Penn Libraries
UNIVERSITY OF PENNSYLVANIA

University of Pennsylvania
ScholarlyCommons

Marketing Papers

Wharton School

December 1969

A Note on the Use of Markov Chains in Forecasting Store Choice

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

John U. Farley
University of Pennsylvania, farleyj@wharton.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

A Note on the Use of Markov Chains in Forecasting Store Choice

Abstract

Ehrenberg's sweeping criticism of Markov brand switching models [3] highlights many shortcomings of these models for aggregate analysis of consumer behavior. While it has been pointed out that some of his criticisms are not entirely correct [13], one of Ehrenberg's themes is unquestionably valid. The models tend to break down empirically due to violations of important Markovian stability assumptions [14]. A situation in which the assumptions of the model appear less restrictive is short-run forecasting of store choice behavior of individual families.

A Note on the Use of Markov Chains in Forecasting Store Choice
By J. Scott Armstrong and John U. Farley

Ehrenberg's sweeping criticism of Markov brand switching models [3] highlights many shortcomings of these models for aggregate analysis of consumer behavior. While it has been pointed out that some of his criticisms are not entirely correct [13], one of Ehrenberg's themes is unquestionably valid. The models tend to break down empirically due to violations of important Markovian stability assumptions [14]. A situation in which the assumptions of the model appear less restrictive is short-run forecasting of store choice behavior of individual families.

1. Switching Behavior and Transition Probabilities

Probability models in general [8] and transition probability matrices in particular have considerable intuitive appeal for organizing sequences of panel data as the following example illustrates. Suppose three families report the/following purchase sequences of a branded product in a time period [6]:

Family 1—AAAAAABACAAA
Family 2—CBBBBBBBBA
Family 3—CCCCCCCBCAA

Early students of brand switching found that these kinds of series were unwieldy and that a summary was needed. One approach was to analyze sets of summary statistics like the share of total purchases represented by a family's favorite brand [2]. Another was to categorize families into "loyalty classes" using purchase shares or similar types of measures [1]. A third alternative was a kind of "sources and destination" approach— that is, given the last purchase, what brand is likely to be purchased next? These data are conveniently cast in a transition matrix; the following matrix summarizes the data shown above as the fraction of purchases of a given brand going to all brands (including itself) at the next purchase.

Brand Purchased At Trial t + 1

		A	B	C
Brand purchased at trial *t*	A	.8	.1	.1
	B	.2	.7	.1
	C	.2	.2	.6

Such an array in and of itself may offer valuable insight into the process; here we see that, if the process remains stable and if the three families represent the entire buying population, brand A is more likely to retain customers than B or C; that A and C each loses customers about equally to the other brands; that B is more likely to lose sales to A than C, etc. Such insight may be useful for managerial purposes, such as identifying particularly dangerous or vulnerable competition.

It is a short step for the mathematically oriented [12] to grasp the potential application of properties of Markov processes if this transition matrix is stable over a long period and if it represents the process generating the data. Under these circumstances' long-run predicted market shares are implied for each brand (in this case, 50 per cent for A, 30 per cent for B, and 20 per cent for C); a variety of subsidiary measures such as variances of these shares and convergence rates to long-run solutions are also implied.

Unfortunately, the assumed stability of the transition matrix is critical for accurate prediction with these long-run shares, and the instability stressed by Ehrenberg turned up in almost all empirical tests. Of course, virtually all marketing activities are aimed at disrupting the stability of the transition probabilities. Further, there are a string of substantial problems such as aggregating purchases with different inter-purchase times [10] and families with unequal probabilities [9], which posed difficulties for both definition and interpretation. As a result, little has been published in which Markov models produce improvements in forecasting.

2. Short-run Disaggregated Forecasts

The fact that the case for aggregate long-run forecasts seems inadequate does not, however, preclude the possibility that the Markov formulation may be useful in forecasting short-run behavior of individual families. This formulation may be useful when 1) sufficient data are generated in a short time period to provide reliable estimates of transition probabilities, and 2) when the time period covered by the analysis is short enough so that the stability assumption is a reasonable approximation of reality; [4]. Supermarket choice seems to provide such a situation. Frequent visits to stores for virtually all families contrast markedly to low purchase rates for even frequently-bought products both in terms of the small proportion of a sample buying the product at all and the relatively low annual purchase rates among those consumers who do buy [7].

a) Forecasting with Markov Chains

The basic procedure for forecasting with Markov chains in this situation involves two stages:

1) Using data for one time period, estimate the transition matrix for the set stores visited by an individual family.

2) With this matrix, solve for the equilibrium shares for each of the stores chosen by the family and use this result for forecasting. This is the so-called "steady state" vector, or long-run proportional occurrence of each state[11] defined as the vector of shares, *t*, such that

$$tP = t$$

where *P* is the transition probability matrix. A "naive model" to use as a bench-mark for evaluation of the Markov model is to forecast that shares of purchases remain constant from period to period. This alternative model provides a test of the increased richness achieved by dropping the zero-order stochastic process (involving no memory) in favor of the first-order Markov process [14] (involving a memory of one purchase).

To provide a convenient link with work in the store loyalty field, *Chicago Tribune* panel data used by Cunningham [3] in his classic study are used in this analysis. Information on the store selections for fifty families in the Chicago-area covered one year; data from the first six months were used to estimate transition probabilities and to generate predictions for the second six month period. To solve serious store-class definition problems, only data from chain and independently affiliated supermarkets were used. Because of the tendency for infrequent visits to produce small and noisy estimates in some cells of the transition probability matrix, the present test is based on forecasting the proportion of purchases made at the most frequently visited store— the modal store. Five of the 50 families were dropped due to inadequate data on purchases in one of the two time periods, so the effective sample is 45 families.

Predictions using the Markov model and the no-change model were made for the percentage of trips that would be devoted to the modal store during the following six month period. This formulation poses a potential problem, dealt with below, associated with regression towards the mean.

b) Results

The no-change model explains only 52 per cent of the variance among family store shares between periods. The Markov model provides slight improvement to this fit, indicated by the correlation coefficient of .40 between actual change and predicted change. While this is significantly better (" = .01) than the no-change model, it reduces the standard error of the forecast by only eight per cent. Whether this is a useful gain depends on the cost-benefit relationships in any decision problem which utilizes the forecast.

There are, however, criteria other than least squares. A sign test was also used to evaluate how often the Markov model predicted the direction of the change [15], but, it showed little advantage for the Markov formulation over the no-change model. It was correct on 19 families and incorrect on 16. No change was predicted for 10 families. The sign test, of course, does not account for the magnitude of the change, and actions by stores (e.g. promotions) would be based on magnitudes rather than direction of change.

The results indicated regression toward the mean as there were 27 decreases in store share and 18 increases. Some 63 per cent of all trips were to the modal store in the first six months but only 59 per cent were to that same store in the second six months. Neither the Markov model nor the no-change model incorporate regression toward the mean, apparently caused by systematic positive error components associated with these shares in the first period. A model taking account of regression alone would predict a decrease for all cases—and thus be correct about the direction of change in 3/5 of the cases.

One final possibility is that the Markov model is really most useful when a family has just completed a change in probabilities. For a test of such extreme cases families were sought which seemed to undergo significant change during this first time period. A runs statistic was used to flag families whose number of subsequent runs to the same store was significantly ($\alpha =$ .025) lower than would be expected by chance using marginal shares as multinomial probabilities [15]. These families exhibit a nonrandom tendency to bunch purchases at first one store, then another. When one such an out-of control point was flagged for a family, the prior data were discarded and the initial transition matrix was developed only for the remainder of that period.

The no-change estimates were also developed from the truncated data base. Unfortunately, the number of sample points which were flagged was very small—six families qualified. The Markov model showed no substantial gain over the no-change model in these six cases. The direction of change was correct on two out of the six cases with one case of no change predicted. The no-change model was slightly better on the criterion of the magnitude of change.

3. Summary and Conclusions

The Markov model showed only slight predictive advantage over the no-change model for short-term forecasting of supermarket choices for a sample of 45 families. While this does not imply a blanket rejection of the Markov technique for forecasting, it is important to recall that this case held to a minimum many of the problems facing Markovian analysis—aggregation of dissimilar units, relatively low purchase rates, and requirement of such long sample periods to build up an adequate sample of events that the critical Markovian assumption of stable probabilities is almost certainly violated. Under these circumstances, the simpler model which says that "nothing changes" performs almost as well as the more refined Markov formulation. It is possible, of course, that the slight advantage of the Markov model will outweigh the increased cost of using such a model, but the no-change model has advantages both with regard to simplicity and to applying control-chart types of procedures to track series for stability over time. The usual qualifications about representativeness of geographic areas, panels, samples of panel members and time periods, of course, apply to this analysis.

References

Brown, G. H., "Brand Loyalty—Fact or Fiction," *Advertising Age*, 23, (January 26, 1953), 75-76.

Cunningham, R. M., "Brand Loyalty—What, Where, How Much," *Harvard Business Review*, 34, (1956), 116-128.

Cunningham, R. M., "Customer Loyalty to Store and Brand," Harvard Business Review, 41, (1962), 127-137.

Cyert, R. M., Davidson, H. J. and Thompson, G. L., "Estimation of the Allowance for Doubtful Accounts by Markov Chains," *Management Science*, 8, (1962), 287-303.

Ehrenberg, A. S. C., "An Appraisal of Markov Brand-Switching Models," *Journal of Marketing Research*, 2, (1965), 347-362.

Farley, J. U. and Kuehn, A. A., "Stochastic Models of Brand Switching," in George Schwartz, ed., *Science in Marketing*, New York: John Wiley and Sons, (1964), 446-65.

Farley, J. U. and Kuehn, A. A., "Brand Loyalty and Consumer Preferences for Branded Goods," unpublished Ph.D. dissertation, Graduate School of Business, University of Chicago, 1962.

Farley, J. U. and Kuehn, A. A., "A Test of the Loyalty-Proneness Hypothesis," Commentary, 8, (1966), 35-42.

Frank, R. F., "Brand Choice as a Probability Process," *Journal of Business*, 25 (1962), 43-56.

Kuehn, A. A., "An Analysis of the Dynamics of Consumer Behavior and Its Implications for Marketing Management," Ph.D. dissertation, Graduate School of Industrial Administration, Carnegie Institute of Technology, 1958.

Kemeny, J. G. and Snell, J. L., *Finite Markov Chains*, Princeton N.J.: D. Van Nostrand Co., Inc., (1960), 69-94.

Lipstein, B., "The Dynamics of Brand Loyalty and Brand Switching," *Better Measurement of Advertising Effectiveness: The Challenge of the 1960's*. Advertising Research Foundation, New York, 1959.

Massy, W. F. and Morrison, D. G., "Some Comments on Ehrenberg's Appraisal of Brand Switching Models," *Journal of Marketing Research*, May 1967.

Massy, W. F., "Order and Homogeneity of Family Specific Markov Processes," *Journal of Marketing Research,* 3, (1966), 48-55.

Siegel, S., *Nonparametric Statistics*, New York: McGraw-Hill, Inc., (1956).