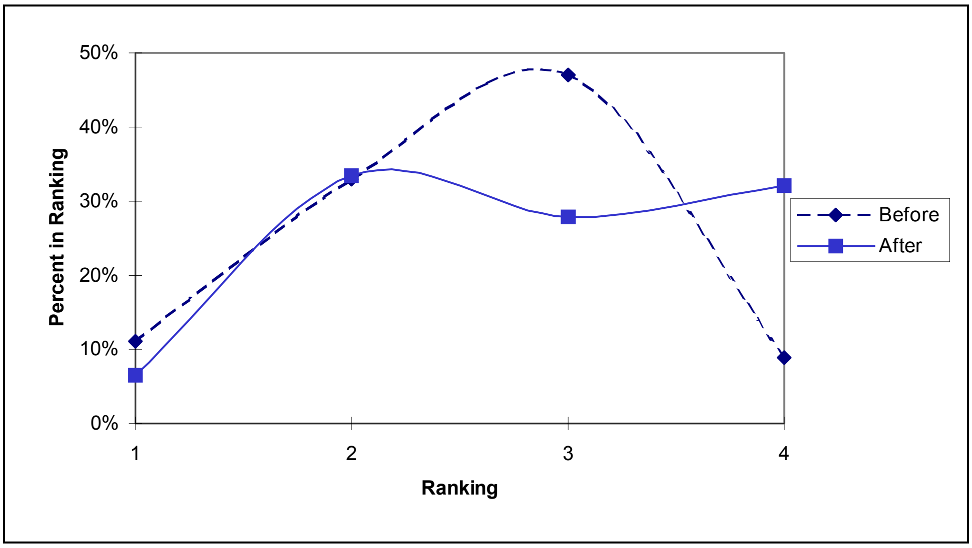
[3] Three customers test-drove the car for 6 days, the longest test drives recorded in the sample.

[4] 15 observations were rejected because the individuals test-drove the cars but did not fill out the survey and did not purchase the cars. We consider them to be curious people rather than serious buyers.

= 1; 5-6 = 2; 7-8 = 3; and 9-10 = 4. The discussion below analyzes the data using this revised rating and provides an interpretation for these rating scales.[5]

Figure 1 graphically shows the distribution of customer ratings before and after the test drives. We see that the assessments before the test drives are represented by a curve that peaks at a rating of '3', with close to half of the customers falling into this category. By contrast, the assessments given after the test drives have a double-peak distribution with modes at '2' and '4'. Table 1 below gives the numerical values associated with these curves. One interpretation of the redistribution of the single-peaked prior assessments to a double-peaked posterior distribution is that test drives are a learning tool that help customers identify whether the car fits their needs or not.

Figure 1: Percent of Customers in Each Rating Before and After Test Drives



[5] We use this revised 4-point scale throughout the paper so that the rating scales used in the equations for estimating test drive duration and the likelihood of buying are consistent with the scale chosen for estimating the Markov transition matrix.

Table 1: Number of Assessments in Each Rating Before and After Test Drives

Assessment Level	1	2	3	4	Total
Before Test Drive	23	72	100	20	215
After Test Drive	14	72	60	69	215

Figure 2 diagrams the relationship between the consumers' final decision to purchase a car and their ratings before and after test drives. The graph indicates that ratings before the test drives (dotted line) are not a good predictor of purchase, which is consistent with Juster (1966) and Ferber and Piskie's (1965). By contrast, the ratings after the test drives are highly correlated with the likelihood of purchasing the product.

Figure 2: Relationship Between Purchases and Prior and Posterior Ratings

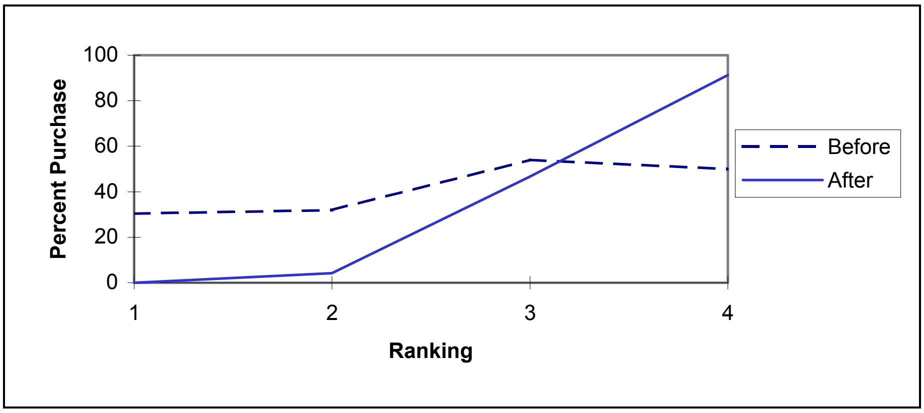


Table 2 gives the same relationship numerically. Of those who gave ratings of '1' and '2' before test drives, 30-32% of them eventually purchased a car, while for those with initial ratings of '3' and '4', 50-54% purchased. Therefore, the relationship between the final purchasing decision and initial rating is weak. By contrast, the relationship between the final purchasing decision and final ratings is very high. Less than 5% of customers with low after-test drive assessments (‘1’ and ‘2’) purchased a car, while over 90% of those with a rating of ‘4’ purchased. The difference in predictability of purchase between pre- and post- test drive assessments suggests an important role for test drives in affecting the purchasing decision.

Table 2: Percent Purchases Given Ratings Before and After Test Drives

Assessment Level	1	2	3	4	Total
Before Test Drive	30.4	31.9	54.0	50.0	43.7
After Test Drive	0	4.2	46.7	91.3	43.7

Model

We develop a model of the use of test drives in order to examine the relationship between test drives and product purchases. We assume the all decisions are made in a sequential manner, with the customer taking past decisions, such as the duration of the test drive, as fixed when re-assessing the car and making the final purchasing decision. The decision-making process affected by test drives is modeled in three stages: 1) determination of test drive duration; 2) the test drive’s impact on updating the customer’s product assessment; and 3) the purchasing decision.

A. Test Drive Duration

Test drives allow customers to experience products before they purchase them. Testing a product provides the customer the opportunity to determine if the product meets his personal standards, needs and tastes, thereby reducing his pre-purchase uncertainty about the product.

Suppose there are J customers considering the purchase of a product. Each customer j has an initial evaluation that is an integer i between 0 and I. This rating reflects the consumer's subjective evaluation of the product's value relative to its price, which we assume to be given, where a rating of I means the product is considered an excellent purchase, while a rating of zero corresponds to the product being a bad purchase. Let the vector $Q_0^j\left(I\times 1\right)$ represent the initial rating of individual j. If the customer's initial rating is the integer i, then the ith element of $Q_0^j=\left(q_{01}^j,\ldots,q_{0I}^j\right)$ is equal to 1 and all other elements are zero. The customer is then exposed to a product demonstration at the store. In the case of a car, the demonstration is a test drive. The duration of the test drive for individual j is D^j and may vary across consumers. In particular, we assume that the log duration of the test drive is a linear function of initial assessment of the product $\left(q_{0i}^j\right)$, product experience $\left(E^j\right)$, and other unobserved factors (ξ^j):

(1) $$\ln D^j = a + bE^j + \sum_{i=2}^{I} c_i q_i^j + \xi^j$$

where a is a constant term and b and c_i are coefficients to be estimated. Although theoretically the lower bound test drive length is zero duration, all of the observations are positive, with two observations found at 10 minutes (the shortest test drives). In addition, there is no specific upper bound, with three customers taking the car for 6 days (the longest test drives). The lack of binding upper and lower boundaries makes a tobit model irrelevant, so OLS was used to estimate the equation.

We develop several hypotheses. Prior knowledge about a product reduces the buyer's value and need for additional search. This is consistent with Lindner et al. (1979) where increased prior knowledge reduces experimentation time.

Regarding the impact of initial assessment on the test drive length, Leland (1994) suggests that there are more benefits from learning when individuals' assessments are mild, rather than strongly favorable or unfavorable. This implies that individuals with either very high or low initial assessments will demand less test drive time than individuals whose assessments are intermediate since extreme assessments reflect greater certainty about a product than do intermediate ones. Therefore demand for test drive time will be an inverted *U* with respect to initial rating.

An alternative hypotheses is that consumers with high initial assessments will be more serious about the product and therefore will be willing to invest more time in test drives before making a purchasing decision. Although sellers do not observe the initial assessments of the customers, they may be able to recognize customers with high initial assessments and are willing to provide longer test drives to the customers that are more serious about the product. This implies that customers with high initial assessments are more likely to be offered longer test drives.

From the above, the test drive duration for each consumer is an outcome that balances consumer demand for test drive time with the seller's willingness (or ability) to provide it to different buyers. The hypotheses from this discussion are:

H_1: Previous experience with the product leads to shorter test drive time, i.e., $b < 0$.

H_{2a}: the relationship between initial assessment and test drive duration is an inverted *U*, i.e., there exists an i^*, $0<i^*<I$, such that $c_i \geq c_{i-1}$ for $i \leq i^*$, $c_i \leq c_{i-1}$ otherwise.

H_{2b}: the relationship between initial assessment and test drive duration is positive, i.e., $c_i \geq c_{i-1}$ for all i.

B. Assessment Modification from Test drives

Although the test drive duration is of interest in our analysis, the essential concerns are the estimation of the effect of test drives on the consumer's product assessment and how these affect the purchasing decision. In this section we will introduce a methodology for estimating the change in consumer assessment that results from the test drive.

Upon experiencing a test drive, the consumer revises his prior assessment of the product and forms a new rating that is represented by a vector Q_1^j. For example, if individual *j* has an initial rating of 3 out of four, then the initial rating is the vector $Q_0^j = (0,0,1,0)$. The Markovian rating modification process for customer *j* is:

(2) $$Q_1^j = P^j Q_0^j + \varepsilon^j$$

where ε^j is the error term, and P^j is a transition matrix of dimension *I x I.* The *n,m* element of this matrix $\left(P_{nm}^j\right)$ represents the expected probability that individual *j* will move from a prior rating of *m* to a posterior rating of *n* as a result of the test drive, and since there are only I possible posterior ratings, $\sum_n P_{nm}^j = 1$ for $m = 1,\dots,I$. The vector of random variables ε^j has dimension *I x 1*, and each component lies between –1 and 1 and is assumed to have mean zero.

The rating modification is assumed to follow a Markov chain that is affected by the test drive duration $\left(D^j\right)$ and the consumer's prior experience with the product $\left(E^j\right)$. The transition matrix for customer *j* can therefore be expressed as

(3) $$P^j = A + Bln\left(D^j\right) + CE^j,$$

where $P^{j}_{n,m} \geq 0$ for all elements of the matrix, and A, B, and C are I x I transition matrices to be estimated. The matrix A, where $\sum_{n} A_{nm} = 1$ for $m = 1,\ldots,I$, is the basic test drive Markov transition matrix. Here, element A_{nm} represents the transition probability of moving from an initial rating m to posterior rating n in the benchmark case where the customer has no prior experience and test drive length is equal to 1 hour $\left(\text{i.e., } ln\left(D^{j}\right) = 0\right)$. Matrices B and C give the effect of test drive duration and previous experience, respectively, on the transition probabilities. The test drive duration and prior knowledge do not affect the sum of the transition probabilities, which must sum to 1, and can be written: $\sum_{n} B_{nm} = 0$, $\sum_{n} C_{nm} = 0$, for $m = 1,\ldots,I$. The model without B and C is a basic Markov chain model. Adding these two explanatory variables expands this basic model and allows us to investigate the impact of test drives and prior experience on the transition probabilities.

To estimate the coefficients within the ME framework, we follow the work of Golan, Judge, and Miller (1996) for parameterizing the system; a method they called generalized maximum entropy (GME). Details of the GME approach for estimating the Markov transition matrix are given in Appendix 2, and the derivation of the standard errors of the estimated coefficients are given in Appendix 3.

Several hypotheses can be tested with the data. The three natural null hypotheses are:

H_3: Test drives do not change customer's assessment of the car, i.e., A is an identity matrix ($A_{nm} = 1$ if $n=m$, and 0 otherwise).

H_4: The duration of the test driving does not affect the customer's assessment updating process, i.e., B is zero matrix ($B_{nm} = 0$ for all n,m).

H_5: Having a prior experience with the car does not affect the customer's assessment updating process, i.e., C is zero matrix ($C_{nm} = 0$ for all n,m).

If test drives affect customer assessments, H_3 will be rejected. Two alternative hypotheses regarding matrix A can also be tested:

H_6: Test drives increase customer's assessment of the car, i.e., A is a positive lower diagonal matrix ($A_{nm} > 0$ if $n \geq m$, and 0 otherwise).

H_7: Test drives decrease customer's assessment of the car, i.e., A is a positive upper diagonal matrix ($A_{nm} > 0$ if $n \leq m$, and 0 otherwise).

Hypotheses 6 and 7 are designed to test whether test drives only have a positive or negative effect on consumer assessment. If H_3, H_6, and H_7 are all rejected, then the test drives have both positive and negative effects on assessments. An interpretation of this would be that test drives act as sorting mechanisms that allow uncertain consumers to identify whether the product fits their needs or not.

If the duration of the test drive affects the change in consumer assessment, H_4 will be rejected. This may be the case if longer test drives give customers more information about the car and allow them to be more confidence about changing their mind. If this is true, the diagonal elements of the matrix would be negative and non-diagonal elements positive. Alternatively, if additional information is randomly positive, negative or consistent with expectations, longer test drives may not be expected to consistently affect consumer assessments in a particular direction and H_4 will not rejected. A weaker test of the hypothesis to see if particular elements of the diagonal in matrix B are different from zero (in particular, negative).

If the customer's prior information about the car affects his change in assessment, H_5 will be rejected. An alternative hypothesis that can be tested is:

H_8: Customers with prior experience are more likely to stick with their initial assessment after the test drive, i.e. $C_{nm} > 0$ for $n=m$.

This hypothesis is consistent with the implications of the general learning hypothesis that people who have gathered more information about a product will be more confident about their assessment and therefore less likely to change it. A weaker version of the test is to see if particular elements of the diagonal in matrix C are positive.

C. Purchasing Decision

In the final stage, the customer decides whether or not to purchase the product after experimenting through the test drive. Let Y^j be the indicator of the decision to purchase by customer *j* where $Y^j = 1$ if the customer purchases and $Y^j = 0$ if he does not. We use a logistic model to estimate the final decision to purchase a car or not. The estimating equation can be expressed as

$$\Pr\left(Y^j = 1\right) = \left[1 + \exp\left[-\left(\alpha + \sum_{i=2}^{I} \beta_i q_{1i}^j + \gamma \ln D^j + \phi E^j + \delta Pos^j + \chi Neg^j\right)\right]\right]^{-1}$$

where $\Pr(Y^j = 1)$ is the customer's probability of purchasing the car, q_{1i}^j is one if it matches customer *j*'s posterior assessment and zero otherwise, ln*D* is the duration of the test drive, *E* is whether the customer had previous experience with the car, *Pos* is a dummy variable equal to one if the customer's posterior assessment is higher than his initial assessment, and *Neg* is a dummy variable equal to one if the customer's posterior assessment is lower than his initial assessment. α is the constant, and $\beta, \gamma, \phi, \delta$, and χ are coefficients to be estimated.[6]

---

[6] We recognize there may be concerns about the endogeneity of the duration of the test drives (lnD) in this equation. However, given the very limited nature of the data, we do not have instruments

Several hypotheses can be tested by this estimation. One key hypothesis is that the information used for the customer's purchasing decision is fully summarized by their final assessment. In this case, the sole determinant of purchasing is the customer's posterior assessment of the car. This implies first that customers with higher posterior assessments will be more likely to purchase the car, and second, that no other variables will significantly affect the purchasing decision.

Alternative theories would suggest that other variables may also affect the purchasing decision. For example, longer test drives may positively influence the purchasing decision for two reasons. First, the posterior ranking is only an estimate of the car's true quality, and there is uncertainty related to this estimate. If longer test drives increase customer knowledge and confidence in the product, they will have a lower threshold for purchase (Lindner et al., 1979) and the effect of test drive duration will be positive. An alternative explanation for expecting a positive effect of test drive duration on purchase is that longer test drives make the customer feel more "obliged" or "committed" to purchasing the product because of the seller's generosity (see Kotler, 1994; Johnson, 1995; Graham, 1995; Anderson, Ross and Weitz, 1998).

The customer's previous experience with the car is also expected to have a positive effect on purchasing a car. Customers that have previous experience with the car may be more knowledgeable and confident about their purchase and therefore have a lower sufficiency threshold for purchase, as discussed in the previous paragraph under test drive duration effects.

---

to substitute for length of demonstration. Moreover, durations were agreed upon and recorded before the test drive and were a lagged variable when the purchasing decision was being made. We also assume that the random element that affects the discrete choice (and thus leads to the logit formulation (Green, 1993; Judge et. al, 1988)) is obtained after the test drive and is uncorrelated with its duration. In this case, simultaneous estimation methods are not necessary for deriving consistent and efficient estimators (Wold, 1953; Green, 1993; Judge et. al, 1988, ch. 13).

Finally, if the test drive disconfirms the prior expectations of the consumer, this "disconfirmation" may affect the likelihood of purchasing the product beyond updating the assessment. A "disconfirmation" experience in our case is defined as an experience that causes the customer to change his prior assessment to a more positive ("positive disconfirmation") or negative ("negative disconfirmation") assessment after the test drive. Hogarth and Einhorn (1992) discuss under which circumstances people give greater weight to initial or more recent information received, and find that this weighting of sequenced information depends on the complexity of the information and whether it is processed at each stage or after all information is received. Viscusi (1997) finds that when there is risk of a bad outcome and people receive conflicting information from different sources, they will give disproportional weight to the negative information received. We can test whether consumers give additional weight to positive or negative changes in assessment when making their purchasing decision. We do not have an a priori expectations of the direction of influence of these variables.

The hypotheses discussed above can be summarized as follows:

H_9: Customers with higher posterior assessments are more likely to purchase the car, i.e., $\beta_i > \beta_{i-1}$.

H_{10}: Joint hypothesis test that all other effects are zero, i.e., $\gamma, \phi, \delta,$ *and* $\chi=0$.

H_{11}: Longer test drives will have a positive effect on car purchase, i.e., $\gamma>0$.

H_{12}: Prior experience will have a positive effect on car purchase, i.e., $\phi>0$.

H_{13}: Positive (negative) disconfirmation will have no effect on car purchase, i.e., $\delta,\chi=0$.

The empirical model can be modified so that the updated assessments are replaced by the initial assessments to compare the effectiveness of prior and posterior assessments in explaining purchasing behavior.

Statistical Results

Our empirical results include the three stages presented in the modeling section. First, we estimate the determinants of test drive length using experience and initial car assessment as explanatory variables. This analysis uses simple linear regression techniques. Second, we estimate the Markov transition probabilities for changing customer assessments due to test drives, and include the impacts of test drive length and prior knowledge on these probabilities. Entropy is used in this analysis. Finally, we estimate the determinants of the purchasing decision using the updated assessments, test drive length, and experience. This is done using a multivariate logistic regression.

A. Test Drive Duration

We used a linear model to estimate the determinants of the duration of the test drive for customers who tested the car model. We found significant heteroskedasticity in the residuals, whose variance depends on prior rating and experience. In particular, customers with a high rating and previous experience with the car had lower variances in their test drive duration than the rest of the sample.[7]

Table 3 below shows the results of the estimation model. The base group is those with a prior rating of '2' and without experience. All of the variables, except for

[7] We present the results of the basic OLS regression since accounting for heteroskedasticity in a GLS estimation did not affect the coefficients, t-statistics, or R-square. The significance of the estimated coefficients and the consistence of the coefficients estimated with subsets of the sample and other tests such as VIF indicate that multicollinearity is not a problem in this estimation.

the prior rating of '1', have coefficients that are statistically different from zero at a significance level of 99 percent or more. The results reveal that individuals with higher initial ratings are likely to have longer test drives. This may reflect the seller's priority to first accommodate the needs of buyers who are more likely to buy the product or indicate the willingness, on the part of buyers with high initial assessments, to take their time with test drives. As predicted, individuals with prior experience with the car tend to have shorter test drives. The statistical significance of the equation as a whole is very high as measured by the F-test, with a P-value of 0.001. The low R-square reflects the fact that a variety of unobserved factors affect the test drive duration.

Table 3: Duration of Test drive as a Function of Prior Rating and Experience

Variables	Coefficients	Standard Errors	t-statistics
Prior rating = 1	0.399	0.415	0.959
Prior rating = 3	0.888	0.289	3.069
Prior rating = 4	1.322	0.466	2.835
Previous Experience	-1.205	0.329	-3.668
Constant	1.431	0.209	6.849
R-square	0.082		
F-test (from mean)	4.698		
P-Value	0.001		

B. Estimating the Markov Process Transition Matrix

In this section the Markov transition probabilities introduced in the modeling section are empirically estimated. The maximization was implemented using GAMS, and the estimated transition probabilities are presented in Table 4. The transition probability matrix *A* represents the effect of having a test drive on the likelihood of a transition from an initial rating (represented by the column) to a posterior rating (represented by the row). Matrix *B* represents the effect of test drive duration, and matrix *C* represents the effect of previous experience on these transition probabilities. Column 1 in matrix *C* is zero because none of the people with experience had an initial rating of '1'.

To demonstrate how to interpret the table, the number 0.273 in element 3,2 of matrix *A* means that, if a customer has an initial rating of '3', he will have a 27.3 percent likelihood of moving to '2' after the test drive. Each column in matrix A sums to one. The columns in matrices *B* and *C* sum to zero since these variables marginally change the probability distributions given in matrix *A*. The interpretation of matrix *C*, for example, is that a person with previous experience who initially gives a rating of '4' would be more likely (by 32 percent) to move to an ex-post rating of '2', and less likely (by 31 percent) to move to '3', than customers who did not have previous experience.

Hypotheses 3, 6 and 7 are related to matrix A and the basic impact of test drives. We reject the hypotheses that test drives have no effect (H_3), a uniformly positive effect (H_6), or a uniformly negative effect (H_7). We are therefore left with the conclusion that test drives have an effect and this effect can be bi-directional. Our interpretation is that test drives help customers identify whether the product is a personal fit for them or not and therefore act as a sorting mechanism. Each element of the diagonal in matrix A is significantly different from 1, indicating that some customers are likely to change their assessment whatever their initial assessment was. The fact that the coefficient for element 3,3 in matrix A (0.348) is smaller than the coefficients in the other diagonal elements suggests that customers who began with an initial assessment of 3 were less certain about their initial assessment and therefore more willing to change their mind.

Matrix *B* shows how the duration of the test drive affects the transition probabilities. Hypothesis 5, that the duration of the test drive has no effect on changes in assessment, cannot be rejected. An examination of the results in matrix *B* shows that time had a significant negative impact on only one transition probability. For those with a low initial rating of'2', longer test drives significantly increased their likelihood of changing their assessment, tending toward an upward transition. The results indicate that in general the duration of the drive has no effect on the direction or likelihood of changing ones assessment which is consistent with the interpretation that new information gained from additional experimentation is not expected to influence customers in a consistent direction.

Matrix *C* gives the change in the transition probabilities from having prior experience with the product. Hypotheses 5 and 8 are related to this transition matrix. H_5 is rejected, indicating that previous experience does affect the customer's updated probabilities. Moreover, H_8 is not rejected, indicating that we cannot reject the hypothesis that customers with previous experience are less likely to change their assessment. In particular, element 2,2 of the diagonal is significantly positive and the others are positive but not significant.

To summarize, the results indicate that: 1) test drives act as a mechanism to sort customers into those for which the product is a fit or non-fit; 2) the duration of the test drive does not affect the transition of assessments; and 3) prior experience reduces the likelihood of the customer changing his initial assessment.

Table 4: Estimated Transition Probabilities

	Matrix A: Fixed test drive effect				Matrix B: Test drive duration effect				Matrix C: Prior experience effect			
	Initial rating				Initial rating				Initial rating			
Posterior Rating	1	2	3	4	1	2	3	4	1	2	3	4
1	0.489 (0.050)[A]	0.017 (0.025)	0.007 (0.029)	0.079 (0.072)	0.008 (0.0202)	0.003 (0.011)	-0.001 (0.010)	-0.016 (0.026)	0	-0.012 (0.063)	-0.002 (0.033)	-0.017 (0.066)
2	0.134 (0.080)[C]	0.580 (0.055)[A]	0.273 (0.035)[A]	0.136 (0.067)[B]	-0.013 (0.0468)	-0.053 (0.025)[B]	-0.005 (0.020)	0.037 (0.043)	0	0.330 (0.079)[A]	-0.060 (0.044)	0.316 (0.075)[A]
3	0.263 (0.102)[B]	0.158 (0.061)[A]	0.348 (0.057)[A]	0.322 (0.106)[A]	-0.020 (0.0472)	0.024 (0.026)	0.006 (0.023)	0.000 (0.053)	0	-0.116 (0.084)	0.064 (0.065)	-0.306 (0.090)[A]
4	0.114 (0.107)	0.245 (0.062)[A]	0.371 (0.065)[A]	0.463 (0.108)[A]	0.025 (0.0490)	0.025 (0.027)	0.001 (0.024)	-0.021 (0.050)	0	-0.201 (0.094)[B]	-0.003 (0.079)	0.007 (0.083)

Numbers in parentheses are standard errors. The superscripts indicate that estimated coefficients are significant at the A=1, B=5, and C=10 percent level.

Table 5: Hypothesis Testing

Hypothesis	*H_0*	*Chi-square*	*Reject? (at 1% significance level)*
H_3: No test drive effect	A=1 on the diagonal and 0 elsewhere	610.45	Yes
H_4: No duration effect	B=0 for each element	6.47	No
H_5: No experience effect	C=0 for each element	174.52	Yes
H_6: Test drives increase assessment	$A_{nm}>0$ if $n \geq m$	110.12	Yes
H_7: Test drives decrease assessment	$A_{nm}>0$ if $n \leq m$	196.68	Yes
H_8: Experience increases stability of initial assessment	C>0 on the diagonal	0.00	No

C. Purchase Decision

We use a multivariate logistic regression to estimate the purchasing decision as a function of the updated rating, test drive duration, prior experience and change in assessment (positive or negative). The results of this analysis are given under "Model 1" (the "Full Updated Assessment Model") in Table 6. The base group for this analysis are the customers with updated (final) ratings of '1' and '2', who had no previous experience with the model, and who ended with the same assessment they began with. The results show that, as expected, customers giving final ratings of '3' and '4' have a significantly increased likelihood of purchasing the car, with a larger coefficient for those giving ratings of '4'. Therefore H_9 is supported.

Model 3 is designed to test hypothesis 10 that posterior assessment of the car contains the total information used for the purchasing decision. The chi-square for comparing models 1 and 3 (H_{10}) is 13.5 with 4 degrees of freedom. Thus H_{10} is only marginally rejected (at the 10% level), suggesting that other variables also influence the final purchasing decision, but this influence is limited.

Test drive duration has a mildly increasing effect on the probability of purchase, thus somewhat supporting H_{11}. This may suggest that customers with longer test drives were more confident of their judgment and therefore more willing to accept a lower sufficiency threshold for purchase the product. An alternative claim is that customers feel obliged to purchase a product if they take a long test drive, or feel committed to buy because of the generous demonstration offered by the seller.

Prior experience with the car brand also significantly increases the probability of purchase, supporting H_{12}. The explanation given earlier for expecting this is that people with product experience will be more confident about their assessment and, therefore, will require a lower sufficiency threshold (product rating) in order to purchase. This translates into a positive purchasing effect.

Finally, we found that if customers increased their assessment after test drives (they were positively surprised by the performance of the car in the test drive), this significantly increased their likelihood of purchase (beyond affecting their final assessment). Thus the hypothesis of no effect (H_{13}) is rejected. Interestingly, we did not find a corresponding negative effect on purchase if customers were negatively surprised (decreased their assessment after test drives). In fact, the negative surprise coefficient is positive, though not significant. Therefore when information received differs from expectations (i.e., when customers change their assessments) customers appear to give greater weight to the positive information when making the purchasing decision. These results appear to differ from those found by Viscusi (1997) in which people receiving contradictory information from two different sources about potential health risks gave greater weight to the negative information.

Table 6: Determinants of the Purchasing Decision

Variables	Model 1: Full Updated Assessment Model	Model 2: Initial Assessment and Experience	Model 3: Updated Assessment Only
Updated rating = 3	3.419 (0.894)[A]		3.187 (0.640)[A]
Updated rating = 4	5.373 (1.024)[A]		5.672 (0.725)[A]
Log of test drive duration	0.235 (0.143)[C]		
Prior experience	1.688 (0.580)[A]	0.512 (0.381)	
Positive change in assessment	0.904 (0.386)[B]		
Negative change in assessment	0.565 (0.632)		
Initial rating = 3		0.784 (0.318)[B]	
Initial rating = 4		0.555 (0.527)	
Constant	-4.790 (0.996)[A]	-0.785 (0.221)[A]	-3.320 (0.586)[A]
McFadden R-square	0.538	0.042	0.492
Log- Likelihood Value	-68.11	-141.19	-74.86
Prediction success table			
Number of correct predictions	184	132	178
Percent correct predictions	85.6%	61.4%	82.8%

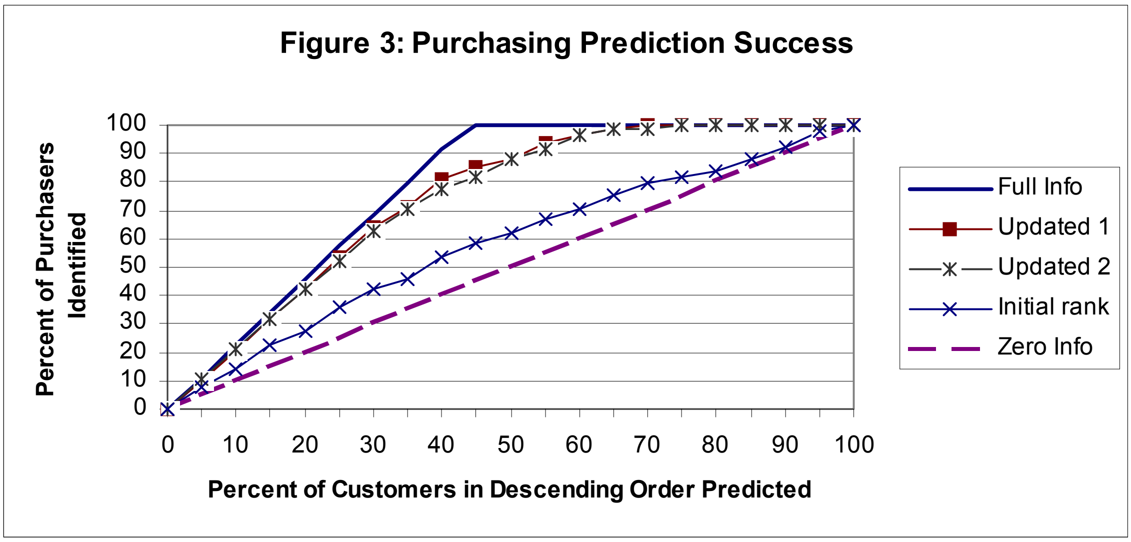
Numbers in parentheses are standard errors. The superscripts indicate that estimated coefficients are significant at the A=1, B=5, and C=10 percent level.

To compare the effectiveness of initial and final ratings in terms of their ability to explain the purchasing decision, we ran Model 2, which uses the initial customer assessments and previous experience with the product to explain product purchase. Those with an initial rating of '3' were found to significantly adopt more than those with low ratings ('1' or '2'), while those with initially very high assessments of '4' did not purchase significantly more. In Model 2, previous experience with the car is not a significant determinant of the purchasing decision.

The models that use updated assessments (Models 1 and 3) both have R-squares around 50%, while the model using initial assessments (Model 2) has an R-square of only 4%. The predictive success of identifying whether a specific customer will buy or not is over 80% for models using updated assessments, while the predictive success from using initial assessments is only 61%, which is only slightly better than the 56% success rate that can be achieved from guessing that no one had purchased the product. These results show the importance of test drives on consumer product assessment and the final purchasing decision.

An alternative measure is to see whether the models correctly rank customers according to their likelihood of purchase given their actual purchasing decision. In this case, it is not the absolute value of an individual's predicted likelihood of purchase that is important, but rather its relationship to the predicted values of other customers. That is, can the model sort out those who are most (or least) likely to purchase the car? Figure 3 provides such a comparison of the models based on the actual purchasing decision. The horizontal axis represents the total percent of the customers, who are selected in descending order of their predicted purchasing probabilities given by the models. The vertical axis represents the cumulative percent of purchasers identified by that selection. The 45-degree (zero information) line represents the effective lower bound for the models since it represents the expected likelihood of correctly identifying buyers if they were chosen randomly. The "full information" line represents the upper bound potential where all of the buyers are selected first and non-buyers are selected last.

The “Updated 1” and “Updated 2” curves represent the effectiveness of sorting out buyers from non-buyers using the predicted values estimated with the updated (post-test drive) ratings from models 1 and 3 respectively. The “Initial” curve represents the effectiveness of sorting out buyers from non-buyers using the initial (pre-test drive) ratings. The predicted values for the models that use the posterior assessments are very effective in sorting out buyers from non-buyers since the “Updated” curves are very close to the full information curve, while the estimates using the initial ratings provide a poor sorting mechanism since the curve lies close to the 45-degree line.



We may also assign “sorting effectiveness” indexes (similar to Gini index) to these curves, which is the proportion of the area under the predicted curves in the area between the full information curve and the 45-degree line. The index has a value of zero for the 45-degree line and one for the full information line. The sorting effectiveness index for the “Updated 1” curve is 0.869, for the “Updated 2” curve is 0.850, and for the “Initial” curve is only 0.279. The similarity between the two updated models, and the large difference between values associated with the updated and initial assessment curves demonstrates that the updated assessments are the primary determinant of purchasing behavior. Therefore test drives provide an effective mechanism for sorting out buyers and non-buyers.

Conclusion

This paper examined the search process for buying a particular car. We model this process in three stages: determining the duration of the test drive, transition of customer assessment due to the test drive, and the customer's decision to purchase the car or not. We show that people update their priors given the information they acquire through a test drive, and that the posterior plays an important role in the final purchasing decision. The test drive acts as both an updating and sorting mechanism in which many of the individuals with more ambiguous (indefinite) prior assessments move to more definite posterior assessments. The changes in preferences are modeled here as a Markov chain. Customers with more experience with the product will be more confident of their assessment and therefore will be less likely to change it and have a lower sufficiency threshold for purchasing the product.

Although the length of the test drive does not affect the change in customer assessments, it does appear to somewhat increase customer likelihood of purchasing the car. This effect on purchases may be interpreted either as customers with longer test drives being more confident in their assessments (the same argument attributed to customers with previous experience) or as customers with long test drives feeling more compelled to purchase the product because of positive goodwill attributed to the seller because of their "generosity" in offering long test drives.

The analysis used maximum entropy to estimate the Markovian transition matrices. This new use of maximum entropy allows us to easily estimate a large number of parameters in a non-stationary Markov process with a limited sample size. There are several managerial implications from this paper. First, prior assessments can be deceiving. There is significant informational value to test drives that make people change their assessments, leading many customers with relatively low initial assessments to increase their assessment and purchase the product. Second, increasing the length of the test drive will not necessarily affect the customer's final assessment of the car because people may discover negative attributes. Nevertheless, longer test drive may increase final sales, either because consumers are more confident of their final assessment and are therefore willing to accept a lower threshold for purchase, or because of increased goodwill that can arise from long test drives that encourage the consumer to buy from the dealer. Third, people with prior product knowledge are the best target group since they self-select (those with very low assessments will not visit the dealer), and the ones that do show up at the showroom require less persuasion (they need less demonstration time) and are more likely to buy than others with the same posterior assessment level.

Our study considers the impacts of test drives on the purchase of one brand. It would be useful to analyze the pre-purchase search process and impact of test drives when considering several brands. We also need to further investigate how the effects of test drives depend on other marketing tools and on the socioeconomic characteristics of the customers.

References

Aaker, David A. "Building a Brand: the Saturn Story." *California Management Review*, (Winter 1994): 114-133.

Anderson, Erin, William T. Ross, Jr., and Barton Weitz. "Commitment and Its Consequences in the American Agency System of Selling Insurance." *Journal of Risk & Insurance,* Vol. 65, No. 4 (Dec. 1998): 637-669.

Bass, Frank M. "A New Product Growth Model for Consumer Durables." *Management Science*, 15 (1969): 215-227.

Blattberg, Robert C., and Subrata K. Sen. "A Bayesian Technique to Discriminate Between Stochastic Models of Brand Choice." *Management Science*, Vol. 21, No. 5 (Feb. 1975). 682-696.

Dockner, Engelbert and Steffen Jorgensen. "Optimal Advertising Policies for Diffusion Models of New Product Innovation in Monopolistic Situation," *Marketing Science*, 34 (1988): 119-130.

Erdem, Tulin and Michael P. Keane. "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science*, 15(1), 1996, 1-20.

Fazio, Russell H. and Mark P. Zanna. "Attitudinal Qualities Relating to the Strength of the Attitude-Behavior Relationship." Journal of Experimental Social Psychology, 14 (1978): 398-408.

Feichtinger, Gustav, Richard F. Hartl and Suresh P. Sethi. "Dynamic Optimal Control Models in Advertising: Recent Developments," *Management Science*, 40(2) (February 1994): 195-226.

Ferber, Robert, and Robert A. Piskie. "Subjective Probabilities and Buying Intentions." *The Review of Economics and Statistics* 47(3) August 1965: 322-328.

Freedman, J. L., and S. C. Fraser. "Compliance Without Pressure: The Foot in the Door Technique," *Journal of Personality and Social Psychology*, Vol. 18 (April, 1966): 195-202.

Givon, Moshe. "Variety Seeking Through Brand Switching." *Marketing Science*, Winter 1984, 3(1): 1-22.

Golan, Amos, George Judge, and Douglas Miller. *Maximum Entropy Econometrics: Robust Estimation With Limited Data*. New York: John Wiley & Sons, 1996.

Golan, A., G. G. Judge, and J. Perloff, "A Generalized Maximum Entropy Approach to Recovering Information From Multinomial Response Data." *Journal of the American Statistical Association*, Vol. 91 (1996): 841-853.

Graham, John R. "Twenty-Four Ways to Increase Your Sales." *Agency Sales Magazine*, Vol. 25, No. 10 (Oct. 1995): 26-29.

Green, William H. *Econometric Analysis*, 2nd edition. New York: Macmillan Publishing Company, 1993.

Hagerty, M.R and David A. Aaker. "A Normative Model of Consumer Information Processing." *Marketing Science*, 3 (1984): 227-246.

Heiman, Amir, Bruce McWilliams, and David Zilberman. "Demonstrations and Money-Back Guarantees: Market Mechanisms to Reduce Uncertainty" *Journal of Business Research*, (2001 forthcoming).

Heiman, Amir, and Eitan Muller. "Using Demonstration in New Product Diffusion—The Theory of Demonstration Duration," *Journal of Marketing Research*, Vol. 33, No. 4 (1996): 422-430.

Hogarth, Robin M., and Hillel J. Einhorn. "Order Effects in Belief Updating: The Belief-Adjustment Model." *Cognitive Psychology*, 24 (1992): 1-55.

Jaynes, E. T. "Information Theory and Statistical Mechanics," *Physics Review*, Vol. 106 (1957a): 620-630.

Jaynes, E. T. "Information Theory and Statistical Mechanics, II." *Physics Review*, Vol. 108 (1957b): 171-190.

Jaynes, E. T. "Information Theory and Statistical Mechanics." In K. W. Ford (ed.) *Statistical Physic.* New York: W. A. Benjamin, (1963): 181-218.

Johnson, John. "Excellence in Distribution - Dolin Supply Co.: Dedicated to Change," *Industrial Distribution*, 84 (12) (December, 1995): 28-29.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl, and T. C. Lee. *Introduction to The Theory and Practice of Econometrics*, 2nd edition. New York: John Wiley & Sons, 1988.

Juster, F. Thomas. "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design." *Journal of the American Statistical Association*, Vol. 61, September, 1966.

Kalish, Shlomo. "A New Product Adoption Model with Price, Advertising and Uncertainty," *Management Science*, 31 (1985): 1569-1585.

Kotler, Philip. *Marketing Management*, 8th edition. Englewoods Cliff, New Jersey: Prentice-Hall, Inc., 1994.

Kullback, J. *Information Theory and Statistics*. New York: John Wiley, 1959.

Lee, T. C., G. G. Judge, and A. Zellner. *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*, North-Holland, 1970.

Leland, J. W. "Generalized Similarity Judgments: An Alternative explanation for Choice Anomalies." *Journal of Risk and Uncertainty*, 9 (October 1994): 151-172.

Levine, R. D. "An Information Theoretical Approach to Inversion Problem." *Journal of Physics*, (1980) A 13, 91-108.

Little, John D. C. "Aggregate Advertising Models: The State of the Art," *Operations Research*, 27(4) (July-August 1979): 629-667.

Meyer, Robert J. and Arvind Sathi. "A Multiattribute Model of Consumer Choice During Product Learning," *Marketing Science*, 4 (1985): 41-61.

Norton, John A. and Frank M. Bass. "A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High Technology Products," *Management Science*, 33 (1987): 1069-1086.

Roberts, John H. and Glen L. Urban. "Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice," *Management Science*, 34 (1988): 167-185.

Scott, Carol A. "The Effects of Trial and Incentives on Repeat Purchase Behavior," *Journal of Marketing Research*, Vol. 13 (August 1976):263- 269.

Shannon, C. E. "A Mathematical Theory of Communication." *Bell System Technical Journal*, Vol. 27 (1948):379-423.

Smith, Robert E., and William R. Swinyard. "Attribute-Behavior Consistency: The Impact of Product Trial Versus Advertising," *Journal of Marketing Research*, Vol. 20, No. 3 (August, 1983):257-287.

Soofi, E. "Information Theory and Bayesian Statistics," in *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, edited by D. A. Berry, K. M. Chaloner, and J. K. Gewek. New York: John Wiley & Sons, Inc., 1996, pp. 179-189.

Stoneman, Paul L. "Intra-firm Diffusion, Bayesian Learning and Profitability." *The Economic Journal* 91 (1981): 375-388.

Stutzer, Michael. "A Simple Nonparametric Approach to Derivative Security Valuation," *Journal of Finance*, Vol. LI, No. 5, December, 1996.

Urban, Glen L., John R. Hauser, and John H. Roberts. "Prelaunch Forecasting of New Automobiles," *Management Science*, 36 (4) (1990): 401-21.

Viscusi, W. Kip. "Alarmist Decisions with Divergent Risk Information," *The Economic Journal*, 107, November 1997, 1657-1670.

Wold, H. *Demand Analysis*. New York: John Wiley and Sons, Inc., 1953.

Zellner, A. "Optimal Information Processing and Bayes's Theorem," *American Statistician*, Vol. 42 (1988):278-284.

Zufryden, Fred S. "Multibrand Transition Probabilities as a Function of Explanatory Variables: Estimation by a Least-Squares-Based Approach," *Journal of Marketing Research*, May 1986, 23(2): 177-183.

Appendix 1

Maximum Entropy

The idea of entropy is briefly described as follows. Suppose the occurrence of an event has N possible outcomes, $x_n, n = 1,...N$, with probabilities p_n such that $\sum_{n=1}^{N} P_n = 1$. Entropy is defined as

$$H = -\sum_{n=1}^{N} p_n \ln(p_n) \quad (1)$$

where $0 \cdot ln(0)$ is defined as zero. It is easy to show that entropy is maximized when there is perfect uncertainty (uniform distribution with $p_1 = ...p_N = 1/N$), with $H = -ln\left(\frac{1}{N}\right) = ln(N)$, and entropy is minimized when there is perfect certainty (one of the probabilities is 1 and the others are zero), with $H = 0$.

ME provides an optimal processing of information and is consistent with Bayesian theory. Zellner (1988) and Soofi (1996) developed an information-theory-based derivation of Baye's Theorem that provides a link between ME procedures and Baye's Theorem.

Shannon (1948) showed that maximum entropy is equivalent to maximizing the number of ways that an outcome can occur. More intuitively, maximum entropy is the same as minimizing the Kullback and Levine distance information measure (Kullback, 1959; Levine, 1980).

How Entropy Works: The Dice Problem

To illustrate how maximum entropy (ME) works, the authors introduce a well-known dice example proposed by Jaynes (1963) in his Brandeis lectures. Consider a die with six faces, and the probability of landing on each side is $p_1, \cdots, p_6$. Suppose the die is rolled many times, and the average number of dots shown is, say, *y*, $1 \le y \le 6$. Suppose the object is to recover the underlying six probabilities given this information. The problem is clearly infeasible under traditional statistical analysis because the number of parameters to be recovered (6) exceeds the two pieces of information: the mean of distribution equal to *y* and the sum of the six probabilities equal to 1. Unless *y* equals 1 or 6, there are generally an infinite numbers of distributions supported on {1, 2, …, 6} such that the mean is *y*.

To solve this problem, Jaynes proposed using the ME approach, which chooses the distribution that maximizes the entropy and is consistent with the average moment constraint

$$\text{Max} \quad -\sum_{n=1}^{6} P_n \ln \left(p_n\right)$$

subject to

$$1p_1 + 2p_2 + 3p + 4p_4 + 5p_5 + 6p_6 = y$$

$$\sum_{n=1}^{6} p_n = 1.$$

Essentially, the ME finds the distribution $p_1, \ldots, p_6$ that is closest to the uniform distribution measured by the K-L distance subject to the data and adding-up constraints. In that sense, the ME estimator is robust, conservative, and works well even when the number of parameters exceeds data points. When *y* is 1.0, ME yields a solution $p_1, \ldots, p_6$ = (1, 0, 0, 0, 0, 0). When *y* = 4.0, the solution is (0.103, 0.123, 0.146, 0.174, 0.207, 0.247).

Appendix 2

The Maximum Entropy Model for Estimating the Markov Transition Matrix

The objective of this estimation is to maximize the joint entropy subject to the Markov chain equations (2) and (3). The A_{nm}'s are transition probabilities that can be estimated directly. Because the arguments of the entropy measures (matrices B and C) are probabilities, the coefficients are reparameterized to be proper probability distributions defined over some support. For example, for each reduced-form coefficient, b_j, first a support space is chosen, which is a set of discrete points $Z = (z_1, z_2, ... z_S)$ of dimension $S \geq 2$, at uniform intervals that span the possible range of the unknown coefficients. The natural bound of probability allows us to determine that the support range is between –1 and 1. In this case three values for z are chosen: $z_1 = -1$, $z_2 = 0$, and $z_3 = 1$. We adopt $S = 3$ since previous studies have shown that increasing the support space to $S = 5$ has little effect on the estimates (Golan, Judge, and Miller, 1996).

After reparameterizing matrices B and C and the error terms, the joint entropy of parameter and noise space are maximized:

(7)

$$\max -\left(\sum_{n,m=1}^{4} A_{nm} \ln A_{nm} + \sum_{n,m=1}^{4} \sum_{s=1}^{3} P_{nms}^{B} \ln P_{nms}^{B} + \sum_{n,m=1}^{4} \sum_{s=1}^{3} P_{nms}^{C} \ln P_{nms}^{C} + \sum_{t=1}^{T} \sum_{n=1}^{4} \sum_{s=1}^{3} w_{tns} \ln w_{tns} \right)$$

subject to data constraint:

(8)

$$\begin{aligned} Q_1 &= PQ_0 + VW \\ P &= A + P^B Z \; ln\,(D) + P^C Z\, E \end{aligned}$$

and the adding-up constraints:

(9) $\sum_{n=1}^{4} A_{nm} = 1, \sum_{n=1}^{4} P_{nm}^{B} = 0, \sum_{n=1}^{4} P_{nm}^{c} = 0,$ for $m = 1, ..., 4,$ and $\sum_{s=1}^{3} w_{tns} = 1,$ for all t and i.

$B_{n,m} = \sum_{s=1}^{s} P_{nms}^{B} z_s$ and $C_{nm} = \sum_{s=1}^{S} P_{nms}^{C} z_s$. P_{nms}^{B} and P_{nms}^{C} are probabilities that correspond to the *S*-dimensional support vectors for each element in matrices *B* and *C*. Obviously, these satisfy the adding-up constraint that the sum of the probabilities equal 1:

$$\sum_{s=1}^{S} P_{nms}^{B} = 1 \text{ and } \sum_{s=1}^{S} P_{nms}^{C} = 1 \ .$$

The error term ε can be similarly reparameterized as $\varepsilon_n^j = \sum_{s=1}^{S} w_{ns}^j v_s$, where *V* is a support space of dimension *S*. Again, since the probabilities naturally lie between zero and one, there is a natural support space *V* that lies between –1 and 1, and *S* = 3: $V = (v_1, v_2, v_3) = (-1, 0, 1)$.

The objective function (7) is maximized subject to data constraint (8) and adding-up constraint (9). This maximization has a negative semi-definite Hessian matrix and yields a globally unique solution for these probabilities. The estimated coefficients for *B* and *C*, and the residuals can be calculated as:

$$\hat{B}_{nm} = \sum_{s=1}^{S} \hat{P}_{nms}^{B} z_s, \ \hat{C}_{nm} = \sum_{s=1}^{S} \hat{P}_{nms}^{C} z_s, \text{ and } \varepsilon_n^j = \sum_{s=1}^{S} \hat{w}_{ns}^j \ v_s \ .$$

The estimated variance-covariance matrix for the ME estimates are based upon large sample theory of maximum entropy estimators (see Appendix 3 for a detailed derivation of the variance-covariance estimates).

Appendix 3

Derivation of Variance-Covariance Matrix and Hypothesis Testing

Under four mild conditions, the ME estimator is consistent, asymptotically normal, and efficient. These four conditions are (1) the error support is symmetric around zero, (2) the supports span the true value for each one of the unknown parameters, (3) the error terms are independently and identically distributed, and (4) the operator *X* is of full rank. Proofs for the consistent and asymptotic normality properties can be found in Golan, Judge, and Miller (1996).

The asymptotic properties of the ME estimator are used to approximate its finite sample estimated variances. After inserting equation (3) into (2), equation (2) becomes a four-equation system and can be rewritten as

$$Q_1 = X\beta + \varepsilon$$

where $X = (Q_o, \ln(D)Q_0, EQ_0)$ [8] and $\beta = (A, B, C)$. The (n,m) elements of the variance-covariance matrix Ω for the four-error terms are estimated as

$$\Omega_{nm} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{nt}\hat{\varepsilon}_{mt}$$

where *m*, *n* = 1,…, 4. Because our model is essentially comprised of equations with linear, adding-up restrictions, the variance-covariance of the ME estimated parameters are

$$V\,\text{cov}(\beta) = C - CR'(RCR')^{-1} RC$$

where $C = (S'\Omega^{-1}S)^{-1}$, *S* is the Kronecker product of *X*, and *I*(4 by 4), *R* is the matrix showing linear restrictions imposed on the parameters (A, B, and C).

---

[8] Since the first column of *E* is perfectly multi-collinear with the intercept, this column is deleted when $V\,\text{cov}(\beta)$ is derived and hypothesis testing is conducted.

To test the hypothesis H_0 : $R\beta = r$ against the alternative hypothesis H_1 : $R\beta \neq r$, we note that since the ME estimates of β follow a normal distribution in the large sample, the following statistic

$$g = \left(R\beta - r\right)' \left(RCR'\right)^{-1} \left(R\beta - r\right)$$

follows a chi-squared distribution with *J* degrees of freedom, where *J* is the number of linear restrictions imposed on the betas under the null hypothesis. For example, when the null hypothesis is that there is no test drive effect (*A* is an identity matrix), then *J* = 12.