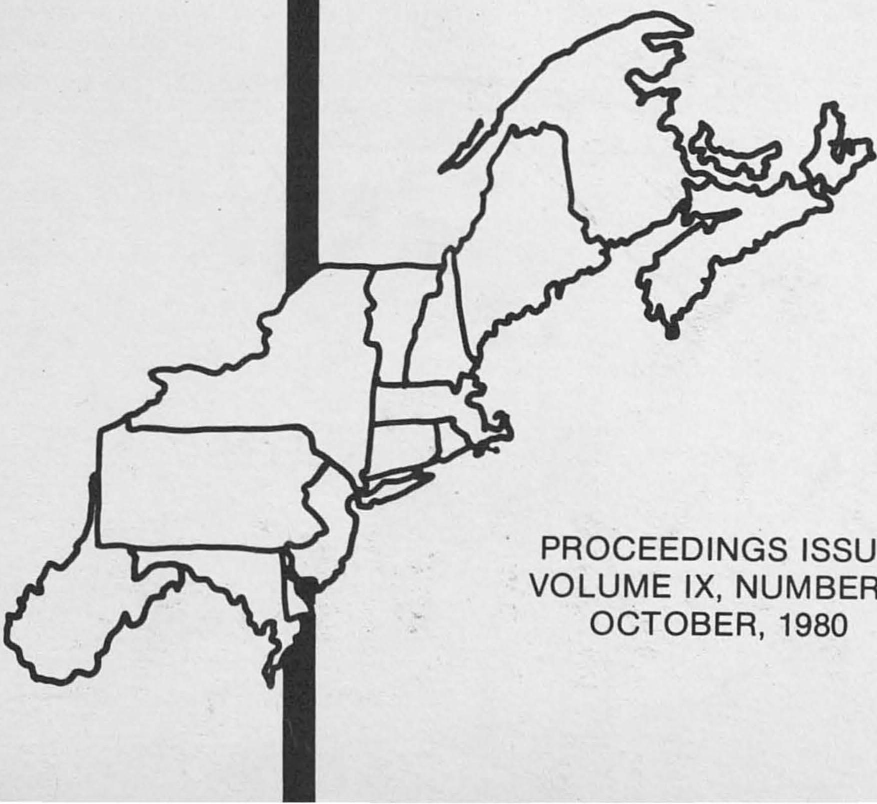
JOURNAL OF THE

Northeastern Agricultural Economics Council



PROCEEDINGS ISSUE
VOLUME IX, NUMBER 2
OCTOBER, 1980

ALTERNATIVE PROCEDURES FOR ESTIMATING THE SIZE DISTRIBUTION OF FARMS

R. N. Stavins and B. F. Stanton

Changes in the number and size distribution of dairy farms in the Northeast have come rapidly in the years since World War II. The objective of this study was to examine some of the newer methods of forecasting changes in this size distribution and ascertain the gains, if any, associated with these methods. Different formulations using Markov processes were compared with simple trend analyses and various functional forms in making projections. During the twenty-year period between 1958 and 1977 the number of farms delivering milk to plants in New York State decreased from slightly more than 45,800 to 16,500, a net decrease of approximately 64 percent. Over the same twenty-year period, annual milk production fluctuated between 9.8 and 11.0 billion pounds with a peak in 1965 and a low point in 1973. During the last five years, 1975-79, the number of farms delivering milk has continued to decline but milk production the the State has increased yearly and is expected to reach an all-time high in 1980. Such structural changes in the dairy industry have stimulated continued interest in problems of milk supply response and future variations in the size distribution of farms.

A twenty-county area in New York State from which essentially all milk produced is sold under the New York-New Jersey Market Order was selected for study. There were 14,272 farms in this area that sold milk at some time in the 10 year period, January 1968 through December 1977. A systematic, randomized list sample of 1,000 farms, stratified by counties and by entry and exit behavior, was drawn. The Market Administrator's Office provided monthly sales data on each farm for the ten-year period examined. To eliminate the effect of seasonality, monthly data were aggregated into annual totals and annual monthly averages calculated.

Farms were classified into size categories using an output measure, pounds of milk sold per farm. The interval used was 20,000 pounds of milk sold per month. This interval is roughly equivalent to 20 cows, assuming average annual sales of 12,000 pounds per cow or 1,000 pounds per cow per month. This allows forming a frequency distribution with ten categories. There is an entry category with 0 production and nine others roughly equivalent to 1-19, 20-39, 40-59 cows, etc. The last category is open ended 160,000 pounds or more (160 cows or more). (Table 2).

At no time during the ten-year period were all of the 1,000 farms selling milk simultaneously. A large number of farms terminated their sales during this interval, while others started or resumed production. As shown in Table 1, net farm numbers declined steadily during the period, while monthly sales per farm increased. In this respect the sample closely approximated aggregate numbers for the twenty-county area and for New York State as a whole.

One way of systematically examining changes in a size distribution over time is presented in Table 2. The original distribution in 1968 is compared with the one ten years later in 1977. Increases in monthly production per farm (and entries of new farms) are reflected in figures located above and to the right of the indicated diagonal, where no change occurred. Decreases and exits are found below and to the left of the diagonal.[1]

Table 1.
Farm Numbers and Average Monthly Production, New York Dairy Study, 1968-1977

Year	Number of farms in sample	Average monthly production per farm
1968	948	30,239
1969	875	33,146
1970	819	35,620
1971	792	38,124
1972	761	39,128
1973	730	38,962
1974	697	41,584
1975	670	43,994
1976	650	45,779
1977	640	47,312
Percent change 1968-77	-32%	+56%

During the ten year span, 64 new farms came into production while 372 discontinued milk sales. Of the 576 farms that produced milk continuously, there were 74 (13 percent) that decreased production enough to drop one or more size classes. Another 246 (43 percent) remained stationary in the same size category, and 256 (44 percent) increased by one or more size groups.

ALTERNATIVE PREDICTION METHODS

Nine alternative procedures[2] for predicting size distributions were investigated:

1. Simple trend analyses of size classes
2. Lognormal distribution
3. Simple negative exponential function
4. Boxley's generalized negative exponential function
5. Constrained Boxley negative exponential function
6. Dovring's transformed negative exponential function
7. Stationary micro-data Markov model
8. Stationary macro-data, Krenz type Markov model
9. Variable micro-data, Markov multinomial logit model

Based upon sample data for the period 1968-1974, each of the nine alternative procedures was used to develop estimates of the 1977 size distribution of farms. Each projected distribution was then compared with the actual 1977 distribution; where several important differences were noted. For example, the distribution which was predicted for 1977 using the lognormal function significantly underestimated the rate of exit in the smaller size

R. N. Stavins is Economist, University of California, Berkeley and B. F. Stanton is Professor of Agricultural Economics, Cornell University. The research is more fully reported in Stavins' thesis. The authors wish to recognize the important contributions of K. L. Robinson throughout the research effort and in editing and revising this article.

[1]The format used in Table 2 is that required for analysis with finite Markov processes (Kemeny, Snell and Knapp; Revuz).

[2]For a complete discussion of the nine methods, see Stavins (1979).

Table 2.
Change in Size of Sample Dairy Farms, 1968-1977

Size-class in 1968	Number of farms in 1968	Exit farms	Milk sold per month 1977 (2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) 0	ENTRY FARMS		19	20	14	4	2	1	1	1	2
(2) 1-19,999	349	213	58	54	11	10	2	0	1	0	0
(3) 20,000-39,999	360	117	35	116	63	21	5	2	0	1	0
(4) 40,000-59,000	154	27	1	22	50	38	9	4	1	0	2
(5) 60,000-79,999	59	11	2	0	9	16	9	5	4	1	2
(6) 80,000-99,999	13	3	0	0	1	2	3	3	0	1	0
(7) 100,000-119,999	8	0	0	0	0	1	1	1	1	2	2
(8) 120,000-139,999	1	0	0	0	0	0	0	0	0	0	1
(9) 140,000-159,999	3	1	0	0	0	0	0	0	0	1	1
(10) 160,000 or more	1	0	0	0	0	0	0	0	0	0	1
Number of farms in 1977		372	115	212	148	92	31	16	8	7	11
TOTAL FARMS (1968) = 948									TOTAL FARMS (1977) = 640		

classes (Table 3). Likewise, it did not project adequately the movement of farms into larger classes in the middle of the distribution.

Table 3.
Lognormal Distribution: Estimate of Size Distribution and Actual Distribution, New York, 1977

Monthly sales per farm	Estimated 1977	Actual 1977	Deviation
pounds	*number of farms*		
1-19,999	131	115	−16
20,000-39,999	206	212	+6
40,000-59,999	125	148	+23
60,000-79,999	69	92	+23
80,000-99,999	39	31	−8
100,000-119,999	23	16	−7
120,000-139,999	14	8	−6
140,000-159,999	9	7	−2
160,000 or more	2	11	+9

EVALUATING THE ALTERNATIVE PREDICTION METHODS

One method of evaluating alternative methods of predicting a size distribution is to compare the square root of the sum of the squared deviations (Ching, Davulis and Frick, 1974). This statistic was calculated for each of the estimated distributions and the nine methods ranked accordingly. This procedure has the advantage of examining overall predictive accuracy. Deviations in all size classes, however, are weighted equally. Thus, errors in estimation at the upper end of the distribution, which proportionately have a greater impact on an estimate of total supply, are counted the same as deviations at the lower end of the distribution. Clearly, this is only one of a number of ways in which results can be compared. Other criteria for evaluation include the costs of doing the analysis, data requirements and timeliness of the alternative predictive methods.

On the basis of minimizing the square root of the sum of squared deviations, the Markov models using individual farm data (micro-data) gave the best results. The four methods using some form of a negative exponential function provided relatively poor estimates. One of the major reasons for this is that the underlying model does not allow for maximum density at any point other than the beginning of the distribution (i.e., the smallest size class must also be the model class). The lognormal distribution and simple trend analysis gave reasonably good short-run results. Furthermore, the data requirements and computational costs of both methods are modest. This combination of characteristics may well explain why these methods are so commonly used in making short-run projections, particularly one or two years into the future.

Table 4.
Relative Accuracy of Alternative Predictions of 1977 Size Distribution of New York Dairy Farms Based Upon 1968-74 Data

Alternative methods	Square root of sum of squared deviations
(9) Micro-data variable Markov multinomial logic model	27.0
(7) Stationary micro-data Markov model	34.5
(2) Lognormal distribution	39.8
(1) Simple trend analyses of size classes	48.6
(8) Macro-data Krenz type Markov model	56.5
(6) Dovring's transformed negative exponential function	155.6
(5) Constrained Boxley negative exponential function	167.9
(3) Simple negative exponential function	176.1
(4) Generalized Boxley negative exponential function	221.8

MARKOV FORMULATIONS BASED ON INDIVIDUAL FARM DATA

The two procedures based on micro-data Markov models yielded results which had the smallest deviations between estimated and actual distributions. These methods, however, require significantly more computing time. The multinomial logit model can not be used without access to a computer and appropriate software packages. A further limitation of the stationary Markov model is that this approach assumes that those forces which caused changes in the size distribution in the base period will continue unchanged during the period over which projections are made. But such an assumption seldom holds when individual cross-sectional data are available and a test is made to determine whether the transition probabilities are in fact constant or stationary, (Colman, Hallberg). A chi-square test for this sample for the 1968-74 period rejected the null hypothesis of constant probabilities, (Colman, Hallberg).[3]

The multinomial logit model (Tyrrell and Mount) builds on Hallberg's work where variable probabilities were incorporated into a Markov framework using least squares regression and a set of exogenous variables to account for some of the variability. In this study, a series of variables[4] which might affect growth, decline, entry and exit of dairy farms were examined. Based upon both logical relationships and empirical evidence the New York State milk-feed price ratio was introduced as the most appropriate

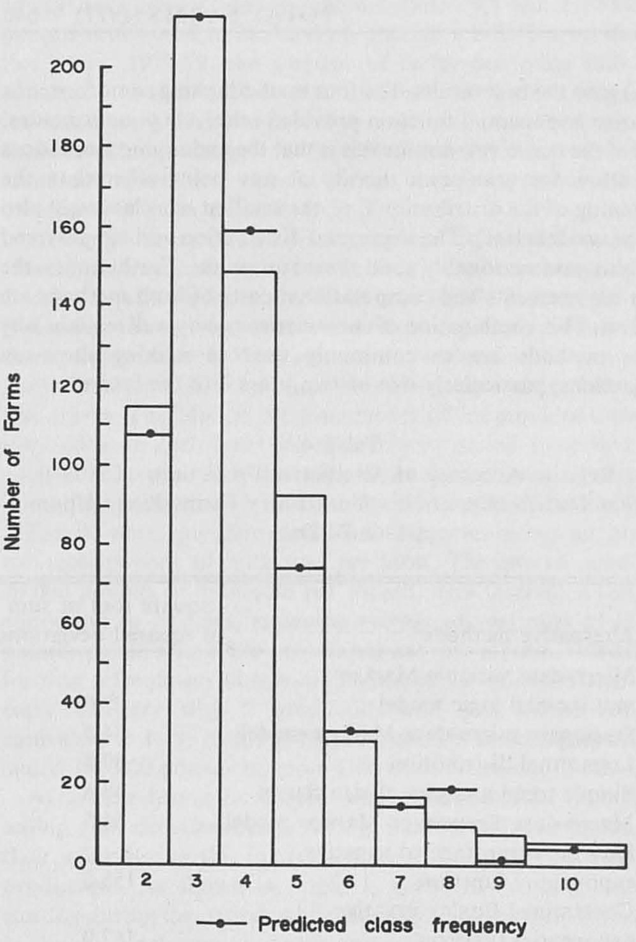


FIGURE 1.
Actual and Predicted 1977 Frequency Distribution of Sample Farms—Micro-Data Variable Markov Multinomial Logit Model

explanatory variable. An increase in the milk-feed price ratio provides additional incentive to increase milk production, keep cows in the herd longer and reduce the culling rate. Conversely a decrease in this ratio discourages grain feeding and encourages higher rates of culling. Historical data on this ratio were incorporated into the multinomial logic model for the period 1968-74, and the sample distribution was estimated for the year 1977. The results of that projection are presented in Figure 1 together with the actual distribution.

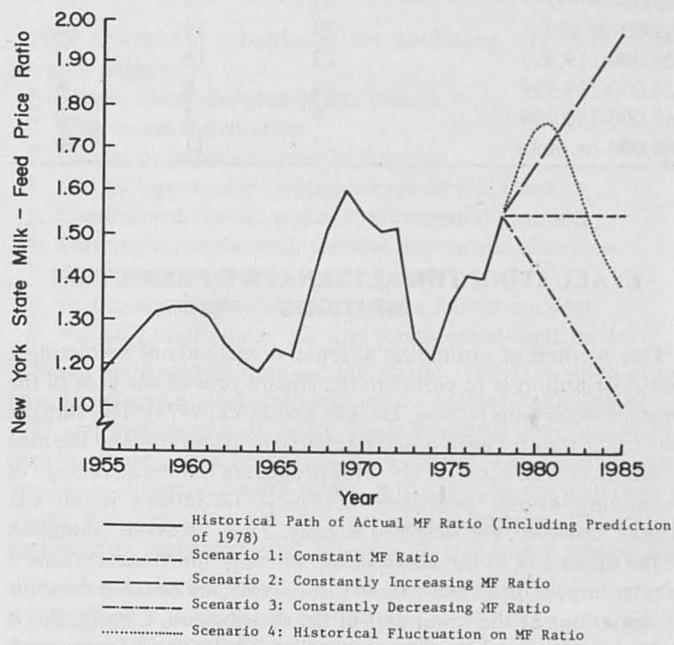
While there are deviations from the actual frequencies in all of the classes, the deviations are generally modest. Perhaps the most important differences in the projections are located in the three largest size classes.

Average monthly production	*Actual number*	*Projected number*
(8) 120,000-139,999	8	21
(9) 140,000-159,999	7	1
(10) 160,000 and over	11	8

In the two preceding size classes, (6) and (7), the projections were essentially equivalent to the actual frequencies. However, in the three largest categories the estimates differed substantially but indicated a total of 30 farms instead of the actual 26 when considered as a total. When two of the size classes are aggregated the projections tend to look better. But there remains the failure to project adequately the shift of some farms to larger sizes.

PROJECTIONS WITH THE MULTINOMIAL LOGIT MODEL TO 1985

The nine methods of projecting size distributions were based on sample data from 1968-1974 in making estimates of the 1977 distribution. These estimates were then checked against the actual distribution for that year. The micro-data Markov multinomial logit model provided the "best" results. Therefore, this method was also used to develop projections of dairy farm size distributions to



Source: New York State Department of Agriculture and Markets. New York State Agricultural Prices and Cash Receipts from Farm Marketings, 1940-1963 and New York Agricultural Statistics 1971, 1973, 1977.

FIGURE 2.
Values of the Exogenous Variable—Milk-Feed Price Ratio for Four Scenarios, Multinomial Logit Model, 1978-1985

[3]The test procedures followed techniques developed by Anderson and Goodman and modified by Colman. Each of the six annual transition probability matrices of the period 1968-74 and the overall transition probability matrix for the seven year period were evaluated. The null hypothesis of stationarity was rejected (2804>522) at the 99 percent level (Stavins, p. 155).

[4]A set of 30 different variables was considered including such diverse items as the consumer price index, price of dairy feed, May rainfall, upstate New York business activity index, and the slaughter price of cows (Stavins, p. 161).

Table 5.
Predicted Size Distribution of Sample New York Dairy Farms Multinomial Logit Model, Fluctuating Milk-Feed Price Ratio, 1978-85

Year	1-19,999	20,000-39,999	40,000-59,999	60,000-79,999	80,000-99,999	100,000-119,999	120,000-139,999	140,000-159,999	160,000 and over
	Average monthly milk sales per farm								
	number of farms								
1978	105	204	145	90	34	16	12	5	13
1979	95	198	138	89	35	17	16	2	16
1980	88	190	133	88	36	18	20	1	17
1981	85	181	132	85	37	18	24	1	17
1982	87	170	136	80	37	17	28	2	17
1983	91	161	141	76	36	16	31	4	17
1984	93	154	143	73	35	15	34	6	17
1985	93	150	142	71	34	15	36	8	17

Table 6.
Predicted 1985 Distributions of New York Dairy Farms, Multinomial Logit Models, Four Milk-Feed Price Ratio Scenarios

Average monthly milk sales per farm	Asumptions about milk-feed price ratios (1) Constant at 1.55	(2) Increasing	(3) Decreasing	(4) Fluctuating
pounds				
1-19,999	89	80	100	93
20,000-39,999	161	162	160	150
40,000-59,999	136	124	147	142
60,000-79,999	77	79	73	71
80,000-99,999	35	37	33	34
100,000-119,999	17	19	15	15
120,000-139,999	37	39	34	36
140,000-159,999	7	3	12	8
160,000 and over	18	21	15	17
Total farms	577	564	589	566
Estimated New York supply (bill. lbs.)	10,96	11.01	10.86	10.66

the year 1985. In the new analysis all of the data from 1968-1977 were incorporated in the predictive model. To do so it was necessary to project the model's exogenous variable, the New York State milk-feed price ratio. Four possible scenarios relative to the milk-feed price ratio were considered. These are shown in Figure 2 along with the historical variation in this key price ratio.

The results obtained in Scenario 4, which assumes a continuing kind of cyclical fluctuation in the milk-feed price ratio, are presented in Table 5. The total number of sample farms continues to fall, but successively more slowly. There is a gradual but steady shift toward larger units. Interestingly, the number of farms in the largest size category becomes virtually stationary in this formulation.

The differences among the various predictions of 1985 size distributions based upon the four scenarios of milk-feed price ratios were quite modest but interesting. An increase in milk prices relative to feed shifted small producers into larger size categories, while decreasing the milk-feed price ratio kept more farms in production but at lower levels of monthly production (Table 6).

If these predicted sample size distributions are converted into estimates[5] of total New York State production in 1985 they provide results which differ by only 3 to 4 percent at the extremes. Perhaps surprisingly, the fluctuating milk-feed price ratio gave the smallest estimate of total production and one of the higher exit rates from dairy production.

[5]An estimate of state-wide production was based on the following assumptions. First the proportion of state milk production represented by farms included in the sample in December 1977 would hold in 1985. The factor used to increase the sample total was 28.1977. Second, the midpoint of each class except the open ended one adequately represented the members of that class. The mean size of farm in the largest class (200,000 pounds of milk sold per month) was used for this class in 1985. When this procedure was used to check the method for 1977 the estimate was 10.215 billion pounds compared to 10.228 billion actually sold.

CONCLUDING OBSERVATIONS

This study considered a set of nine alternative methods of predicting size distributions of dairy farms in New York State. Monthly sales data for a stratified, random sample of 1,000 farms in a twenty-county area over a ten-year period were used to make the projections. To test these alternatives, a set of projections was first made based on the initial seven years of data. These results were then compared with the actual distribution in the tenth year, 1977. The Markov multinomial logit model most closely approximated the actual distribution in 1977 and was therefore used to make predictions to the year 1985. This method allowed for the use of variable transition probabilities in making the estimates and incorporated the milk-feed price ratio as an exogenous variable in the model for developing the final estimates of size distributions.

REFERENCES

Aitchison, J., and J. A. C. Brown, *The Lognormal Distribution*, Cambridge, England: Cambridge University Press, 1957.

Anderson, T. W., and Leo A. Goodman, "Statistical Inference About Markov Chains," *The Annals of Mathematical Statistics*, 28(1957):89-110.

Boxley, Robert F., "Farm Size and the Distribution of Farm Numbers," *Agricultural Economics Research*, 23(1971):87-94.

Ching, C. T. K., J. P. Davulis, and G. E. Frick, "An Evaluation of Different Ways of Projecting Farm Size Distributions," *Journal of the Northeastern Agricultural Economics Council*, 3(1974):14-22.

Colman, David R., "The Application of Markov Chain Analysis to Structural Change in the Northwest Dairy Industry," *Journal of Agricultural Economics* (British), 18(1967):351-361.

Conneman, George J., and D. H. Harrington, "Implications of Exit and Entry of Farm Firms in Agricultural Supply Analysis," *Agricultural Economics Research*, 21(1969):40-44.

Dent, Warren T., and Richard Ballintine, "A Review of the Estimation of Transition Probabilities in Markov Chains," *Australian Journal of Agricultural Economics*, 15(1971):69-81.

Dovring, Folke, "Distribution of Farm Size and Income: Analysis by Exponential Functions," *Land Economics*, 49(1973):133-147.

________, *Income and Wealth Distributions: The Exponential Functions*, Department of Agricultural Economics, AE-4212, University of Illinois, 1969.

Duncan, George T., and Lizbie, G. Lin, "Inference for Markov Chains Having Stochastic Entry and Exits," *Journal of the American Statistical Association*, 67(1972):761-767.

Furniss, I. F., and Bengt Gustafson, "Projecting Canadian Dairy Farm Structure Using Markov Processes," *Canadian Journal of Agricultural Economics*, 16(1968):64-78.

Hallberg, Milton C., "Estimation of Regression Parameters with the Predicted Dependent Variable Restricted in a Certain Range: A Reply," *American Journal of Agricultural Economics*, 52(1970):615.

________, "Projecting the Size Distribution of Agricultural Firms—An Application of a Markov Process with Non-Stationary Transition Probabilities," *American Journal of Agricultural Economics*, 51(1969):289-302.

Hart, P. E., and S. J. Prais, "The Analysis of Business Concentration: A Statistical Approach," *Journal of the Royal Statistical Society*, Series A. 119(1956):150-181.

Judge, G. G. and E. R. Swanson, *Markov Chains: Basic Concepts and Suggested Uses in Agricultural Economics*, Department of Agricultural Economics. AERR-49, University of Illinois, 1961.

Kemeny, J. G., J. L. Snell, and A. W. Knapp, *Denumberable Markov Chains*, Princeton: D. Van Nostrand, 1966.

Krenz, Ronald D., "Projection of Farm Numbers for North Dakota with Markov Chains," *Agricultural Economics Research*, 21(1964):77-83.

Lee, Tsoung-Chao, "Estimation of Regression Parameters with the Predicted Dependent Variable Restricted in a Certain Range: A Comment," *American Journal of Agricultural Economics*, 52(1970):613-615.

Lee, T. C., G. G. Judge, and A. Zellner, *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*, Amsterdam: North-Holland Publishing Company, 1977.

Power, A. P., and S. A. Harris, "An Application of Markov Chains to Farm-Type Structural Data in England and Wales," *Journal of Agricultural Economics* (British), 22(1971):163-177.

Quandt, Richard E., "On the Size Distribution of Firms," *American Economic Review*, 56(1966):416-432.

Revuz, D., *Markov Chains*, Amsterdam: North-Holland Publishing Company, 1975.

Salem, Ali Ben Zaid, "Firm Size and Changes in Industrial Structure: An Econometric Analysis of the Dairy Industry in New York State," Ph.D. Thesis, Cornell University, 1973.

Salkin, Michael S., Richard E. Just, and O. A. Cleveland, Jr., "Estimation of Nonstationary Transition Probabilities for Agricultural Firm Size Projection," *Annals of Regional Science*, 10(1973):71-82.

Stanton, Bernard F., and Lauri Kettunen, "Potential Entrants and Projections in Markov Process Analysis," *Journal of Farm Economics*, 49(1967):633-642.

Stavins, Robert N., "Forecasting the Size Distribution of Farms: A Methodological Analysis of the Dairy Industry in New York State," M.S. thesis, Cornell University, 1979.

Tyrrell, Tim and Tim Mount, *An Application of the Multinomial Logit Model to the Allocation of Budgets in the U. S. Consumer Expenditure Survey for 1972*, Department of Agricultural Economics, A. E. Staff Paper No. 78-20, Cornell University, 1978.