FACTORS RELATED TO FARMERS' USE OF CREDIT: LEAST-SQUARES ANALYSIS OF SAMPLE SURVEY DATA

BY

EMANUEL MELICHAR

Reprinted from JOURNAL OF FARM ECONOMICS
Vol. 47, No. 5, December, 1965

Factors Related to Farmers' Use of Credit: Least-Squares Analysis of Sample Survey Data

EMANUEL MELICHAR

THE 1960 Sample Survey of Agriculture marked the first time that farm operators in a representative national sample were asked to report all their outstanding farm debts.[1] Each operator was also asked about such factors as his age, tenure, and size of farm, thereby permitting study of the relationships between these characteristics and various aspects of farm indebtedness. For instance, one might examine how each factor was related to the likelihood that a farmer had debt, to how heavily indebted he was, or to the likelihood that he owed a particular kind of debt or was indebted to a particular lender group. This paper discusses the application of least-squares analysis to the first of these areas.

The most direct approach to determination of the relationship between any single factor and the presence of debt is to sort the farmers by the different values or classes of that factor and then calculate the percentage who were indebted in each class. As is well known, however, this simple approach often cannot provide the desired knowledge because other farm characteristics also affect the likelihood of indebtedness, and there are unequal proportions of farmers with these other characteristics in each of the classes of the factor being studied. We therefore wish to perform an analysis that takes account of these intercorrelations and so yields what might be termed a "net relationship" between a factor and the probability that an operator was indebted, as contrasted to the "gross relationship" obtained in the initial simple calculation.

The usual procedure in such investigations is cross-classification, but unfortunately this procedure has serious practical limitations. Most basically, when several correlated factors are to be studied, even a large sample may not permit cross-classification by all of them simultaneously. The analyst must settle for study of a series of two-way or three-way tables, but if more than three correlated factors are involved, the results of such study obviously may be inconclusive. Furthermore, the research report is often tedious because the analysis does not produce summary statistics that describe the nature of the net relationships, but rather

[1] *Farm Debt—Data from the 1960 Sample Survey of Agriculture,* Washington, D.C., Board of Governors of the Federal Reserve System, 1964, and *1960 Sample Survey of Agriculture,* 1959, Vol. V, Part 5, Spec. Rep., Washington, D.C., Bureau of the Census, 1962.

EMANUEL MELICHAR *is a member of the staff of the Board of Governors of the Federal Reserve System.*

forces the analyst to evaluate many rows and columns of data pertaining to a single factor.

There is, however, an alternative least-squares regression method that tends to overcome these difficulties. In this procedure, each class of each factor to be studied is represented by a separate independent variable in a regression equation; six types of farms, for instance, are represented by six variables. The coefficient estimated for each variable measures the net effect of membership in that class on the dependent variable. The group of coefficients estimated for a single factor such as type of farm provides a concise description of the net relationship between that factor and the dependent variable. Finally, a coefficient of partial determination can be calculated to measure the importance of each factor, and an F-ratio calculated to test its statistical significance.

This technique has been employed extensively in the biological sciences.[2] It was recently introduced to economists by Suits[3] and has since been employed in several published studies, the most extensive work being by Morgan et al.[4] Among the econometrics texts, the subject is treated by Goldberger[5] and Johnston.[6]

The mechanics of the regression approach are not difficult. The independent variable assigned to each class of a factor such as geographic area is given a value of either 1 or 0 for each sample farm, depending on whether the farm does or does not belong in the class represented by the variable. The dependent variable may be numeric, such as the amount of debt, or may also be a dummy variable—in our case, either 1 or 0, depending on whether the farmer did or did not have debt. After coding the observations in this fashion, one wishes to estimate an equation such as

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4Z_1 + b_5Z_2 \quad (1)$$

in which Y is the presence of debt and, as an example, the three X's represent three geographic areas into which the nation is divided, and the two Z's represent the farmers who did and who did not use production contracts. For technical reasons, this equation cannot be estimated without imposing a constraint upon each of the two groups of independent vari-

[2] Walter R. Harvey, *Least-Squares Analysis of Data with Unequal Subclass Numbers*, USDA ARS-20-8, July 1960, and E. L. LeClerg, *Mean Separation by the Functional Analysis of Variance and Multiple Comparisons*, USDA ARS-20-3, May 1957.

[3] Daniel B. Suits, "Use of Dummy Variables in Regression Equations," *J. Am. Stat. Assn.* 52:548–551, Dec. 1957.

[4] James N. Morgan, Martin H. David, Wilbur J. Cohen, and Harvey E. Brazer, *Income and Welfare in the United States*, New York, McGraw-Hill Book Co., 1962.

[5] Arthur S. Goldberger, *Econometric Theory*, New York, John Wiley and Sons, Inc., 1964, pp. 173–177, 218–231, 248–255.

[6] J. Johnston, *Econometric Methods*, New York, McGraw-Hill Book Company, 1963, pp. 123–130, 221–228.

ables. Constraints that lead to coefficients that can be readily interpreted are as follows:

$$(2) \qquad x_1b_1 + x_2b_2 + x_3b_3 = 0$$

$$(3) \qquad z_1b_4 + z_2b_5 = 0$$

in which x_1 is the number of farms in the class represented by X_1, x_2 the number in the class represented by X_2, and so forth. Solving (2) for one of the coefficients, such as b_3, doing the same for b_5 in (3), and substituting the results into (1) yields:

$$(4) \qquad Y = a + b_1\left(X_1 - \frac{x_1}{x_3} X_3\right) + b_2\left(X_2 - \frac{x_2}{x_3} X_3\right) + b_4\left(Z_1 - \frac{z_1}{z_2} Z_2\right).$$

Equation (4) contains only three independent variables, each of them being a transformation of the original independent variables in equation (1). Equation (4) can be estimated by a standard regression program, and then b_3 and b_5 are obtained by substituting the results into (2) and (3), respectively. Because of the particular constraints used, the constant term in (4) is equal to the mean of the dependent variable; in our case, it is .64, the national average proportion of farmers with debt. The coefficient estimated for each independent variable is therefore readily interpreted as the net difference from that mean which is associated with membership of a farm in the class represented by that independent variable.

Measures of the importance and significance of the net contribution of a factor toward explanation of the variation in the dependent variable can be obtained by calculating the coefficient of determination for equation (4), then calculating the coefficient of determination for (4) after omitting the group of independent variables that represents the factor being studied, and finally employing this information in the formulas given by Johnston[7] to obtain a coefficient of partial determination and an F-ratio. Goldberger[8] has demonstrated that these results are the same as those that would have been obtained through more usual analysis-of-variance methods.

In the study reported here, data were prepared as described above for the 8,891 commercial farms in the national sample. Each sample farm represented from 1 to 900 actual farms and was replicated that many times in the analysis, but significance tests were based on the sample number.

[7] Johnston, pp. 123–130.
[8] Goldberger, pp. 227–231.

The 12 factors listed in Table 1 were used in studying the probability that a farmer had debt. First, a separate equation was estimated for each of the factors taken individually, and these gross relationships were all found to be significant. Next, all 12 factors were included in a single equation, being represented therein by 50 independent variables. The R^2 for this equation is .127, and the equation F-ratio of 25 is significant at the .01 probability level. The equation was then re-estimated 12 times, with one factor omitted each time, to obtain the partial determination coefficients and F-ratios shown in the last two columns of Table 1. The latter indicate that 10 factors exhibit a net relationship significant at the .01 probability level.

Table 1. Importance of selected characteristics in explaining presence of debt among commercial farm operators in December 1960

Characteristics	Gross relationship		Net relationship	
	R^2	F-ratio	Partial R^2	F-ratio
Years on farm	.030	40**	.015	20**
Age of operator	.049	91**	.015	27**
Tenure of operator	.018	53**	.012	35**
Value of farm operated	.036	55**	.008	12**
Type of farm	.013	24**	.006	11**
Per-acre value of farm	.003	8**	.005	11**
Value of annual sales	.042	78**	.005	8**
Off-farm work	.011	20**	.005	8**
Off-farm income	.004	7**	.002	3**
Race of operator	.004	16**	.001	5**
Geographic area	.012	53**	.001	4*
Use of production contracts	.002	17**	.000	3

* Significant at probability level of .05.
** Significant at probability level of .01.
Note: In calculating the above statistics, complications arising from stratification and clustering in the sample design and from the presence of heteroscedasticity were ignored.

The nature of the more significant of these relationships is illustrated in Fig. 1. The gross relationships are charted on the left, and on the right are plotted the regression coefficients obtained for the equation that contained all 12 factors. These coefficients represent net differences from the national average proportion of operators with debt (.64).

The handling of numerical independent variables should be noted. For instance, age was represented in the equations by six variables, one for each of the five class-intervals into which the range of ages was divided and the sixth for "age not reported," which allowed these farmers to be retained in the analysis. Also, the functional form of the relationship did not have to be a subject of experimentation, as the optimum transformation, within the restrictions imposed by the class intervals used, is in effect indicated by the coefficients themselves.

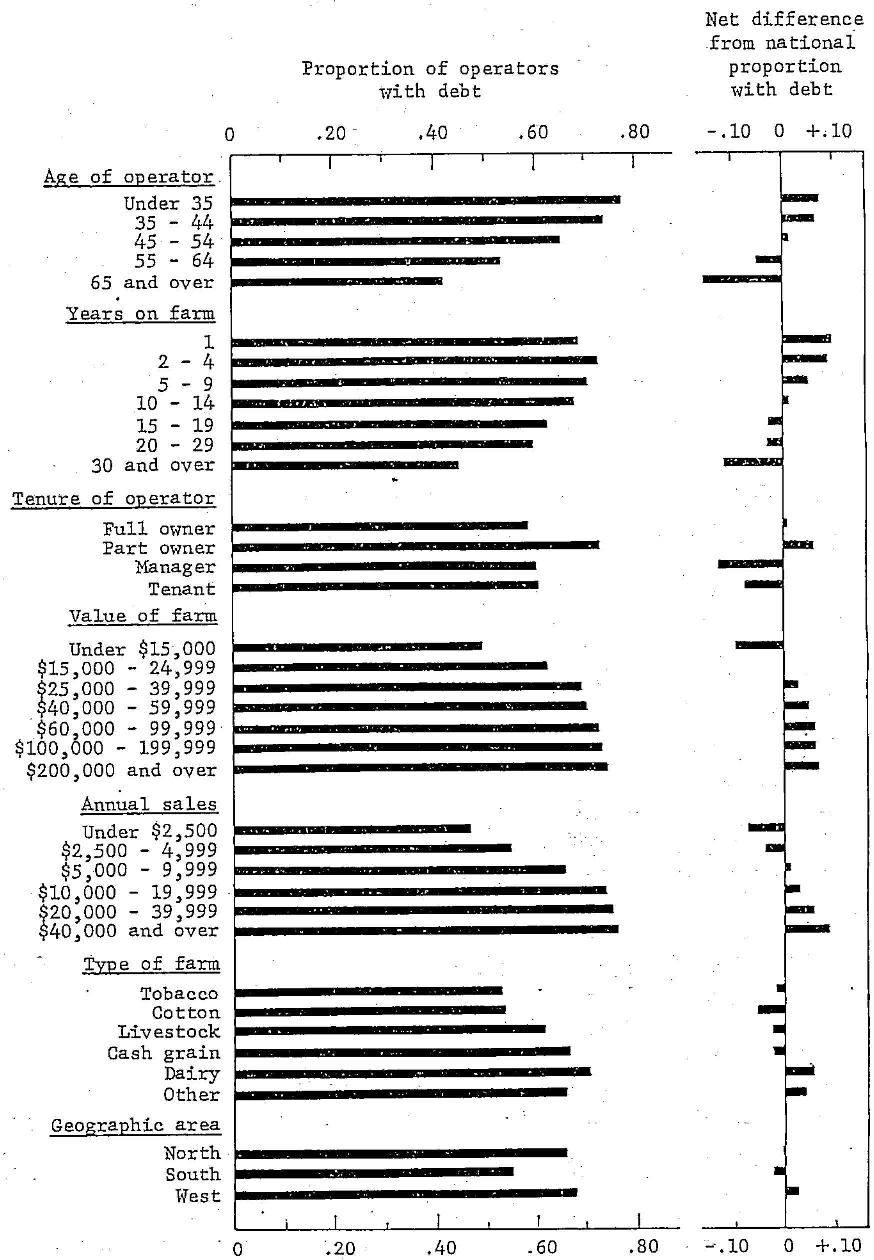

Figure 1. Gross and net relationships between selected farm characteristics and the probability that a commercial farm operator was indebted in December 1960

A discussion of the reasons for and the implications of the findings reported here is beyond the scope of this paper, and will be presented elsewhere along with the results of similar analyses of other aspects of farmers' indebtedness. It is evident, however, that the analytical procedure has yielded clear and concise information upon which such further discussion can be based. As such, it appears to be a useful addition to the tools of survey analysis.