BEREND WIERENGA*

The author presents a new method for estimating the parameters of the linear learning model. The procedure, essentially a least squares method, is easy to carry out and avoids certain difficulties of earlier estimation procedures. Applications to three different data sets are reported, as well as results from a goodness-of-fit test. A simulation study was carried out to validate the method. The outcomes are compared with those obtained from the minimum chi square estimation method. The results of the new method appear to be satisfactory.

A Least Squares Estimation Method for the Linear Learning Model

INTRODUCTION

The linear learning method (LLM) has proved to be a successful tool for modeling consumer choice phenomena. A broad spectrum of applications of this model, with respect to both brand choice and store choice, has been reported in the marketing literature [see e.g. 1, 3, 5, 8–11, 13]. The author presents an iterative least squares method for estimating the parameters of the LLM. In this regression procedure individual consumers represent data points. Their purchase histories constitute the independent variables and actual brand choices are the dependent variables.

First a brief description of the LLM is given. Then the new estimation method is described in some detail. Reports are presented on empirical estimation obtained with the new method for different data sets. Also a goodness-of-fit test is applied to the data. The results of a simulation study, carried out to validate the new estimation method, are given, and the consistency of the procedure is discussed. For the data sets, the LLM paramters were estimated also by the minimum chi-square estimation method developed by Massy et al. [10, Chapter 5]. These results are compared with those from the least squares method and some general theoretical and practical aspects of both procedures are discussed.

*Berend Wierenga is a senior lecturer of the Marketing Department of the Wageningen Agricultural University, The Netherlands, who visited Stanford University during 1974/75. This research was supported by grants from Ahold N.V. (the "Albert Heyn Studie Beurs" was awarded to the author) and the Fulbright-Hays Program. Useful comments on a draft of the article by Prof. Srinivasan and doctoral student Shelby McIntyre, both of Stanford University, and comments from *JMR* referees are gratefully acknowledged. Jaap Bijkerk carried out part of the computer work for this article.

LINEAR LEARNING MODEL

The linear learning model for brand choice processes is defined by the following operators [see e.g. 10, Chapter 5; 13, Chapters 3–5]:

(1) $p = \alpha + \beta + \lambda p_{t+1}$ (purchase operator)

and

(2) $p = \alpha + \lambda p_{t-1}$ (rejection operator).

The market is assumed to contain only two brands, indicated here as brand 1 and brand 0. p_t stands for the probability that brand 1 is chosen at purchase occasion t. After a purchase, the probability of a consumer purchasing brand 1 is transformed according to equation 1 if brand 1 is chosen; otherwise the transformation is according to equation 2.

The parameters α, β, and λ are non-negative and because the p_t are probabilities the constraint

(3) $(\alpha + \beta + \lambda) \leq 1$

must hold.

The LLM can also be formulated as follows. Let $\{X_t\}$ denote the (stochastic) brand choice process. X_t can only take the values 1 and 0, corresponding with a brand 1 and a brand 0 purchase, respectively.

Journal of Marketing Research
Vol. XV (February 1978), 145–53

Then:

(4) $p_t = \text{Prob}(X_t = 1 | x_{t-1}, p_{t-1}) = \alpha + \beta x_{t-1} + \lambda p_{t-1}$

(x_{t-1} is the purchase made at $(t - 1)$, which is also either 1 or 0).

Thus equation 4 is a complete definition of the LLM and includes equations 1 and 2. Now according to equation 4:

$$p_{t-1} = \alpha + \beta x_{t-2} + \lambda p_{t-2}. \tag{5}$$

Substituting this in equation 4 gives:

$$p_t = \alpha(1+\lambda) + \beta(x_{t-1} + \lambda x_{t-2}) + \lambda^2 p_{t-2}. \tag{6}$$

Repeating this type of substitution $(k - 1)$ times gives:

$$p_t = \alpha \sum_{j=0}^{k-1} \lambda^j + \beta \sum_{j=0}^{k-1} \lambda^j x_{t-j-1} + \lambda^k p_{t-k}. \tag{7}$$

Equation 7 plays a major role in the estimation method to follow.

LEAST SQUARES ESTIMATION METHOD

Basic Approach

Because consumers generally cannot give their current or even past probability of a brand 1 purchase, the LLM parameters cannot be estimated directly from equation 4. Therefore an indirect way is followed that has equation 7 as its starting point. In equation 7 the probability of a brand 1 at purchase occasion t is expressed as a function (with parameters α, b, and λ) of (1) the purchase history during the k most recent purchases and (2) the probability of brand 1 purchase at $(t - k)$. Because $0 \le \lambda \le 1$ the effect of the latter part (the third term on the right side of equation 7) becomes smaller and smaller as k increases. Only in the extreme case of no learning at all ($\lambda = 1$) is this phenomenon not true. In applications of the LLM the values found for λ are mostly within the range .3 to .7 [1; 10, Chapter 5; 13, Chapter 4]. Now when $k = 10$ for example, one has $(0.3)^k = 0.000$ and $(0.7)^k = 0.027$. To obtain the contribution of the last term in equation 7, λ^k has still to be multiplied by p_{t-k}, a number between zero and one, which makes it even smaller. So for k that does not have too small a value, the last term in equation 7 is always very small.

Therefore in this section this term is omitted:

$$p_t = \alpha \sum_{j=0}^{k-1} \lambda^j + \beta \sum_{j=0}^{k-1} \lambda^j x_{t-j-1}. \tag{8}$$

The influence of the remote past of a consumer's purchase history on his current purchase probability is neglected. (In the next section the effect of this remote past is explicitly included.)

After defining

$$v_k(\lambda) = \alpha \sum_{j=0}^{k-1} \lambda^j \tag{9}$$

and

$$w_{kt}(\lambda) = \sum_{j=0}^{k-1} \lambda^j x_{t-j-1} \tag{10}$$

one can write equation 8 as:

$$p_t = v_k(\lambda) + \beta w_{kt}(\lambda). \tag{11}$$

Now for the time being the value of λ is fixed.

Suppose one has the purchase histories of a sample of N consumers, referring to their last $(k + 1)$ purchase. Let the index t of equation 11 correspond with the most recent purchase. Now a consumer's sequence of purchase at $(t - 1)$, $(t - 2)$, ..., $(t - k)$ can be conceived of as his purchase history x_{t-1}, x_{t-2}, ..., x_{t-k}. His most recent purchase at t is the currently observed brand choice. Let the latter purchase be indicated as y. Of course y is either 1 or 0. Different consumers mostly have different purchase histories and for each consumer a value of $w_{kt}(\lambda)$ can be computed with equation 10. So for each consumer i one then has two data points: y_i and $w_{kt}(\lambda)$, where $w_{kt}(\lambda)$ depends on i. Deleting the subscripts k and t because they are the same for all consumers and remembering that v and w depend on λ, one can write equation 11 for the individual consumer i as:

$$p_i = v + \beta w_i.$$

Now:

$$Ey_i = p_i$$

so that

$$Ey_i = v + \beta w_i.$$

This is equivalent to:

$$y_i = v + \beta w_i + u_i \tag{12}$$

where $Eu_i = 0$.

So with equation 12 one has a regression model in two variables.

Given a series of pairs of observations (y_i, w_i), v and β can be estimated by the well-known method of least squares [see e.g. 7] in which the sum of the squared deviations between p_i and y_i is minimized. Of course α can be derived directly from v with equation 9.

So far the value of λ has been fixed so that the squared deviations are minimized for a given value of λ. Now the least squares procedure is repeated for different values of λ, until that λ has been found for which the overall sum of squared deviations is as small as possible. This is also the λ for which the coefficient R^2, referring to equation 12, is maximum. Besides estimates for α, β, and λ, the least squares procedure also gives the value for R^2 and estimates for the standard deviation of α and β.[1]

Bayesian Treatment of Remote Purchase History

So far the effect of purchases made more than k purchase occasions ago has been neglected. Theoretically the error made by omitting the term on the extreme right of equation 7 can be made as small as desired by making k large enough. However, for large values of λ (i.e., large carryover effect), the length of the empirical purchase histories required can become too great for practical estimation. Moreover, the value of λ is not known in advance. Therefore in this section the effect of the remote past is taken into account when k is not that big.

Assume that the procedure of the last section has been carried out for a data set, where a particular value for k is chosen. Then the estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$ are available. Now consider a consumer i with a certain purchase history h_i. (For $k = 10$, 1,024 different purchase histories are possible). Then given this purchase history and the parameter values $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$, what can be said about this consumer's probability of a brand 1 purchase k purchase occasions ago, i.e. his p_{t-k}?

The probability of a certain value for this chance is given by Bayes' Theorem:

$$\text{(13)} \qquad \operatorname{Prob}(p_{t-k} \mid h_i) = \frac{\operatorname{Prob}(h_i \mid p_{t-k}) \operatorname{Prob}(p_{t-k})}{\operatorname{Prob}(h_i)}.$$

If one assumes that p_{t-k} can only take r different values ranging from m_1 to m_r, equation 13 can be written for a particular value, viz. m_l, as:

$$\text{(14)} \qquad \operatorname{Prob}(p_{t-k} = m_l \mid h_i) = \frac{\operatorname{Prob}(h_i \mid p_{t-k} = m_l) \operatorname{Prob}(p_{t-k} = m_l)}{\sum_{d=1}^{r} \operatorname{Prob}(h_i \mid p_{t-k} = m_d) \operatorname{Prob}(p_{t-k} = m_d)}.$$

The *likelihood* of purchase history h_i when $p_{t-k} = m_l$, i.e.,

$$\operatorname{Prob}(h_i \mid p_{t-k} = m_l)$$

for the LLM-parameter values $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$ can be computed directly by use of the definition of the LLM. This computation can be done for all r possible values of p_{t-k}. Suppose no prior information so that all values of p_{t-k} are *a priori* equally likely, i.e.:

$$\text{(15)} \qquad \operatorname{Prob}(p_{t-k} = m_l) = \frac{1}{r} \; (l = 1, \ldots, r).$$

Now all quantities in equation 14 are known and the *posterior* probabilities for the p_{t-k}-values—given purchase history h_i—can be determined. In this way a posterior distribution over the r values of p_{t-k} can be obtained for each consumer.

In the present case it is not necessary to represent the distribution of p_{t-k} very finely, i.e., with a large number of different values. Because in equation 7 λ^k is small, the contribution of the term $\lambda^k p_{t-k}$ is sufficiently determined when only the first digit of p_{t-k} is known. Therefore take $r = 10$ and split up the (0, 1) interval (the range of p_{t-k}) in 10 parts of equal size, each represented by its midpoint: 0.05, 0.15, ..., 0.95. For each of these values the *a priori* probability is 0.10.

The Complete Method

In the complete method, iterative regression neglecting the effect of the remote past is carried out first. This step produces initial estimates for α, β, and λ. Then the p_{t-k} distributions are updated in the Bayesian way. From the updated p_{t-k} distribution for each consumer the modus is taken as the value to be inserted in equation 7. When defining:

$$\text{(16)} \qquad z_{kt}(\lambda) = \lambda^k p_{t-k},$$

equation 7 can be written as:

$$\text{(17)} \qquad p_t = v_k(\lambda) + \beta w_{kt}(\lambda) + z_{kt}(\lambda),$$

which is the same as equation 11 but with the effect of the remote past taken into account. For given λ, $z_{kt}(\lambda)$ is known for each consumer. Deleting the subscripts k and t, one obtains from equation 17 for individual consumer i:

$$\text{(18)} \qquad p_i = v + \beta w_i + z_i.$$

Now defining:

$$y_i^* = y_i - z_i,$$

estimates of v and β can be obtained by means of least squares regression of y^* on w. This process is repeated for a new value of λ, etc. So the same iterative procedure as before is used to obtain new estimates for α, b, and λ, the only difference being that y is replaced by y^*. When the new estimates are close to the old ones, the procedure stops. Otherwise the p_{t-k} distribution is updated once again and the estimation is repeated. So there are two iteration cycles. In the first one the initial p-distribution is varied, and in the second one the LLM-parameter λ takes different values. The second cycle is carried out within each step of the first one.

---

[1]Because the dependent variables are probabilities, generalized least squares could be used instead of ordinary least squares. The author decided not to do this because (1) the true p-values with which the weighting must be done are not known and one would have to use estimates instead; (2) consumers with p near one or zero would get a heavier weight than consumers with less extreme p values and this would imply that loyal consumers (either to brand 1 or 0) contribute more to the estimation results than consumers who switch more frequently.

Table 1
LEAST SQUARES ESTIMATION RESULTS FOR THREE FOOD PRODUCTS

Product	*Brand 1*	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\sigma}_{\hat{\alpha}}$	$\hat{\sigma}_{\hat{\beta}}$	$\Sigma e_i^2/n$	R^2	*n*
Beer	B1	.009	.242	.745	.004	.007	.075	.675	613
	B2	.004	.251	.717	.003	.009	.050	.579	613
	B3	.000	.234	.757	.002	.005	.021	.788	613
	B4	.005	.322	.653	.004	.008	.058	.714	613
Fopro	F1	.018	.586	.384	.006	.010	.038	.840	666
Marg	M1	.002	.243	.755	.002	.004	.031	.835	850

EMPIRICAL RESULTS

The estimation method was applied to brand choice data for three frequently purchased consumer products in the Netherlands: beer, fopro (a pseudonym for a food product), and margarine. The purchases were made during 1967 and 1968 by members of the Dutch Attwood Consumer panel. The numbers of households of which the brand choice data were used are 613, 666, and 850 for beer, fopro, and margarine, respectively. More details about the data can be found in [13]. For each product the biggest brand in the market was called brand 1. For beer, fopro, and margarine these are indicated as B1, F1, and M1, respectively. All other brands in a market are designated as brand 0. The beer data also were examined with three other brands in turn indicated as brand 1. These brands are called B2, B3, and B4, respectively.

For every household the first 10 purchases made in the reporting period are considered to be the purchase history; the eleventh purchase constitutes the subsequently observed brand choice (the y_i of the previous sections). So k is taken to be 10 here.

In Table 1 the estimates obtained for α, b, and λ, corresponding standard deviations, the mean square deviation ($\Sigma e_i^2/n$), and R^2 are presented. Table 2 shows some additional statistics: the correlation coefficient of $\hat{\alpha}$ and $\hat{\beta}$; the ratios coefficient/standard deviation for $\hat{\alpha}$ and $\hat{\beta}$ and lower and upper limits for p.

To give an idea about the effect of taking into account the remote part of the purchase history, Table 3 shows the parameter estimates for the first and last iteration rounds described in the previous section. An iteration was terminated when the third digit of $\hat{\lambda}$ did not change from one iteration to the next.

To illustrate the dependence of the sum of squared deviations (Σe_i^2) on λ within an iteration round, Figure 1 depicts the relationship between these two quantities for the first iteration round of beer, B1. The pattern in Figure 1 is typical for all other cases in that Σe_i^2 always had an unambiguous lowest point without problems of local minima.

Comments

1. The estimates found for the LLM parameters α, β, and λ look reasonable and are in agreement with restriction (3). If this were not so, the fit of the LLM would be questionable here. In the LLM the range of values of p (the probability of a brand 1 purchase) is limited. The lower limit is $p_L = \alpha/(1-\lambda)$; the upper limit is $p_U = (\alpha+\beta)/(1-\lambda)$. In Table 2 these limits are given for the parameter values of Table 1. These boundary values look reasonable—p becomes very small after a great many brand 0 purchases and very large after many brand 1 purchases. The probability of a brand 1 purchase never becomes exactly equal to 0 or 1, however. Theoretically, when there are more than two different brands in a market, the slope λ of the LLM should be equal for all brands [see 6]. For beer this condition can be checked. It appears that the estimated λs are rather close together. They range from 0.6528 to 0.7565, and three of them are within a range of 0.04. The value for B4 is the most extreme, probably because B4, unlike all other brands distinguished, is not a real brand in the market, but a conglomerate—i.e., defined as all brands that are not B1, B2, or B3.
2. The standard deviations for the $\hat{\beta}$s are relatively small. For the $\hat{\alpha}$s they are large, especially when $\hat{\alpha}$ is small. This result has implications for the accuracy of long-term market shares predictions.

Table 2
ADDITIONAL STATISTICS WITH RESPECT TO THE LLM PARAMETERS, OBTAINED BY LEAST SQUARES METHOD

Product	*Brand 1*	$\hat{\rho}(\hat{\alpha},\hat{\beta})$	$\hat{\alpha}/\hat{\sigma}_{\alpha}$	$\hat{\beta}/\hat{\sigma}_{\beta}$	P_L	P_U
Beer	B1	−.70	2.04	35.66	.033	.980
	B2	−.43	1.39	28.98	.014	.901
	B3	−.38	.21	47.61	.002	.961
	B4	−.59	1.10	39.07	.013	.941
Fopro	F1	−.62	2.99	59.09	.029	.980
Marg	M1	−.55	.89	65.56	.007	.999

Table 3
PARAMETER ESTIMATES FOR FIRST AND LAST ITERATION ROUNDS

		First iteration round			*Last iteration round*			*Number of rounds required*
Product	*Brand 1*	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	
Beer	B1	.009	.238	.765	.009	.242	.745	3
	B2	.004	.240	.743	.004	.251	.717	3
	B3	.001	.234	.774	.000	.234	.757	3
	B4	.005	.329	.649	.005	.322	.653	3
Fopro	F1	.018	.586	.384	.018	.586	.384	2
Marg	M1	.002	.225	.796	.002	.243	.755	3

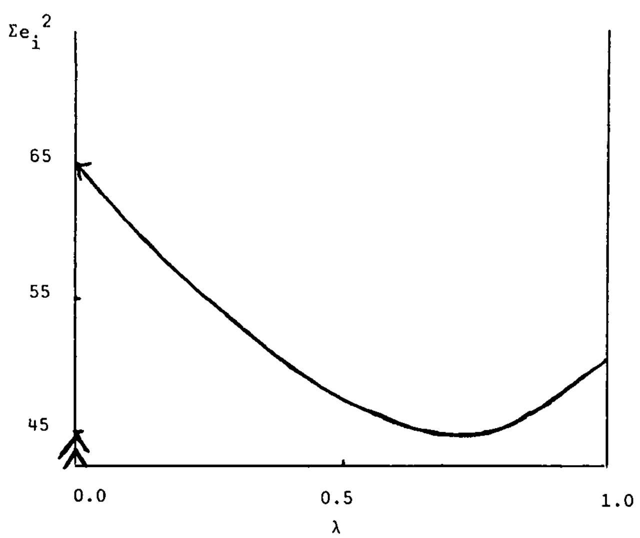
As shown in [10, p. 148] the expression for the long-term market share of brand 1 is:

$$M = \frac{\alpha}{1 - \beta - \lambda}.$$

Here the parameter α plays a crucial role and a large variance of $\hat{\alpha}$ causes a large variance of the predicted long-term market share. There is a negative covariance between $\hat{\alpha}$ and $\hat{\beta}$, clearly related to the tendency of $(\alpha + \beta + \lambda)$ to be near 1.

3. Table 3 shows that the effect of incorporating the remote part of the purchase history on the estimated values for α, b, and λ is not great for the data considered here. Sometimes in the first iteration round the restriction (3) is violated, however.
4. The values of R^2 are generally good. There is a clear influence of the purchase history on current brand choice. However, with the interpretation of R^2 caution is required. Here the value of R^2 is not only a function of the goodness of fit of the data to the LLM estimated, but also a function of the particular parameter values of this model itself. In general Σe_i^2 tends to be smaller and R^2 tends to be larger as the particular LLM allows the values of p to become closer to 0 and 1. (For the case in which p has a beta distribution, the upper bound for R^2 can be computed; see [12]). Therefore the comparison of R^2 values for different data sets does not provide a complete picture of the relative goodness of fit of the LLM.

Figure 1
DEPENDENCE OF Σe_i^2 ON λ FOR BEER, BRAND B1



A TEST FOR GOODNESS OF FIT

Once the LLM parameters are estimated, for each consumer i one can compute the predicted p_i, i.e., the probability of choosing brand 1, which can be compared with y_i, the actual outcome of the brand choice. (Here p_i and y_i have the meaning from the previous paragraphs). The hypothesis to be checked is:

$$\text{Prob}(y_i = 1) = p_i \, (i = 1, \ldots, N). \tag{19}$$

A test described by Cox [4, Section 4.4] is used to examine this hypothesis. The test consists of two parts and uses the logistic transform, extensively discussed by Cox. The first part is concerned with the question of whether the p_i are systematically too high or too low. The test statistic is:

$$T_1 = \Sigma y_i.$$

Under the null hypothesis "the p_i are neither too high or too low," the statistic

$$S_1 = \frac{T_1 - ET_1}{\sqrt{\text{Var}\, T_1}} \tag{20}$$

is distributed approximately normally.

The second part of the test is concerned with the question of whether the p_i are too clustered or too dispersed. Here is a test statistic T_2 is used, which after transformation into

$$S_2 = \frac{T_2 - ET_2}{\sqrt{\text{Var}\, T_2}}$$

is distributed approximately normally under the hypothesis: "the p_i are neither too clustered nor too dispersed." When the p_i are too dispersed, T_2 is smaller than expected under the null hypothesis; when the

p_i are too clustered, T_2 is larger. For the definition of T_2 and more information about these tests the reader is referred to Cox [4].

Results obtained with these tests for the author's data are presented in Table 4. From the low values in this table it appears that the p_i predicted by the LLM, using the least squares parameter estimates, perform very well. None of the p_i values is too low or too high, or too clustered or too dispersed.[2]

VALIDATION OF THE PROCEDURE AND CONSISTENCY OF THE ESTIMATORS

Simulation Study

To examine how well the least squares estimation method reproduces the true LLM parameters, a simulation study was run in which the parameters were estimated from purchase histories generated by LLM brand choice processes with known true parameters. Three different parameter sets were used, the "classic" Snow Crop case [10, p. 172–5] and the cases of B1 and F1.

First the situation of 500 households and a history of five purchases per household was considered. Then the effects of increasing the number of households to 1,000 and increasing the number of purchases to 10 were examined. The results for the estimated parameter values are summarized in Table 5. For each combination in this table 10 simulation runs (with different random starting values) were carried out. For individual households the probabilities of choosing brand 1 at the first purchase were obtained as drawings from beta distributions corresponding to the respective parameter sets (the μ_1 and μ_2 values given in [10, Table 5.5.] for Snow Crop and in [13, Table 4.23] for beer and fopro were used to determine the beta parameters). Table 5 shows that the reproduction of the original parameters is satisfactory. In the case of 500 households and five purchases per household the mean error ranges from .008 to .096.

The last line in Table 5 gives an impression about possible biases in the estimator. Because 30 simulations were run for each parameter set (10 for each *N-NN* combination), for an unbiased estimator one would expect 15 estimated parameter values to be under the true values. Table 5 shows that the observed numbers of underestimates are close to this expected number with the exception of $\hat{\alpha}$ for parameter set B. This value was systematically too low, probably because of the small true value of α in this parameter set (< 0.01).

Table 4

RESULTS OF COX TEST FOR GOODNESS OF FIT OF LLM, LEAST SQUARES ESTIMATED PARAMETERS

Product	*Brand 1*	S_1	S_2	*n*
Beer	B1	−.00	.00	613
	B2	−.02	−.13	613
	B3	−.03	−.46	613
	B4	−.01	−.05	613
Fopro	F1	−.01	−.04	666
Marg	M1	−.03	−.37	850

[2]The p_i used in the test are not given *a priori* but are computed from parameter estimates obtained from the data. Strictly speaking, the variance of the test statistic should be corrected for errors in these estimates. Because such a correction generally would increase this variance, the resulting values of S_1 and S_2 would become even smaller, thus strengthening the conclusion.

Consistency of the Estimators

In terms of the symbols of equation 12 the true relationship between current purchase and purchase history is:

$$y_i = \nu + \beta w_i + \lambda^k p^i_{t-k} + u^T_i \tag{21}$$

where u^T_i is the true error term for consumer i. In the first iteration round of the estimation procedure, however, the model is specified as:

$$y_i = \nu + \beta w_i + u_i \qquad (= \text{equation 12})$$

Thus the error term in this equation—u_i—is composed of the true error term and a term dependent on p_{t-k}, i.e.,

$$u_i = u^T_i + \lambda^k p^i_{t-k}.$$

Now according to equation 12

$$p^i_{t-k} = (x^i_{t-k} - u^i_{t-k}) \qquad (\text{because by definition } y^i_t = x^i_t),$$

and according to equation 10

$$w_i = \sum_{j=0}^{k-1} \lambda^j x^i_{t-j-1}.$$

So in equation 12 x_{t-k} appears both in w_i and in the error term u_i. Because for consistent estimators these quantities ought to be uncorrelated, the question is whether this dependence causes a consistency bias in the estimates obtained by the procedure proposed here. The following remarks can be made.

First, for $k \rightarrow \sim$ the discrepancy between equations 12 and 21 disappears, so the procedure is consistent in the sense that when the length of the observed purchase histories increases the estimates converge in probability to their true values. But increasing the number of households, keeping the length of the purchase history fixed, will not automatically remove the dependence. Second, even for short purchase histories the impact of the correlation between w_i and u_i can be assumed to be very small. The term x_{t-k} is only one of the components of w_i and in fact the component with the smallest weight (λ^{k-1}, see equation 10). Third, in the second round (and in all additional rounds) of the iteration procedure the term

Table 5
SIMULATION RESULTS FOR THREE PARAMETER SETS

	Parameter set A: *Snow Crop* $\alpha = .015$ $\beta = .305$ $\lambda = .612$			Parameter set B: *beer BI* $\alpha = .009$ $\beta = .242$ $\lambda = .745$			Parameter set C: *fopro FI* $\alpha = .018$ $\beta = .586$ $\lambda = .384$		
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$N = 5$									
$NN = 500$									
MES	.012	.328	.586	.008	.283	.704	.016	.597	.380
MER	.008	.045	.063	.009	.059	.067	.008	.087	.096
MSE	.008	.062	.083	.010	.080	.089	.009	.099	.107
$N = 5$									
$NN = 1{,}000$									
MES	.012	.313	.612	.004	.268	.727	.020	.620	.347
MER	.005	.025	.026	.006	.040	.044	.005	.067	.073
MSE	.006	.030	.033	.006	.048	.059	.006	.091	.101
$N = 10$									
$NN = 500$									
MES	.019	.321	.594	.007	.204	.787	.016	.600	.374
MER	.005	.021	.038	.002	.039	.043	.007	.081	.087
MSE	.006	.028	.042	.003	.053	.059	.007	.100	.111
Number of times estimated parameter value is equal to or below true value	15	12	16	22	16	13	15	14	15

N = length of observed purchase history.
NN = number of households in the simulated sample.
MES = mean estimate.
MER = mean error.
MSE = mean square error.
Number of simulation runs per combination = 10.

$\lambda^k p^i_{t-k}$ is restored. Although the p^i_{t-k}s used here are approximations of the true values, the consistency problem is considerably reduced.

Table 6 illustrates these points. For simulated LLM purchase histories (for which the true p^i_{t-k} values were known) the mean value of the product of error term and w_i was calculated for (1) the true error term u^T_i, (2) the error term u_i in the first iteration round, and (3) the error term u_i in the second iteration round.

The specific LLM parameters used in the simulation are the Snow Crop values: .015, .305, and .612. The effect of the dependence in the error term can be assessed by the extent to which the product $\Sigma u_i w_i$ differs more from zero in the case of dependent error terms than in the situation of the true error terms. (In the latter case the expected value is zero.)

As is clear from Table 6 the impact of the dependence is very small, even when the length of the purchase history is only five. Moreover, the dependence effect decreases quickly as N (the length of the purchase history) increases. Also the effect is much less severe in the second than in the first iteration round.

These results are in agreement with the good quality of the estimates in the simulation study. Therefore

Table 6
IMPACT OF CONSISTENCY BIAS FOR SIMULATED PURCHASE HISTORIES

N	NN	(1) $\left\lvert\dfrac{\Sigma u_i^T w_i}{NN}\right\rvert$	(2) $\left\lvert\dfrac{\Sigma u_i w_i}{NN}\right\rvert$ *first iteration round*	(3) $\left\lvert\dfrac{\Sigma u_i w_i}{NN}\right\rvert$ *second iteration round*	(2) − (1)	(3) − (1)
5	500	.0240	.0435	.0200	.0195	−.0040
10	500	.0119	.0136	.0115	.0017	−.0004
15	500	.0109	.0108	.0109	−.0001	.0000
5	1000	.0021	.0141	.0102	.0120	.0081
10	1000	.0053	.0066	.0047	.0013	.0006

the consistency issue is not much of a problem, although one should keep in mind that the purchase histories should not be taken too short. (As a guideline: five purchases still gave satisfactory results in this study.)

COMPARISON OF THE LEAST SQUARES METHOD WITH THE MINIMUM CHI SQUARE METHOD

The most prominent estimation method for the LLM parameter is the minimum chi square procedure developed by Massy et al. [10, Chapter 5]. It was used extensively in [10], and other authors have used the method subsequently [e.g. 1, 9, 11, 13]. Carman [3] described an estimation method which is essentially a regression procedure, but very different from the least squares method presented here. Carman's procedure does not take into account the fact that different consumers will have different p-values at the beginning of the observation period.

Haines [5] used various procedures to estimate the LLM parameters. However, he dealt with a simplified version of the LLM and also made the assumption that at the start of the observation period all consumers have the same p-value. Here only the minimum chi square method is considered further. This is the first method whereby—in an ingenious way—the heterogeneity of the consumer population with respect to p at the start of the observation period is taken fully into account. Each consumer is assumed to have his own initial p-value p_0, and p_0 has a distribution in the consumer population with density function $f(p_0)$ and first four moments μ_1, μ_2, μ_3, and μ_4.

The input for the procedure consists of the relative frequencies with which the 16 different purchase sequences of length four occur in observed purchase histories. According to the theory of the LLM the expected values of these 16 relative frequencies are a function of the LLM parameters α, β, and λ and the four moments of the p_0 distribution μ_1 to μ_4. The procedure searches for those values of α, β, λ, μ_1, μ_2, μ_3, and μ_4 that minimize the differences between observed and expected relative frequencies in a chi square sense. The method is described in detail in [10, Chapter 5].

In the following discussion the minimum chi square method (MCS) is compared with the least squares (LS) procedure.

In [13] the LLM parameters for the beer, fopro, and margarine data which were used here to apply the LS method have been estimated by the minimum chi square method, where the first 10 purchases of each household made in the reporting period are used as observations. The resulting estimates of the LLM parameters (indicated by a ~, in contrast to the LS estimates which are indicated by a ^) are reproduced in Table 7.

By use of the Bayesian procedure described previously, for each household the probability of a brand 1 purchase after the 10 observed purchases can be computed for the LLM parameter estimates $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\lambda}$. These computed p-values can be compared with actual brand choices at the eleventh purchase, so that the mean square deviation $(\Sigma e_i^2/n)$ and the test statistics S_1 and S_2 can be calculated in the same way as for the LS estimates. These also are given in Table 7. The following comments can be made.

1. From the χ^2 values it appears that according to the MCS criterion the fit of the LLM is satisfactory for beer and fopro. For margarine the fit is less satisfactory.
2. The MCS estimates are to be compared with the LS estimates of Table 1. The estimated values for α are small in both tables. Moreover, as seen before, the variance of these estimates is relatively large.

   Therefore and because of the phenomenon that estimates of β and λ are approximately complements of each other (caused by the tendency of α, β, and λ to sum to unity), a comparison of the estimated values of λ found by both methods is most appropriate. In three cases $\hat{\lambda}$ and $\tilde{\lambda}$ are rather close together (difference < .1); in three other cases the differences are somewhat bigger, although never more than 0.2. Such differences are not very large when one considers the general properties of minimum chi square estimates, especially with respect to their variance [see 2]. For the biggest difference in λ-values, viz. beer, B3, the value of the chi square criterion was computed using the LS parameter estimates instead of the MCS estimates. Because the LS method produces no corresponding values for μ_1 to μ_4, these values must be guessed. After some trial and error,

Table 7
MINIMUM CHI SQUARE PARAMETER ESTIMATES AND ADDITIONAL STATISTICS

Product	*Brand 1*	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{\lambda}$	χ^2_8	$\hat{\lambda}-\tilde{\lambda}$	$\Sigma e_i^2/n$	S_1	S_2
Beer	B1	.017	.403	.562	6.20	.183	.079	.05	−.61
	B2	.003	.285	.690	6.20	.027	.060	.09	−1.45
	B3	.001	.424	.555	4.98	.202	.023	.78	−1.80
	B4	.008	.398	.572	5.97	.081	.059	−.72	−.25
Fopro	F1	.005	.439	.545	2.67	−.161	.039	1.60	−1.12
Marg[a]	M1	.001	.168	.831	23.35	−.075	.031	−.08	1.37

[a]The results for margarine in [13] are somewhat different from those given here, because the former ones were based on the first 20 purchases of each consumer instead of 10.

values of μ_1 to μ_4 were found which, combined with the estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\lambda}$, produced a chi square value for B3 as low as 6.72. Thus for B3 the LS estimates are practically as good as the MCS estimates when the value of the MCS criterion is considered. As already noted, theoretically the values of λ for different brands in the same market have to be equal. Thus the smaller range of λ values for the four beer brands when LS estimates are compared with MCS estimates (0.10 versus 0.15) is a point in favor of the former. A comparison of $\Sigma e_i^2/n$ in Tables 7 and 2 shows that the mean square deviations for the MCS estimates, although always slightly higher, are surprisingly near to those for the LS estimates.

3. The test statistics S_1 and S_2 in Table 7 are always larger (absolutely) than the corresponding values for the LS estimates in Table 4. For fopro the MCS method tends to produce p-values which are systematically too high (the probability of finding a higher value for S_1 under H_0 is 3.5%). Although in the other cases individual test statistics are further removed from significance, the better fit for the least square estimates on the whole is evident. Also theoretically the LS method has advantages over the MCS method. With the MCS method when more purchase sequences from the same household are taken as observations, these observations are not independent as they should be. Moreover, the distribution of p_0 (the probability of choosing brand 1 at the start of an observed purchase sequence) is generally not constant when more sequences from the same consumer are used. Another consideration is the practical point that in the MCS method a highly nonlinear function in seven unknowns must be minimized, whereas in the LS method only three parameters must be estimated, by a straightforward regression procedure, complemented with a simple routine for updating the initial p-values of the households. The LS method also can be used to extend the linear learning model. For example, it enables the user to handle an LLM model for brand choice with different operators in different stores.

SUMMARY

The purpose of this article is to present a quick and efficient method for estimating the parameters of the linear learning model. The least squares method uses purchase histories of individual consumers as input. The method is demonstrated for empirical brand choice data from three different data sets. Reasonable parameter estimates were obtained, which obeyed the constraints of the linear learning model. In a test for goodness of fit these estimates produced very good results. A simulation study to validate the new estimation method also gave a favorable outcome.

Empirical findings for the least squares method were compared with those obtained by the minimum chi square method from the same data. The parameter estimates from the two methods were close in three cases (differences < 0.1) and the differences were somewhat bigger in the three other cases examined (between 0.1 and 0.2). In the goodness-of-fit test the minimum chi square estimates did not perform as well as the least squares estimates.

REFERENCES

1. Aaker, D. A. and J. M. Jones. "Modelling Store Choice Behavior," *Journal of Marketing Research*, 8 (February 1971), 38–42.
2. Blattberg, R. C. and S. K. Sen. "An Evaluation of the Application of Minimum Chi-Square Procedures to Stochastic Models of Brand Choice," *Journal of Marketing Research*, 10 (November 1973), 421–7.
3. Carman, J. M. "Brand Switching and Linear Learning Models," *Journal of Advertising Research*, 6 (June 1966), 23–31.
4. Cox, D. R. *The Analysis of Binary Data.* London: Methuen & Co., 1970.
5. Haines, G. H., Jr. *Consumer Behavior, Learning Models of Purchasing.* New York: The Free Press, 1969.
6. Herniter, J. D. and R. A. Howard. "Stochastic Marketing Models," *in Progress in Operations Research Vol. 2.* New York: John Wiley and Sons, Inc., 1964.
7. Johnston J. *Econometric Methods,* 2nd ed. New York: McGraw-Hill Book Company, 1972.
8. Kuehn, A. A. "Consumer Brand Choice—A Learning Process?" *Journal of Advertising Research*, 2 (December 1962), 10–7.
9. Lilien, G. L. "Application of a Modified Linear Learning Model of Buyer Behavior," *Journal of Marketing Research*, 11 (August 1974), 279–85.
10. Massy, W. F., D. B. Montogomery, and D. G. Morrison. *Stochastic Models of Buying Behavior.* Cambridge, Massachusetts: MIT Press, 1970.
11. McConnel, D. "Repeat Purchase Estimation and the Linear Learning Model," *Journal of Marketing Research,* 5 (August 1968), 304–6.
12. Morrison, D. G. "Upper Bounds for Correlation Between Binary Outcomes and Probabilistic Predictions," *Journal of the American Statistical Association*, 67 (March 1972), 68–70.
13. Wierenga, B. *An Investigation of Brand Choice Processes.* Rotterdam: Rotterdam University Press, 1974.