AGRICULTURAL ECONOMICS RESEARCH

VOL. 23, NO. 4, OCTOBER 1971

Farm Size and the Distribution of Farm Numbers

By Robert F. Boxley[1]

Size-of-farm data for 1964 were fitted to the function $\ln Y = \ln a\text{-}bX$. Results showed that the percentage distribution of farms by size classes tends to follow the distribution of an inverse exponential function. Furthermore, empirical size distributions seem to have an underlying stability across time and geographic areas. These features have several applications, one of which is prediction of future size distributions of farms. A method of making such a prediction is illustrated with census data on farm numbers in 1935 and 1964.

Key words: Farms; distribution by size; Gini ratio.

One standard measure of the economic status of the farm sector is average size of farm. Despite the shortcomings of land area as a measure of economic well-being, changes in farm size are closely followed in both popular and technical farm literature. The doubling of average farm size between 1935 and 1964, for example, was classified by writers for the Bureau of the Census as "one of the significant developments in agriculture in the United States in the twentieth century" (*5*, p. 242).[2]

Despite the interest popularly attached to farm size measured in acreage, it is not immediately obvious how changes in this parameter should be interpreted or, indeed, whether much importance should be attached to it. Average farm size for the conterminous United States rose from 154.8 acres in 1935 to 350.8 acres in 1964 (while total acreage of land in farms remained about constant); Nikolitch and McKee note that one interpretation would attribute the change to "an ever-decreasing number of increasingly larger farm organizations" (*3*, p. 1549). Yet such a change in farm size could have been achieved in several ways: By the expansion of a relatively few 1935 farms into giant operations; by the outmigration of every other 1935 farm operator across all size classes (which would leave concentration, in the Lorenz curve sense, unaffected); or by outmigration of all of the smallest farm operators (requiring proportionately modest expansion by the farms remaining to absorb the land thus freed). In truth, elements of all three explanations appear involved in the farm size changes of 1935-64 and, depending upon which explanation is favored, the distributional consequences can be interpreted about as ominously or as auspiciously as one likes.[3]

This observation is, of course, generally true. Most statistical measures such as averages or medians have economic or social meaning only within some distributional context. Thus, we are interested in changes in farm size not in some absolute context but in relation, say, to access to farming opportunities or competitive structure of the industry. In a similar manner, we are interested in projected capital needs as they relate to special requirements of the very large or very small farms. Or, we may be less interested in explaining why median income is at some level than we are in explaining why a particular group, with incomes below that level, persists.

The purpose of this paper is to explore some of the statistical relationships involved in the concepts of average farm size and the distribution of farm numbers and, from these relationships, infer some distributional consequences of changes, past and future, in U.S. farm sizes. Farm size measures are given particular attention

[1] The patient assistance of Lynn Pollnow, Economic Research Service, in working out concepts used in this paper is gratefully acknowledged.

[2] Italic numbers in parentheses refer to the Literature Cited, p. 94.

[3] A significant part of the change in average size was largely the statistical consequence of a very high rate of outmigration by farmers in the smallest size classes. Another, smaller part of the growth was attributable to an absolute increase in the number of farms of 500 acres or more. Finally, there was a very general outmigration of farm operators (out of agriculture or into other size classes) across the remaining size classes of less than 500 acres. See, for example, (*4*). Because of this general outmigration, the relative distribution of farmland among farm operators in 1964 was not greatly changed from the distribution in 1935. The concentration ratio for the conterminous United States was 0.65 in 1935. It rose to 0.67 in 1940 and 0.70 in 1945 and remained at that approximate level, being 0.71 in 1964. (All calculations are based on data obtained from (*5*) or equivalent earlier census volumes.)

because physical limitations on possible configurations of farms within a finite space suggest that the distribution of farm numbers is characterized by a specific functional form. (There is some evidence that this particular function also describes a broader class of property and wealth distributions.) Finally, we illustrate a specific application of a distribution function as a means of predicting future farm size distributions.

Functional Distribution of Farm Numbers

Past processes of fragmentation and consolidation of farms have resulted in a distribution of farm sizes ranging from very small to very large acreages, even in States that were originally homesteaded in quarter- or half-section units. It is obvious, however, that these processes have not been completely random. In general, any change in the number of farms of a given size requires either a change in the land base or offsetting changes in other size categories. If the changes occur over a constant land base, the possible farm size combinations are physically constrained by that land base and by the fact that the maximum number of farm units that can be created of a given size is inversely related to that size. Additional constraints on possible size combinations are imposed as other parameters (e.g., total farm numbers, median farm size) of the distribution are specified.

The inverse relationship between frequency and farm size categories has led Folke Dovring (*1,2*) to suggest that a "normal" size distribution of farm numbers should resemble the inverse exponential function, e^{-x}. This function can be viewed as representing a decumulative size distribution by writing $Y = e^{-x}$, where Y equals the percentage remaining above a given size limit, x. At $x = 0, e^{-x} = 1.0$ (or 100, if interpreted in percentages). As the size limit increases ($x > 0$), values of the function decline smoothly, becoming infinitesimal in the vicinity of $x = 10$. Exponential functions, as the antilogarithm of a natural logarithm, plot as a straight line on semilogarithmic paper (figure 1). On logarithmic paper, the functions plot as a curve (figure 2). The latter representation is a particularly convenient form for graphic analysis of distributive phenomena.

One reason to consider farm size distributions as exponential functions is that some State distributions coincide with or closely follow the e^{-x} distribution. An example is the 1964 farm size distribution for Indiana, which is also plotted on figures 1 and 2.[4] An evaluation

[4] The distribution for Indiana was fitted by defining x as relative farm size and setting it equal to 1.0 at average size

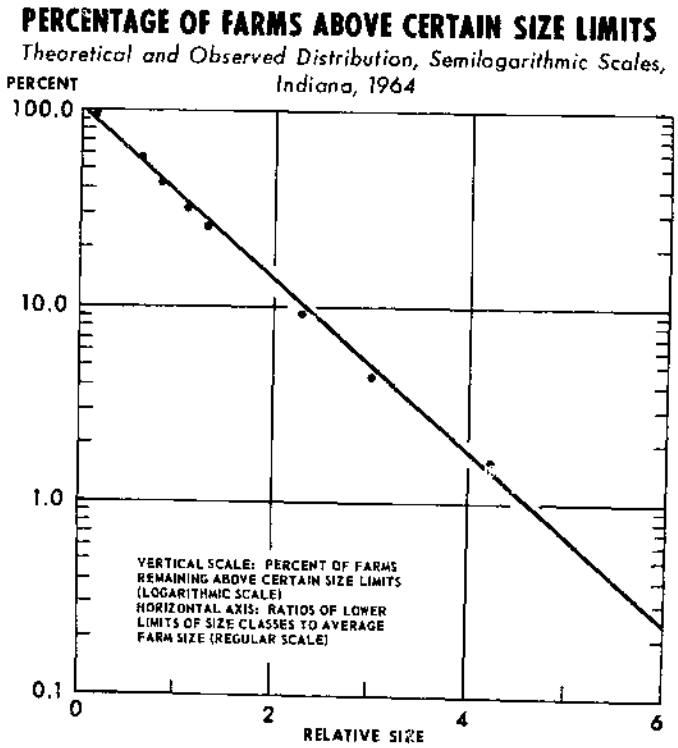

Figure 1

PERCENTAGE OF FARMS ABOVE CERTAIN SIZE LIMITS

Theoretical and Observed Distribution, Logarithmic Scale, Indiana, 1964

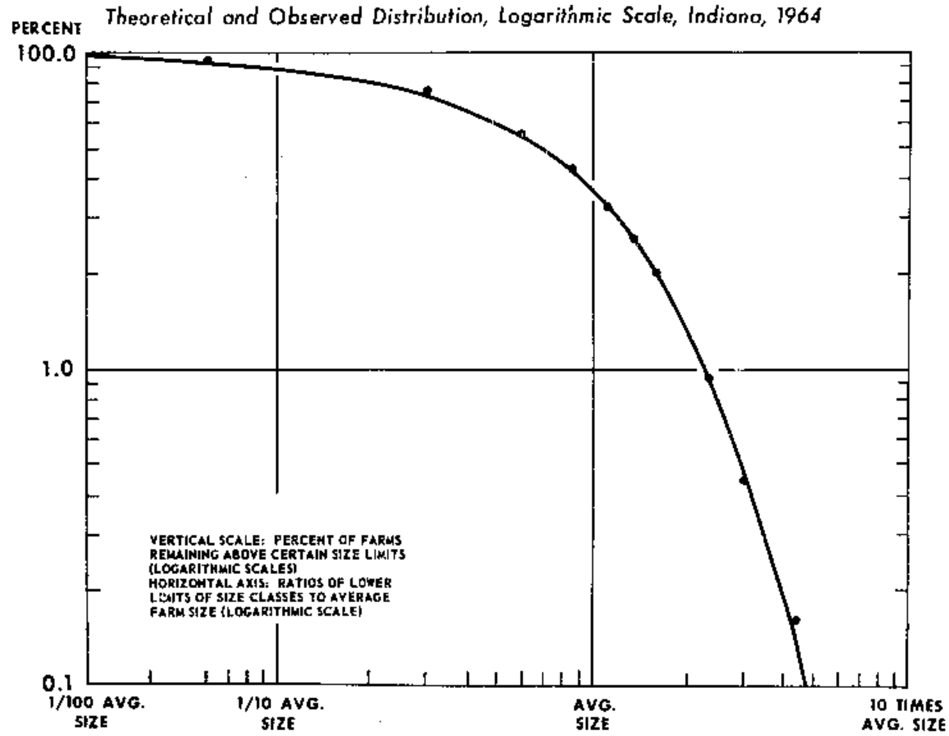

Figure 2

of the agreement between the exponential distribution and the 1964 farm size distribution for Indiana can also be obtained by comparing actual and indicated class frequencies (table 1). The largest difference between actual and estimated frequency (obtained by reading directly from a table of the inverse exponential distribution) occurs for the 50-to-99-acre class and is equivalent to underestimating by 1,800 the 22,600 farms in this class.

Indiana in 1964 illustrates a particularly close agreement between the exponential and empirical distributions of farms by size. Some other States—notably Maine, New Hampshire, New York, Pennsylvania, Ohio, Kentucky, Kansas, and Missouri—also closely follow (but do not exactly coincide with) the exponential distribution. While there may be no reason to expect exact coincidence,[5] viewing the exponential distribution as a

Table 1.—Actual and estimated percentage of farm numbers by size class, Indiana, 1964

Size class	Actual[1]	Estimated[2]	Difference
Acres	*Percent*	*Percent*	*Percent*
Under 10	4.5	5.8	- 1.3
10-49	19.0	20.1	- 1.1
50-99	20.9	19.2	+1.7
100-139	12.5	11.7	+0.8
140-179	10.1	9.2	+0.9
180-219	7.2	7.3	- 0.1
220-259	5.8	5.9	- 0.1
260-379	10.7	10.7	—
380-499	4.8	5.2	- 0.4
500-699	2.9	3.4	- 0.5
700-999	1.1	1.3	- 0.2
1,000 and over	0.5	0.2	+0.3

[1] Computed from 1964 Census of Agriculture data (*5*).

[2] Computed from a table of values for e^{-x} at selected values of x.

(165.9 acres). Points on the x axis were then located by expressing the lower limit of the census size categories as fractions or multiples of average size and plotted against the decumulative percentage distribution of farm numbers by size categories.

[5] In the distribution of $Y = e^{x}$, the median occurs at $x = 0.69$. Whenever data series are found in which the median is 0.69 of the average ($x = 1.0$), the rest of the series distribution tends to be identical with that of the inverse exponential function. The distributions of data series for which the median/mean ratios vary from 0.69 tend to vary also from the inverse exponential distribution in predictable ways (*2*, p. 3).

norm can still be instructive. In Indiana, for example, the positive discrepancy between the actual and estimated distribution in table 1 occurs in the range of 50 to 179 acres. This "heaping" may be observed generally for the Midwestern States and appears to be a residual from the original settlement patterns which favored farms of this size under the rectangular survey (*2*, p. 9). The negative discrepancies are then explicable as offsets. The positive discrepancy for the 1,000-acre-or-larger class, on the other hand, presumably reflects some other past or present force at work at this level.

There may also be reason to view the exponential distribution as a limiting distribution for farmland and farm numbers. States with empirical farm size distributions relatively close to an exponential form are generally found in the long-settled areas east of the Mississippi. This geographical distribution has led Dovring to advance the hypothesis that the resemblance of the size distribution to an exponential distribution comes about only over time and through the processes of farm consolidation and division which smooth irregularities associated with the original settlement patterns (*2*, p. 9). There have been very few studies of the dynamics of farm size changes, but Walrath's work (*6, 7*) indicates that the processes of farm consolidation and fragmentation are complex and that major but offsetting changes may occur simultaneously in a given area. From his studies, it is possible to see how the smoothing might come about.

Classes of Distributions

Although it is intriguing to compare empirical distributions with the inverse exponential distribution, it is probably more useful to view these distributions as members of a class that might be described as exponential-type distributions. Members of this class—which includes a number of measures of income or wealth in addition to farm size—are characterized by the general functional relationships plotted in figures 1 and 2 (i.e., linear in semilog, curvilinear in double-log), indicating that they are of the same family as the e^{-x} distribution.[6]

Dovring has done considerable work in classifying distributions and in developing transformations of empirical distributions as an analytical tool by which phenomena can be gaged relative to the function, e^{-x}. For the purposes of this paper, however, it may suffice to use some simple measures of farm size distributions. One means of doing this is to view the function:

$$Y = e^{-x}$$

as a special case of the class of functions:

$$Y = ae^{-bx}$$

where a and b both equal 1.0. By taking natural logarithms the general function can be expressed as:

$$\ln Y = \ln a - bX.$$

An appropriate measure of goodness of fit for this function is the simple correlation coefficient, r^2.[7]

Size-of-farm data for 1964 for the 48 conterminous States were fitted to the general function and the following distribution of r^2 was obtained:[8]

r^2 range	Number of States
0.980 - 1.000	10
.950 - .979	11
.900 - .949	11
.800 - .899	12
.730 - .799	4

The lowest r^2's were obtained from West Coast and Rocky Mountain States—Oregon, California, Washington, Arizona, Utah, and Nevada. For these States, the exponential-type distribution gives the poorest fit, but this perhaps should not be surprising, since the land-in-farm base is still expanding in these areas and the size distributions are relatively new. Thus,

[6] This statement is based partially on some cursory investigations of data from various sources but is largely drawn from Dovring's work which indicates that exponential-type distributions may characterize a wide range of income and wealth distributions and, thus, have a number of important analytical applications.

[7] In this paper we interpret r^2 as a measure of goodness of fit for the general function in an absolute sense (a perfect fit having an r^2 of 1.0). We have also compared r^2 for alternative functions fitted to the farm size distributions namely: $y = a + bX$ and $y = aX^b$. The arithmetic form gives the uniformly poorest fit. The logarithmic form fits best in those States where the semilogarithmic form fits least well, but it is generally inferior to the semilogarithmic function.

[8] The empirical farm size distributions tend to depart significantly from linearity at the upper limits of the distribution (approximately, values of $x > 10$). To maintain comparability in the regression measures for Eastern and Western States, the State regressions were computed using only values of $x < 10$. Practically, this means that the upper limit of the size distribution for States east of the Mississippi was the class of 1,000 acres or more and, for States west of the Mississippi, the class of 2,000 acres or more.

the experiences in the Western States do not necessarily contradict broader statements of functional size relationships. Florida was next in the ranking of States with low r^2 (0.853), and it too has experienced recent increases in its land-in-farm base.

The next group of distributions with low r^2's include the States of Louisiana, Mississippi, Alabama, and South and North Carolina. The regionalism of this grouping suggests that the relatively poor fit (all r^2's less than 0.90) may reflect some remnant of former sharecropping and plantation systems. The best fits (highest r^2's) were found, generally, in the Midwestern and Northeastern States.

These patterns reflect another characteristic of the farm size distributions: A tendency for State distributions to fit into groupings on a geographic or regional basis.[9] The regional similarities can be noticed in graphic comparisons, as well as in the regression parameters for the individual State distributions. One possible set of regional groupings, put together from consideration of both sources, is listed in table 2. This grouping varies in several instances from more commonly used regionalizations based on type of farm or other geographic considerations.[10] Presumably, these regional similarities reflect the common influence of factors such as time and pattern of original settlement, topography, and various institutional factors.

The general form of the exponential distribution was also fitted to the regional groupings (table 2). The r^2's indicate the strength of the regional associations, while the intercept and slope coefficients provide measures of regional differences in the farm size distributions. These regressions were computed with logarithms to base 10. By way of comparison, a regression of $Y = e^{-x}$ to a $\log_{10}$ base, scaled in the same manner, should have an r^2 of 1.00, an intercept of 2.00 ($\log_{10}$ 100), and a slope coefficient of -0.434 ($-1.0 \log_{10} e$).

[9] For simplicity we report only work using State distributions as the basic unit of observations. In practice, better fit on a unit or regional basis can be obtained by using counties as the basic unit of observation and splitting States on the basis of known intra-State differences in types of farming or farm organization, topography, or other features. For example, fits for several Great Plains States can be improved by an east-west split reflecting the transitional nature of agricultural production in the region.

[10] Our criterion was to arrange the States in a "reasonable" way to minimize the number of regions while maintaining contiguity within regions.

Table 2.—Regional groupings of farm size distributions and measures of fit by regression analysis, 1964

Region and States	r^2	Intercept	Beta (standard error)
New England (Maine, N.H., Conn., Mass., R.I.)	0.970	1.830	-0.314 (.018)
Northeast and Lake (Vt., N.Y., Pa., Ohio, Mich., Ind., Ill.)	.988	1.965	- .414 (.015)
Mid-Atlantic (N.J., Del., Md., Va., N.C.)	.929	1.683	- .241 (.022)
Southeast (S.C., Ga., Ala., Miss., La.)	.890	1.681	- .257 (.030)
Florida	.853	1.632	- .426 (.059)
Appalachia (W.Va., Ky., Tenn.)	.943	1.739	- .270 (.022)
Upper Central (N. Dak., Minn., Wis., Iowa)	.939	1.870	- .328 (.026)
Lower Central (S.Dak., Nebr., Kans., Mo., Okla., Ark.)	.942	1.868	- .350 (.028)
Texas	.894	1.803	- .451 (.049)
Mountain (Mont., Idaho, Wyo., Colo.)	.914	1.871	- .593 (.058)
Arid (Nev., Utah, Ariz., N.Mex.)	.780	1.746	-1.076 (.181)
West Coast (Calif., Oreg., Wash.)	.776	1.547	- .259) (.044)

Distributional Shifts, 1935-64

Analytically, one potentially useful characteristic of the State distributions of farm numbers is the relative stability of many of these distributions over time. The lack of major shifts in the size distributions is especially remarkable in view of rates of change and outmigration that, in New England for example, resulted in only a fourth as many farms in 1964 as existed in 1935. Several States had size-of-farm distributions in 1964 that were virtually unchanged from their 1935 distributions. This was true of States that have experienced only small changes in their land base, notably Illinois, Iowa, and North Dakota, as well as some Northeastern States–New Hampshire, New York, and Vermont–which experienced substantial declines in their land-in-farm base

Most States did experience some shifts in the distribution over the three decades–as evidenced by the previously noted increase in concentration ratios and in absolute numbers of farms of more than 500 acres. On the other hand, part of the aggregate change in both of these measures was due to the westward migration of farm production and farm numbers that also occurred during the three decades. In table 3, we have attempted to quantify these distributional shifts, on a regional basis, between 1935 and 1964. For this table the 1964 distribution of farm numbers was estimated from the 1935 curves, using known 1964 total farm numbers and average size. The regions in table 3 correspond generally to those in the 1964 Census of Agriculture.[11]

The estimated distributions by regions in table 3 indicate how the actual number of farms in 1964 would have been distributed if average farm size in each region were at the 1964 level and farm numbers were functionally distributed as in 1935. This implies that any changes in the land base occurred in such a way as to leave the 1935 distribution unchanged.[12]

One implication of table 3 is that a researcher in 1935, who correctly predicted 1964 regional farm numbers and average size, could have further estimated 1964 size distributions within a range of 88 to 128 percent of actual class numbers for the conterminous United States–an average error of estimate of 12.4 percentage points.[13] By way of contrast, a "naive"

[11] Estimates were made on the basis of the previously identified regional groupings and then aggregated to the three census regions. This resulted in New Jersey being included in the South rather than the North.

[12] The land-in-farm bases for 1964 were 94, 92, and 146 percent of the base in 1935 for the North, South, and West, respectively.

[13] Most of the difference between actual and estimated class numbers is due to shifts in the distribution between 1935 and 1964. However, some interpolation error between adjacent classes may have been generated in reading from the 1935 curve (the estimates were derived using graphical techniques).

Table 3.–Actual and estimated number of farms by size category, conterminous United States, 1935 and 1964, and regions, 1964

Size class	Farm numbers									Ratio, estimated to actual[1]			
	United States			North		South		West					
	Actual 1935	Actual 1964	Estimated	Actual 1964	Estimated	Actual 1964	Estimated	Actual 1964	Estimated	U.S.	North	South	West
	Thou.	*Thou.*	*Thou.*	*Thou.*	*Thou.*	*Thou.*	*Thou.*	*Thou.*	*Thou.*	*Pct.*	*Pct.*	*Pct.*	*Pct.*
0-99	4,138	1,358	1,197	440	376	769	696	149	126	88	85	90	84
100-139	754	325	337	163	173	145	147	17	16	104	106	101	98
140-179	684	308	286	187	169	104	109	17	11	93	90	104	64
180-219	294	191	245	116	146	66	83	9	10	128	126	125	112
220-259	212	164	187	109	114	47	65	8	12	114	105	138	146
260-379	473	297	348	192	209	89	118	17	21	117	108	134	123
380-499		154	190	97	113	46	61	11	15	123	117	132	143
500-699	167	124	137	69	80	42	42	12	16	111	115	99	130
700-999		87	90	45	45	30	27	12	17	103	101	89	145
1,000 and over	89	145	137	51	44	46	37	48	56	95	87	80	116
Total[1]	6,812	3,153	3,153	1,469	1,459	1,383	1,383	300	300	–	–	–	–

[1] Computed from unrounded data.

projection of the 1964 size-of-farm distribution at the same percentage distribution as in 1935 yields an average error of 38.4 points and, more importantly, completely fails to anticipate either the large decline in numbers of small farms or the increase in the very large farms.

Divergencies between actual and estimated farm numbers by size classes provide a general indication of regional shifts in the size distribution over the three decades and are consistent with the earlier observation that changes have occurred mainly at the extremes of the distribution. The actual numbers of farms in 1964 of less than 100 acres or more than 1,000 acres were more than expected on the basis of functional estimates derived from the 1935 distribution, and the numbers of farms in the intermediate classes were consequently fewer. The only regional exception to this pattern was in the West, where a significant part of the increase in average size came about through additions to the land base. Excluding the West, however, the net differences between the actual and estimated 1964 distributions are remarkably small. In the North, for example, only 89,000 farms were misclassified on net[14]—less than 6 percent of the 1.5 million farms of the region in 1964. Results were nearly identical in the South, involving 85,000 of a total of 1.4 million farms.

Future Size Distributions

As the above exercise suggests, one useful application of knowledge about the current size distributions of farms is the estimation of probable future size distributions. For any given unit of observation (county, State, region), this requires, basically, estimating the expected number of farms and land in farms and then extrapolating the future distribution from the current one. Based on past changes, we may have more confidence in projections for some regions than others but, in general: "It remains a sound proposition to say that if the same kind of economic and related forces are at work in the future as in the past, further development over the foreseeable future should be such that it could be projected by extrapolating the experience of recent past. In the projection of farm size *distribution*, it is not even necessary to pin down any particular year when such a structure will have taken the place of the present one. Assuming that some time in the near future farm numbers will have declined to the point where a certain average size has been attained; it is then possible to project, approximately, how farms and farmland will be distributed by size classes at that time" (*1*, p. 8).

As an example of this application, we have extrapolated from the aggregate U.S. distribution of 1964, two possible future distributions when, it is assumed, farms will average 500 and 700 acres (table 4). Aggregate farm numbers would be 2.2 and 1.6 million, respectively, assuming that total farm acreage remains in the vicinity of 1.1 billion acres.

The "potential distributions" of table 4 illustrate one way in which the assumed farm-size increases can be accommodated within the present distributional framework. Under ceteris paribus conditions, the extrapolation indicates that another doubling of average farm size for the Nation can occur without either the complete disappearance of small farms or the overwhelming dominance of giant-sized farms. Farms in the medium-size categories could continue in the majority (albeit of a much diminished number). The doubling of average size, given the assumed distribution, could be effectuated largely through a continued decline (but not disappearance) in the number of farms of less than 500 acres, and would entail only a moderate increase in the number of farms of more than 1,000 acres.

The above exercise is not a specific projection since we do not take into account even obvious regional differences in farm size distributions, nor have we considered the difficulties of projecting farm numbers or land in farms. As a practical matter, we would more likely have reason to project farm numbers or land in farms at some point in time (rather than a projection of a distribution per se). Nevertheless, many projected trends or trend changes may carry important distributional consequences and this technique would seem to provide one useful means for specifying them.

Conclusions

The percentage distribution of farms by size classes during 1935-64 remained relatively stable despite large increases in average size of farm. The rapid increase in farm size can be explained primarily by the outmigration of farm operators over a wide range of size classes (with subsequent consolidation of the agricultural lands thus released) and—of lesser importance—by a westward shift in farm numbers and the agricultural land base.

Work with the exponential-type distribution function indicates that farm size distributions—at least at the State level—are characterized by an underlying regularity. This is illustrated by both (a) the relative stability of the actual distributions over time for a large number

[14] This is the sum of the differences between estimated and actual numbers for either all classes where actual > estimated or all classes where actual < estimated.

Table 4.–Land in farms and number of farms by size, conterminous United States, 1964 and projections

Item	1964 (average farm size 350.8 acres)	Potential distributions if average farm size becomes–	
		500 acres	700 acres
Land in farms . . . . 1,000 acres	1,105,866	1,100,000	1,100,000
Number of farms . . . . farms	3,152,611	2,200,000	1,571,000
Size class:			
Under 10 acres . . . . farms	179,967	99,000	53,400
10-49 acres . . . . do.	635,824	354,200	198,000
50-99 acres . . . . do.	542,157	277,200	168,200
100-139 acres . . . . do.	324,543	204,600	103,700
140-179 acres . . . . do.	308,104	162,800	100,600
180-219 acres . . . . do.	191,199	162,800	122,600
220-259 acres . . . . do.	164,151	121,000	72,300
260-499 acres . . . . do.	451,144	426,800	380,300
500-999 acres . . . . do.	210,378	233,200	199,600
1,000-1,999 acres . . . . do.	84,971	96,800	97,400
2,000 acres and over . . . . do.	60,173	61,600	75,400

of States and (*b*) the apparent smoothing over time of the distribution toward an exponential-type curve for other States. This tendency toward stability suggests a number of useful analytical and predictive applications. The projection of future size distributions is one example. The stability also suggests that a certain amount of determinism may exist in the distribution of land, conditioned on the initial distribution. Perhaps this may hold also for other forms of wealth. Finally, it may be possible to use the technique to evaluate the impact of exogenous factors, such as farm programs, on the size distribution of farm numbers.

Literature Cited

(1) Dovring, F. Farm size data: Frequency distribution, interpolation, and projection. Dept. Agr. Econ., Univ. Ill. Col. Agr., Res. Rpt. AERR-50, May 1962.

(2) ______ Income and wealth distributions: The exponential functions. Dept. Agr. Econ. Univ. Ill. Col. Agr., AE-4212, June 1969.

(3) Nikolitch, R., and D. E. McKee. The contribution of the economic classification of farms to the understanding of American agriculture. Jour. Farm Econ., Vol. 47, No. 5, Dec. 1965.

(4) Reinsel, R. D. Changes in farm size. Farm Real Estate Mkt. Devel., U.S. Dept. Agr., Econ. Res. Serv., CD-69, June 20, 1967.

(5) U.S. Bureau of the Census. U.S. Census of Agriculture, 1964: Size of farm. Statistics by Subjects, Vol. II, Ch. 3, 1967.

(6) Walrath, A. J. Rural land ownership and industrial expansion. Va. Agr. Expt. Sta., Va. Polytechnic Inst., Bul. 527, May 1961.

(7) ______ Rural land ownership and economic development of a three-county area. Res. Div., Va. Polytechnic Inst., Bul. 10, Sept. 1967.