Projection of Farm Numbers for North Dakota With Markov Chains

By Ronald D. Krenz

The Markov process was first applied in economics to problems of relative structure, as for example in the analysis of income and wage distributions. The extension of such distributions over time was an initial step in projection. The analysis in this paper takes a further step by developing a method for making projections of absolute numbers from an appropriately modified Markov base. This new departure represents both a weakness and a strength; weakness because the modifying assumptions abstract considerably from reality, but strength because the technique covers a broader area of obvious need.

THE RESULTS of the 1959 Census of Agriculture have given rise to much speculation regarding the future number of farms. Current literature contains numerous analyses of census data and discussions of the implications for agricultural policy and farm-income problems (*3*, *6*, *7*, and *10*).[1]

The article reports the use of the Markov-chain process to project future farm numbers in North Dakota from census data. Such projections are useful in studies of economic adjustments and area development.

Concept of Markov Chains

The concept of Markov chains was introduced around 1907, but did not come into general use by economists until fairly recently. In the last few years the technique has seen considerable use in the analysis of income and wage distributions (*9*) and in studies of the size distributions of firms in the steel industry (*1*). In the field of agriculture it has been used to study sizes of hog-producing firms (*5*) and farm tenure in Illinois (*8*) and wheat yields in Montana (*2*).

For those who are not familiar with Markov chains, the study of wheat yields in Montana by Bostwick (*2*) gives a good general explanation of the technique.

The process assumes that any population of firms or individuals can be classified into various groups or "states" and that movements of firms or individuals between states over time can be regarded as a stochastic process. With a given set of states ($S_1, S_2 \dots S_n$), it is assumed possible to estimate the probabilities (p_{ij}) of firms moving from S_i to S_j. These probabilities of movements during a given time period can be expressed in a transition matrix P:

$$P = \begin{array}{c} \\ S_1 \\ S_2 \\ \cdot \\ S_n \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & \dots & S_n \end{array} \\ \left[\begin{array}{cccc} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \cdot & \cdot & & \cdot \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{array}\right] \end{array} \qquad (1)$$

The states used in this analysis are the size groups used by the U.S. Census Bureau classification of farms, as follows:

State	*Farm size (acres)*
S_0	No farms.
S_1	10–99.
S_2	100–179.
S_3	180–259.
S_4	260–499.
S_5	500–999.
S_6	1,000 and over.

Various types of projections can be obtained from the transition matrix and the initial distribution of farms. Among other items, this method will give estimates of the equilibrium distribution and number of farms, the mean lifetime of a farm within a size state, and indexes of mobility of farms among states. Our principal concern is with projections of farm numbers at various future dates.

Estimating a Transition Matrix

To estimate a transition matrix, data are needed that describe the movements of individual firms over time. With such data the transition probabilities can be estimated by averaging these movements. However, in the problem at hand, and for

[1] Italic numbers refer to Literature Cited, p. 83.

Table 1.—*Farm numbers in North Dakota, 1935–59*

Year	Size state						Total
	S_1	S_2	S_3	S_4	S_5	S_6	
	Farms	*Farms*	*Farms*	*Farms*	*Farms*	*Farms*	*Farms*
1935	3, 948	13, 572	5, 552	35, 133	19, 891	5, 250	83, 346
1940	2, 985	10, 415	4, 491	29, 620	19, 371	6, 405	73, 287
1945	2, 272	6, 654	3, 670	26, 198	22, 004	7, 975	68, 773
1950	1, 900	5, 205	3, 340	23, 317	22, 138	8, 831	64, 731
1955	1, 632	4, 462	2, 831	20, 337	21, 999	9, 925	61, 186
1960	1, 451	3, 277	2, 166	15, 596	20, 672	11, 364	54, 526

Data from U.S. Census of Agriculture.

many similar cases, the data on individual movements are not available. The quinquennial U.S. Census of Agriculture, the best known source of data on farm numbers in North Dakota, records only the number of farms in each of several size groups as of the date of enumeration (table 1). With census data alone it is not possible to determine whether the farms in a given size group at a given time are the same farms that were in that size group at a preceding time or are farms that have "moved" from other size groups. In the absence of detailed data on individual movements, the Markov process loses some of its usefulness as an analytical tool. Detailed data on farm movements would provide the means for a useful analysis of farming careers along with the changes in farm size. However, with appropriate assumptions the data available can be recast in such a way that a Markov chain analysis is possible.

To utilize census data, assumptions must be made regarding the movements of farms between size groups or states. These assumptions must be based on whatever knowledge the analyst may have regarding the population.

The following assumptions seem reasonable:

1. Operators of any size of farm in North Dakota would expand their acreage, if possible. This is a commonly expressed opinion of farmers and is consistent with production cost data which indicate some economies of size in farming (*4*).

2. The farms most likely to expand are those that are initially larger than average. This is likely to be true for the size groups here dealt with. The available cost data indicate that these larger farms have lower production costs and hence should be in a better financial position to purchase land. They may also gain further cost economies by further expansion.

3. Increases in farm size are likely to come about by gradual increases in acreage. This is true because of the problems of finding additional land for sale in the community and of financing land purchases.

4. Decreases in size of farms are not likely to occur. Again, because of the existing economies of size, decreases in size of units are not likely to be made voluntarily. A farm is more likely to disappear as a farming unit than to become smaller.

These assumptions give rise to the following rules for determining the transition of farms from size state to size state.

1. Farms in the largest category, S_6, remain in this category.

2. Increases in number of farms in any state S_i come from the next smaller state, S_{i-1}.

3. Any decrease in number of farms in any state (other than from rule 2 above) results in a movement to S_0, i.e., they are assumed to go out of business rather than move to S_{i-1}.

An appropriate base period is an important consideration for any set of projections. The choice of a base period for a transition matrix becomes especially important when the rate of change has been as variable over time as in this instance (table 2). Although several base periods were used in constructing the transition matrix, results presented here were based on the 1935–60 period.[2]

[2] As indicated elsewhere in this paper, there are really two kinds of base periods: one is the base for determining the transition matrix, the other is the base used as a starting point in making number projections. Both are necessary.

TABLE 2.—*Percentage rates of decline in farm numbers in North Dakota*

Period	Decline in 5-year period
	Percent
1935 to 1940	12. 07
1940 to 1945	6. 16
1945 to 1950	5. 88
1950 to 1955	5. 48
1955 to 1960	10. 88

The heart of the Markov process is the matrix of probabilities of movements of farms from size state to size state. This transition matrix is determined by the following procedure. Using the above rules for farm movements and the census data for 5-year intervals, a flow chart is developed for each 5-year interval. Each flow chart, when completed, shows the estimated movements of farms from state to state during a period. Table 3 illustrates the estimated transition for the period from 1935 to 1940. The beginning and ending numbers of farms in each group are known (column and row totals). From this we proceed to estimate the individual movements that took place between time periods. We start with S_6 (this is not necessary but it is convenient). To obtain 6,405 farms in S_6 in 1940, we retain the 5,250 farms that were in S_6 in 1935, and show a movement of 1,155 farms from S_5 in 1935 to S_6 in 1940. This leaves 18,736 farms in S_5 for 1940. However, we need a total of 19,371 in S_5 for 1940 so we must transfer 635 farms from S_4 in 1935 to S_5 in 1940. This leaves 34,498 farms in S_4 but we only need 29,620 in S_4 for 1940. Hence the remaining 4,878 farms are transferred to S_0. This procedure is continued for the remaining cells until the totals all balance.

The same procedure is repeated for each 5-year interval. When all the tables are completed (for five intervals in this case), like items in all of the tables are added. For instance, the items S_{56} (row 5, column 6) from table 3 will be added with S_{56} from each of the other tables. The same addition is performed on the row totals. The individual items in the table of totals are then divided by the row totals. This gives an average of the transitions between states as a percentage of the firms in each state. These percentages are then entered in the transition probability matrix.

The transition matrix, with 1935 to 1960 as the base period, is as follows:

$$P=\begin{array}{c} \\ S_0 \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{array}\begin{array}{c} \begin{array}{ccccccc} S_0 & S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ .1958 & .8042 & 0 & 0 & 0 & 0 & 0 \\ .2554 & 0 & .7446 & 0 & 0 & 0 & 0 \\ .1307 & 0 & 0 & .8301 & .0392 & 0 & 0 \\ .0997 & 0 & 0 & 0 & .8491 & .0512 & 0 \\ 0 & 0 & 0 & 0 & 0 & .9420 & .0580 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad (2)$$

TABLE 3.—*Estimated transitions of farms in North Dakota from 1935 to 1940*

Size group in 1935	Size group in 1940							
	S_0	S_1	S_2	S_3	S_4	S_5	S_6	Total
S_0	0		0	0	0	0	0	0
S_1	963	2, 985	0	0	0	0	0	3, 948
S_2	3, 157	0	10, 415	0	0	0	0	13, 572
S_3	1, 061	0	0	4, 491	0	0	0	5, 552
S_4	4, 878	0	0	0	29, 620	635	0	35, 133
S_5	0	0	0	0	0	18, 736	1, 155	19, 891
S_6	0	0	0	0	0	0	5, 250	5, 250
Total	10, 059	2, 985	10, 415	4, 491	29, 620	19, 371	6, 405	83, 346

This matrix is typical of the transition matrices obtained regardless of the base period used. The coefficients indicate the percentage of the farms in S_i that will probably be in S_j after 5 years. The matrix is fairly sparse because of the type of data available and the need for making assumptions on the movements of farms. The sum of the coefficients in any row must always be unity if all farms are to be accounted for. Row S_0 indicates that no new firms are entering farming and column S_0 indicates that some farms from S_1, S_2, S_3 and S_4 are going out of business. State S_0 is an "absorbing state" as farms enter this state but do not leave it.

State S_6 (farms of 1,000 or more acres) also is an absorbing state. The number of farms in this category has been steadily increasing since 1935.

The matrix reflects the initial assumptions on movements of farms. The principal diagonal contains fairly large coefficients, which would indicate stability of farm size. But in this example the size of the coefficient is partially due to the assumptions on movements of farms. Data on actual farm movements would give a transition matrix containing many more nonzero elements and would thus reflect less stability. Certainly these movements take place but they cannot be identified from census data. One could "fill" these cells by changing the assumptions about movement of farms, but the resulting coefficients would have little meaning. It was found that other sets of assumptions regarding movements of farms produced different projections for the individual states but had little effect on the projected totals. The change in the totals is related to the historical average change in farm numbers in the base period.

The coefficients in the transition matrix provide some worthwhile information not readily available from other types of projections. For instance, the coefficients in row S_3 indicate that approximately 13 percent of the farms of 180 to 259 acres will probably go out of farming in any 5-year period. Of this same group, 83 percent will remain in the same size state and 4 percent will expand to 260–499 acres in any 5-year period. In comparison, 5.1 percent of the farms in S_4 will probably expand to S_5 in 5 years and in the same time period 5.8 percent of the farms in S_5 will probably expand to S_6. On the other hand, farms smaller than S_3 are not likely to expand. However, the transition matrix also indicates that farms of 500 acres or more (S_5 and S_6) do not go out of business. Clearly, it would be much more meaningful if these items of information were based on data for individual farm movements instead of assumptions.

Projecting Farm Numbers

The projections of farm numbers can be obtained with the use of the transition matrix by either of two slightly different methods. The most commonly used method is to multiply the P matrix by itself n times to obtain the probability of movements during n time periods. Individual elements of the desired P matrix are then multiplied by the farm numbers in their respective size states in the base year, or period, used for projection. An alternative method is to multiply the distribution of farms in the base year for projection by P to obtain the projections for one period and continue to postmultiply the results by P for the desired number of periods. The latter procedure has the advantage of giving projections for each time period, in this case every 5 years, and is simplest in this case since the transition matrix is relatively sparse.

In this paper, the first projection base is 1935. The first year in the base period for estimating the transition matrix was used as the projection base. But the P^n matrix could be applied to any other year or base that seemed appropriate. It might sometimes be more reasonable to make this application to the most recent information about the number of farms in different size groups. Taking an earlier base may give estimates for later years which do not agree with known numbers. In the present study, using a 1935 base for projections gives estimates of farm numbers for 1960 that agree very closely with the actual numbers in 1960.

The resulting estimates of the total number of farms for 1975 and 2000 show some differences between projections that are due to selection of different base periods (table 4). The rate of decline in total number of farms is related to the average rate of decline in the base period, although this is not a 1-to-1 relationship.

The projected distribution of farms for 1975 and 2000 indicates differential rates of decline in farm numbers by size categories (table 5). Comparisons with table 1 indicate the magnitude of changes that can be expected for each size group. The pro-

TABLE 4.—*Projected number of farms in North Dakota using different base periods for the transition matrix*

Base period	Projected number of farms for—		Projected decline in number of farms from 1960 to—		Average rate of decline per year
	1975	2000	1975	2000	
	Farms	*Farms*	*Farms*	*Farms*	*Percent*
1935 to 1960	46, 814	41, 247	7, 712	13, 279	1. 37
1945 to 1960	46, 854	40, 057	7, 672	14, 469	1. 38
1950 to 1960	45, 614	38, 507	8, 912	16, 019	1. 58
1955 to 1960	42, 511	35, 275	12, 015	19, 251	2. 18

jections have many implications for present farmers and for young people contemplating farming as a career.

The Markov process also allows the estimation of several other parameters that may be of interest. With two absorbing states, S_0 and S_6, the transition matrix (2) can be rearranged as in equation (3).

If the southeast submatrix of equation (3) is designated Q, then where I is an identity matrix, $[I-Q]^{-1}$ gives the mean number of periods in each transient state for each initial nonabsorbing state. Multiplying this matrix by the scalar 5 (5-year periods) converts these estimates to years. Equations (3) and (4) are given below.

TABLE 5.—*Projected number and distribution of farms in North Dakota (transition base=1935–60)*

Size of farm	1975	2000
	Farms	*Farms*
S_1 (10 to 99 acres)	691	232
S_2 (100 to 179 acres)	1, 282	294
S_3 (180 to 259 acres)	924	364
S_4 (260 to 499 acres)	9, 925	4, 471
S_5 (500 to 999 acres)	19, 236	15, 924
S_6 (1,000 acres and over)	14, 756	19, 962
Total	46, 814	41, 247

$$
P=\begin{array}{c}S_0\\S_6\\S_1\\S_2\\S_3\\S_4\\S_5\end{array}
\left[\begin{array}{cc|ccccc}
S_0 & S_6 & S_1 & S_2 & S_3 & S_4 & S_5\\
1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0\\
\hline
.1958 & 0 & .8042 & 0 & 0 & 0 & 0\\
.2554 & 0 & 0 & .7446 & 0 & 0 & 0\\
.1307 & 0 & 0 & 0 & .8301 & .0392 & 0\\
.0997 & 0 & 0 & 0 & 0 & .8491 & .0512\\
0 & .0580 & 0 & 0 & 0 & 0 & .9420
\end{array}\right] \quad (3)
$$

$$
[I-Q]^{-1}\,(5)=\begin{array}{c}S_1\\S_2\\S_3\\S_4\\S_5\end{array}
\left[\begin{array}{ccccc}
S_1 & S_2 & S_3 & S_4 & S_5\\
25.54 & 0 & 0 & 0 & 0\\
0 & 19.58 & 0 & 0 & 0\\
0 & 0 & 29.43 & 7.64 & 6.75\\
0 & 0 & 0 & 33.13 & 29.25\\
0 & 0 & 0 & 0 & 86.21
\end{array}\right] \quad (4)
$$

Thus for a farm in S_3 the mean number of years before absorption by S_0 or S_6 is 29.43 years in S_3, 7.64 in S_4, and 6.75 years in S_5. For farms in S_2 the mean number of years before absorption is 19.58. These estimates are another indication of the rate of change in number of farms by size of farm. These estimates do not mean, for example, that in 19.58 years the number of farms in S_2 will be reduced to half the initial number. These estimates are *mean number of years* of a firm in a state and *not the years needed to reach the mean number*. Actually, the number of farms in S_2 will be reduced to 50 percent of the initial number in approximately 12 years, and to 5 percent in 50 years.

Another set of estimates available with the Markov process is estimates of the equilibrium distribution of firms. Equilibrium in a Markov process has been interpreted by Adelman as follows:

> [Equilibrium] may be defined as that distribution for which the average number of corporations entering a given stratum per period equals the average number of businesses leaving it. Our concept of equilibrium is statistical in nature for the industry, and dynamic for the individual firm. In other words, equilibrium in this paper does not imply that there is no movement of enterprises between strata. On the contrary, the stochastic conception of equilibrium explicitly requires that firms move in and out of each class. But on the average, forces acting to increase the number of enterprises in a given size range are exactly counterbalanced by those tending to decrease it (*1*, pp. 895–96).

With absorbing chains, all equilibrium distributions will consist of firms in only the absorbing states, in this case S_0 and S_6. In other words, it is assumed that all farms will either go out of business or consist of 1,000 acres or more. Hence the estimates of the equilibrium distributions with absorbing states lead to trifling results.

However, it may be of interest to know what the number of surviving farms may be. This can be arrived at by first estimating the probability that firms in each nonabsorbing state will end up in each absorbing state. If the southwest submatrix of equation (3) is designated R, then:

$$[I-Q]^{-1}R = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \end{array} \begin{array}{c} \begin{array}{cc} S_0 & S_6 \end{array} \\ \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ .9217 & .0783 \\ .6607 & .3393 \\ 0 & 1 \end{array}\right] \end{array} \qquad (5)$$

The elements of equation (5) indicate the probabilities of absorption of states S_1 to S_5 by state S_0 and S_6. As shown, the probability is 1 that S_1 and S_2 will be absorbed by S_0. This was obvious from the transition matrix. S_5 eventually will be absorbed into S_6. Farms in S_3 and S_4 will be absorbed by either S_0 or S_6. Ninety-two percent of the farms in S_3 and 66 percent of the farms in S_4 will be absorbed by S_0. Thus the number of remaining farms will be approximately 8 percent of the farms in S_3, 34 percent of S_4, and all of the farms in S_5 and S_6, a total of 37,500 farms (1935–60 base period). In contrast, use of the 1955–60 base period gives an estimate of 32,400 as the number of surviving farms.

Reliability of Estimates

Statistical tests of the reliability of these types of projections have not been developed. The method assumes that the probability for the farms in any one size group moving to another remains constant over time. But this assumption becomes increasingly unrealistic as times goes on. Conceivably, some farmers may have been adjusting rapidly to a change in technology during the base period. The remaining farmers may not adjust in the same way or at all. But the procedure assumes that they behave in each size group just as they did during the base period.

Another possible inconsistency has to do with the relationship between average numbers and average sizes of farms in each size group. With a relatively fixed total acreage of farmland in North Dakota, the projected increase in farm numbers in the larger groups means that the average size in one or more of the size groups must fall. This may be all right, but it raises a question about whether the Markov process results in the most reasonable allocation of the changes between numbers and sizes.

Regression techniques using a single independent variable can be used to make similar projections by extrapolation. Such methods embody many of the same sets of assumptions as those implied in the Markov chain. Regression techniques do permit the calculation of errors of estimate or confidence intervals for estimates within the range of the data. But when the technique is used to project an estimate outside of the range of the original data, the confidence limits are no longer applicable. One advantage of simple regression analysis is that the procedure is already understood by most economists.

The estimates presented in tables 4 and 5 are conservative in the opinion of the writer. Many people would predict even fewer farms for 1975 and 2000 than any of the above estimates. One's thinking is likely to be influenced most by more recent events and hence one might be more inclined to accept estimates based on the recent period, 1955–60. Even these estimates look fairly reasonable, although they indicate a more drastic decline in farm numbers than projections based on the longer periods.

Concluding Remarks

The procedure, as illustrated, has several shortcomings: (1) No statistical measures of reliability are available, (2) only net movements can be measured, and (3) for North Dakota more size categories in the census classifications would be desirable. The 1959 Census of Agriculture included an enumeration of farms of 2,000 or more acres. This is highly desirable since the projections indicate that eventually most of the farms in North Dakota will consist of 1,000 acres or more. With data on actual farm movements between size states much more useful information could be obtained.

The use of Markov chains to project future number of farms has several advantages over traditional procedures. The following are the most important:

1. Projections can be made more conveniently for each size category of farms.
2. The method provides other types of estimates which are not readily obtainable with traditional techniques.

Literature Cited

(*1*) Adelman, Irma G.
1958. A Stochastic Analysis of the Size Distribution of Firms. Jour. Amer. Statis. Assoc. 53: 893–904. December.

(*2*) Bostwick, Don.
1962. Yield Probabilities as a Markov Process. Agr. Econ. Res. 14(2): 49–56.

(*3*) Clawson, Marion.
1963. Aging Farmers and Agricultural Policy. Jour. Farm Econ. 45(1): 13–30.

(*4*) Hole, Erling, and James Vermeer.
1963. Wheat Growers' Machinery Costs, by Size of Farm, in Central North Dakota. U.S. Dept. Agr., Agr. Econ. Rpt. 24, 21 pp.

(*5*) Judge, G. G., and E. R. Swanson.
1961. Markov Chains: Basic Concepts and Suggested Uses in Agricultural Economics. Ill. Agr. Expt. Sta. Res. Rpt. AERR 49, 17 pp.

(*6*) Kanel, Don.
1963. Farm Adjustments by Age Groups, North Central States, 1950–1959. Jour. Farm Econ. 45(1): 47–60.

(*7*) McElveen, Jackson V.
1963. Farm Numbers, Farm Size and Farm Income. Jour. Farm Econ. 45(1): 1–12.

(*8*) Reiss, F. J., R. C. Hughes, and G. G. Judge.
1963. Changes in Farm Tenure: A Markov Process Analysis. Ill. Agr. Econ. 3(2): 9–16. July.

(*9*) Solow, Robert.
1951. Some Long-Run Aspects of the Distribution of Wage Incomes. Econometrica 19: 333–334. July.

(*10*) Tolley, G. S., and H. W. Hjort.
1963. Age-Mobility and Southern Farmer Skill—Looking Ahead for Area Development. Jour. Farm Econ. 45(1): 31–46.