Finance

1976

EXPLAINING FARM OPERATORS DEBT: AN APPLICATION OF THE AUTOMATIC INTERACTION DETECTOR TECHNIQUE

By

David A. Lins
Timothy R. Donaldson

Submitted as a
Contributed Paper
Summer AAEA Meetings
University Park, Pennsylvania
August 15-18, 1976

Introduction

Financial intermediaries and policy makers are interested in understanding how socio-economic factors affect the demand for farm loans. Special Census surveys and other publications provide cross-tabulations between farm operators debt and a multitude of variables which are believed to influence the amount of such debt. However, correlation and interaction among the variables make it extremely difficult to determine from cross-tabulations the importance of these variables in explaining variations in the level of debt. Empirical approaches such as regression, factor and analysis, and the automatic interaction detector (AID) technique provide a means of dealing with correlation and/or interactions among variables.

Most empirical analyses of factors associated with farm debt have been based upon time series analyses (Herr, Lins). These studies have concentrated on explaining the aggregate amount of farm debt and how it changes over time. However, knowledge of factors influencing the amount of debt of individual farm operators in also important to decision makers. Cross-sectional studies of this nature appear to be less common.

The purpose of this paper is to discuss what factors influence the amount of debt of individual farm operators in the U. S. The analysis relies on applications of the (AID) technique to cross-sectional data. Results reported here are based upon the work of Donaldson.

Source of Data

Data used in this study are from the 1970 Survey of Agricultural Finance, conducted early in 1971 by the Bureau of Census, to cover the calendar year 1970. For this study the data were from report form (69-A9.1), which recorded all outstanding farm debts for farm operators. In addition, it provided information regarding such factors as the operator's age, race, size of farm, operating expenses, value of land and buildings, and so forth. This permits an examination of the relationship between these characteristics and various aspects of farm indebtness.

Methodology

The methodology selected for use in this study was the automatic interaction detector (AID) technique developed by Sonquist and Morgan. The AID technique is basically a searching process. A group of independent variables, selected on the basis of theory, are subjected to the searching process to determine the best model through statistical inference. The technique does not provide statistical tests of the significance of variables.

AID is a technique for determining what variables and categories within them combine to produce the greatest discrimination among group means. The program divides the sample, through a series of binary splits, into mutually exclusive subgroups. The group means account for more of the total sum of squares of the dependent variable than the means of any respectively associated pair of subgroups.

Independent variables must be specified in coded intervals (classes) with a maximum of 31 class intervals per variable. The dependent variable is assumed to be continuous.

In operation the program splits the parent group into two subgroups which provide the greatest reduction in error sum of squares for the dependent variable. This is accomplished in the following four steps: (See Figure 1).

(1) Compute the total sum of squares (TSS) for the parent group.

(2) For each independent variable, find the division of classes that results in two non-overlapping subgroups which provide the largest reduction in unexplained variance, i.e. a binary split which maximizes the between sum of squares (BSS).

(3) Divide the BSS for each independent variable by the TSS for the parent group.

(4) Choose that independent variable with the largest BSS/TSS ratio. The initial iteration is completed when the sample is split into two groups, accordingly.

Once the parent group is split into two groups, the program treats each subgroup as a separate population and the same process is repeated. It is possible that the program will split on the same independent variable more than once.

The number of iterations in the program is controlled by the following three stopping rules which constrain the number of subgroups formed:

(1) Minimum sample size -- Each subgroup must contain a minimum sample size to be eligible for further splitting.

(2) Split eligibility -- A subgroup must contain a minimum percentage of the total original sum of squares if it

Figure 1: Diagrammatical Procedures Used in the Automatic Interaction Detector Technique

Step (1) Parent Group

Compute the total sum of squares (TSS) for the parent group. $TSS = \sum_{i=1}^{N} (Y_i - \bar{Y})^2$

Step (2) Variable X_1 [1, 2 | 3, 4, 5, 6, 7]

For each independent variable split at the point where the between sum of squares (BSS) is maximized

$= BSS_1$

Variable X_2 [1, 2, 3 | 4, 5, 6, 7, 8]

$= BSS_2$

$$BSS = n_1 \bar{y}_1^2 + n_2 \bar{y}_2^2 - N\bar{y}^2$$

where

$n_1 + n_2 = N$

Variable X_n [1 : : : | : : 31]

$= BSS_n$

$y_1 + y_2 = y$

Step (3)

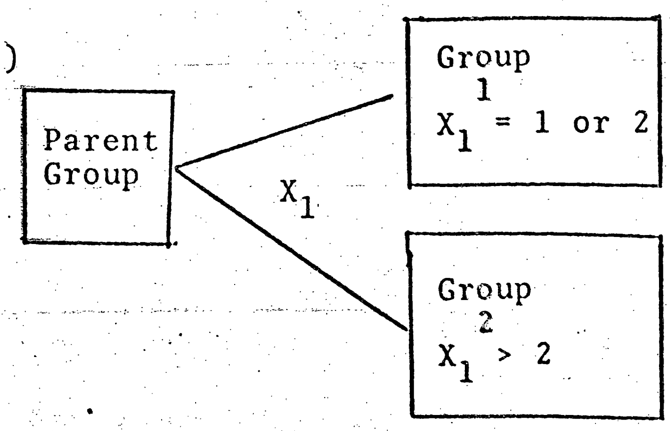
$BSS_1/TSS = .19$

$BSS_2/TSS = .07$

$BSS_n/TSS = .09$

Compute the ratio of BSS/TSS

Step (4)



Split the sample accordingly

is to be eligible for further splitting. This requirement prevents subgroups with little variation from being further split.

(3) Split reducibility -- The between sum of squares for the i_{th} group has to be a minimum percentage of the total original sum of squares. This criterion is applied when none of the independent variables in the group sufficiently reduce the unexplained variance.

Independent variables may be entered in the AID program as either "free" or "monotonic", depending upon whether the researcher desires to allow the ordering of the independent variables to be rearranged or maintained during the partioning process. When specified as "monotonic", independent variables will have the order of their coded values (0, 1, 2...31) maintained during the partition scan. An independent variable which is designated as monotonic is assumed to have either a monotonically increasing or decreasing relationship with the dependent variable.

When the independent variable is specified as "free" the order of class values for it may be rearranged to find that partition which maximizes the sum of squares between the subgroups formed. An independent variable which is designated as free is assumed to have a non-monotonic relationship with the dependent variable.

There is an option in AID, which permits the research to exclude or include extreme cases. Extreme cases are defined in terms of the number of standard deviations from the mean of the group (either parent or subgroup) in question.

The least-squares criterion used in AID (and other empirical techniques) is very sensitive to extreme cases. Extreme cases often involve either errors of measurement or conceptual problems. Genrally, one does not want his findings to be dominated by a few extreme cases. The exclude outliers option was used in this study to omit observations beyond 10 standard deviations from the mean.

Variables Used

The choice of variables for inclusion in this study was restricted to those available from Census report forms. Theoretical determinants of demand for loan funds were considered in selecting independent variables. Independent variables used in this study are shown in Table 1. In addition, the table shows the manner in which these variables were entered in the AID program (monotonic versus free) and their hypothesized relationships to the dependent variable. The independent variables were subjectively grouped into six major categories: (a) geographic, (b) size measured in physical or value terms, (c) expenditures, (d) income, (e) demographic and (f) structural.

AID Results

Three different AID runs were made. In the first run the total debt of U.S. farm operators was analyzed. In the second and third runs total real estate debt and total nonreal estate debt respectively were analyzed.[1/] Tables 2, 3, and 4 present the results of these three AID runs. In addition, Figure 2 shows a tree diagram of the results for the total debt run.

1/ Stopping criteria specified for these runs were: (a) minimum sample size = 50, (b) split eligibility = 0.2%, (c) split reducibility = 0.8%.

Variables Selected

For the run on total debt, the original sample size was 24,658.[2/] The amount of expenditures paid by the operator (V_{24}) was the first variable selected to split the parent group (Group 1). The parent group was divided into two subgroups; Group 2 contained 15,810 observations consisting of farm operators who had less than $40,000 in expenditures paid, while Group 3 contained 8,823 observations consisting of farm operators who had $40,000 or more in expenditures paid. This split accounted for 25.6 percent of the variation in the original sample.

Next the program split group 3 into groups 4 and 5 based upon the total value of land and buildings owned by the operator (V_4). Subsequent splits were based upon size of farm, economic class of farm, total operating expenses and total capital purchases.[3/] Nine final groups were formed and 46.6 percent of the variation in the dpendent variable was accounted for.

For the real estate debt run, only three explanatory variables were found to be important; total value of land and buildings owned by operator, total operating expenses and size of farm (Table 3). Approximately 42 percent of the variation in the dependent variable was explained. Seven final groups were formed.

Five variables were important in explaining variations in nonreal estate debt. Seven final groups were formed and roughly 40 percent of the variation was explained. Unlike the previous two runs, type of

2/ Each observation was weighted by the inverse of the rate of sampling for observations in that category.

3/ Debt as measured in the survey is related to the land owned and operated, whereas V_4 includes only the value of land and buildings owned. Size of farm is based upon the Census definition of "acres in this place."

TABLE 1

DESCRIPTION OF VARIABLES, THE TYPE OF PREDICTORS AND HYPOTHESIZED RELATIONSHIP WITH THE DEPENDENT VARIABLE

	Independent Variable	Type of Predictor	Hypothesized Relationship With Dependent Variable	Type of Variable
X_1	Region	Free		Geographic
X_3	Acres Owned by Operator	Monotonic	Positive	Size
X_4	Total Value of Land and Buildings Owned by Operator	Monotonic	Positive	Size
X_5	Acres Rented from Others	Monotonic	Positive	Size
X_6	The Value of Land and Buildings Rented from Others	Monotonic	Positive	Size
X_7	Acres Rented to Others	Monotonic	Positive	Size
X_8	Value of Land and Buildings Rented to Others	Monotonic	Positive	Size
X_9	Acres in This Place	Monotonic	Positive	Size
X_{10}	Total Purchases of Land and Buildings	Monotonic	Positive	Expenditure
X_{11}	Cash Purchases of Land and Buildings	Monotonic	Positive	Expenditure
X_{14}	Total Capital Purchases	Monotonic	Positive	Expenditure
X_{15}	Total Capital Purchases Excluding Purchases of Land and Buildings	Monotonic	Positive	Expenditure
X_{16}	Total Capital Purchases in Cash	Monotonic	Positive	Expenditure
X_{17}	Total Capital Purchases in Cash, Excluding Cash	Monotonic	Positive	Expenditure
X_{22}	Total Operating Expenses	Monotonic	Positive	Expenditure
X_{23}	Expenditures Paid by Contractor	Free		Expenditure

TABLE 1 Continued

Independent Variable	Type of Predictor	Hypothesized Relationship With Dependent Variable	Type of Variable
X_{24} Expenditures Paid by Operator	Monotonic	Positive	Expenditure
X_{25} Operating Expense Paid in Cash	Monotonic	Positive	Expenditure
X_{28} Economic Classes of Farms	Monotonic	Positive	Income
X_{29} Total Off-farm and Other Income	Monotonic	Positive	Income
X_{30} Total Value of Machinery and Equipment	Monotonic	Positive	Size
X_{31} Types of Farm	Free		Structural
X_{32} Types of Organization	Free		Structural
X_{33} Operators Age	Free		Demographic
X_{34} Size of Farm	Monotonic	Positive	Size
X_{35} Tenure of Operator	Free		Structural
X_{36} Days Operator Works Off the Farm	Monotonic	Positive	Income
X_{37} Race of Operator	Free		Demographic
X_{39} Operator's Net Cash Farm Income	Monotonic	Positive	Income
X_{40} Total Operator's Net Cash Income	Monotonic	Positive	Income

Figure 2. AID Analysis of Total Operator Debt

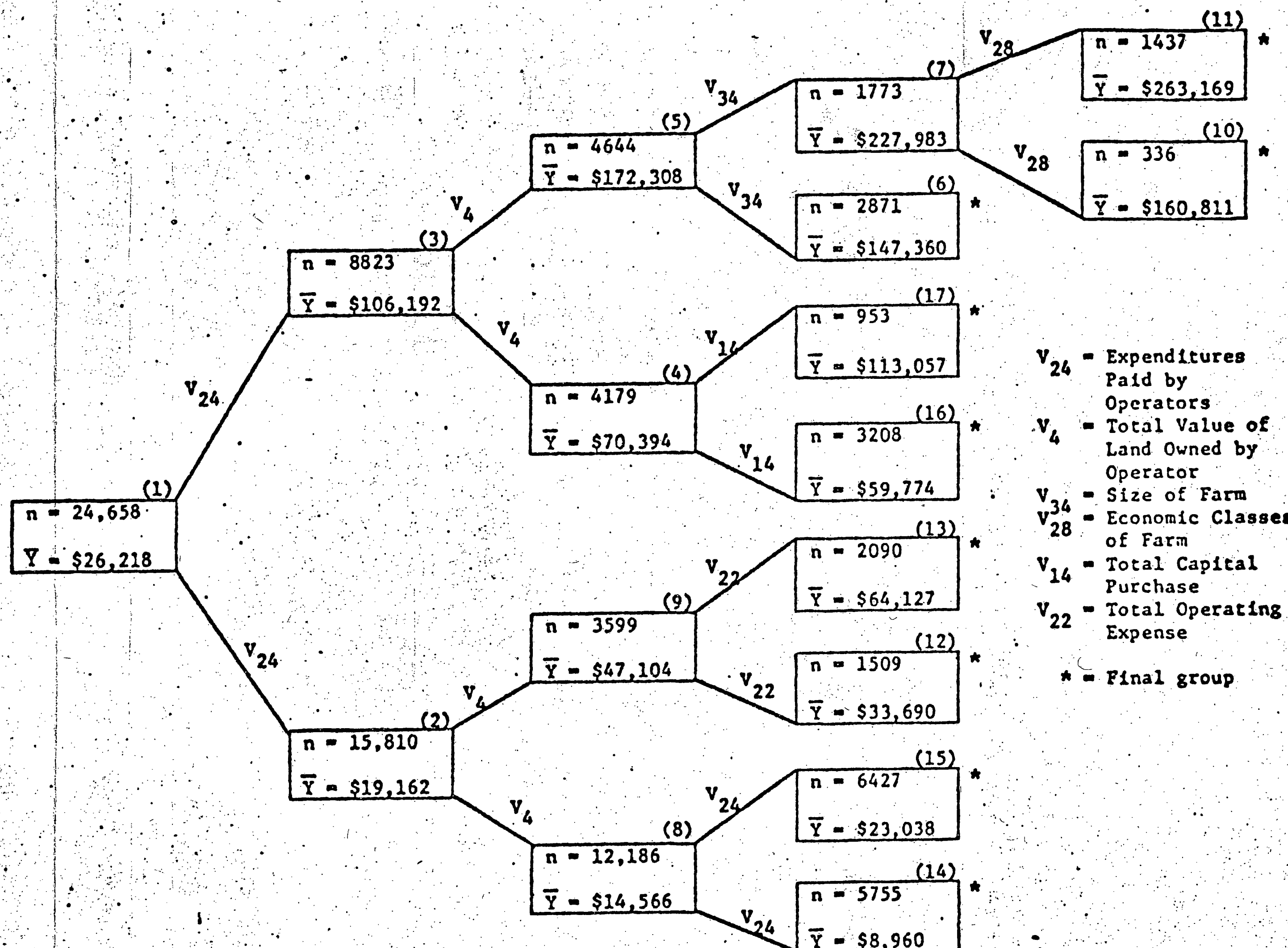


TABLE 2

AID ANALYSIS OF U.S. FARM OPERATORS TOTAL DEBT

Group Split	Group Number	Sample Size	Mean Value	Standard Deviation	Predictor Variables	Variable Values	Percentage of Variance Explained
	1	24,658	$ 26,218	$ 47,019			
1a/	2	15,810	19,162	25,726	Expenditures Paid by Operator (V24)	< $40,000	
	3	8,823	106,192	112,743		$40,000 or more	25.6
3	4	4,179	70,394	63,753	Total Value of Land and Buildings Owned by Operator (V4)	< $200,000	
	5	4,644	172,308	148,136		$200,000 or more	8.7
	6*	2,871	147,360	124,502	Size of Farm (V34)	< 2000 acres	
	7	1,773	227,983	178,468		2000 acres or more	1.8
2b/	8	12,186	14,566	16,484	Total Value of Land and Buildings Owned by Operator (V4)	< $100,000	
	9	3,599	47,104	42,724		$100,000 or more	5.5
7	10*	336	160,811	149,118	Economic Class of Farm (V28)	< $100,000	
	11*	1,437	263,169	182,526		$100,000 or more	.9
9	12*	1,509	33,690	31,838	Total Operating Expenses (V22)	< $16,000	
	13*	2,090	64,127	48,355		$16,000 or more	1.3
8c/	14*	5,755	8,960	9,547	Expenditures Paid by Operator (V24)	< $10,000	
	15*	6,427	23,038	20,549		$10,000 to $39,999	1.7
4d/	16*	3,208	59,774	54,232	Total Capital Purchases (V14)	< $25,000	
	17*	953	113,057	74,063		$25,000 or more	1.1
							46.6%

a/ 25 outliers excluded from Group 1
b/ 25 outliers excluded from Group 2
c/ 4 outliers excluded from Group 8
d/ 18 outliers excluded from Group 4
* Final Group

TABLE 3

AID ANALYSIS OF U.S. FARM OPERATORS REAL ESTATE DEBT

Group Split	Group Number	Sample Size	Mean Value	Standard Deviation	Predictor Variables	Variable Values	Percentage of Variance Explained
	1	17,274	$ 25,477	$ 39,854			
1[a]	2	11,942	19,349	20,374	Total Value of Land and Buildings Owned by Operator	< $200,000	
	3	5,307	92,166	96,141		$200,000 or more	26.0
3	4*	1,072	60,058	59,385	Total Operating Expenses	< $40,000	
	5	4,235	125,315	113,949		$40,000 or more	5.6
5	6*	2,603	106,456	92,801	Size of Farm	< 2000 acres	
	7*	1,632	168,042	142,369		2000 acres or more	2.1
2	8	8,374	14,983	13,497	Total value of Land and Buildings Owned by Operator	< $100,000	
	9	3,568	41,764	31,793		$100,000 - $199,999	5.6
9	10*	905	30,720	24,654	Total Operating Expenses	< $16,000	
	11*	2,663	50,805	34,051		$16,000 or more	.9
8	12*	3,151	8,998	6,920	Total Value of Land and Buildings Owned by Operator	< $40,000	
	13*	5,223	20,711	15,609		$40,000 - $99,999	1.7
							41.9%

a/ 25 outliers excluded from Group 1

* Final group

TABLE 4

AID ANALYSIS OF U.S. FARM OPERATORS NON-REAL ESTATE DEBT

Group Split	Group Number	Sample Size	Mean Value	Standard Deviation	Predictor Variables	Variable Values	Percentage of Variance Explained
	1	18,572	$ 13,339	$ 27,255			
	2	13,349	10,438	16,194	Economic Class of Farm	< $100,000	
1/	3	5,198	80,628	83,982		$100,000 or more	26.4
	4	3,730	67,504	74,041	Size of Farm	< 2000 acres	
3	5	1,468	120,952	98,637		2000 acres or more	2.9
2b/	6	9,758	7,615	10,175	Expenditures Paid by Operator	< $25,000	
	7*	3,566	28,864	27,875		$25,000 or more	7.0
	8*	2, 691	59,393	64,932	Total Capital Purchases Excluding Land and Buildings	< $25,000	
4	9*	1,039	92,012	92,278		$25,000 or more	.8
	10*	238	69,646	78,607	Type of Farm	Cash Grain or Cotton, Other Field Crops,	
5	11*	1,230	134,067	98,986		Vegetable, Fruit & Nut, Poultry, Dairy, Livestock	.9
6c/	12*	4,228	3,974	5,156	Expenditures Paid by Operator	< $6,000	
	13*	5,520	12,196	12,428		$6,000 - $24,999	1.9
							39.9%

a/ 25 outliers excluded from Group 1
b/ 25 outliers excluded from Group 2
c/ 10 outliers excluded from Group 6

* Final group

farm was found to be an important variable in explaining variations in nonreal estate debt.

The stopping rule invoked in all runs was the split reducibility creterion which was set at 0.8 percent. No other splits could be formed that would reduce the variation in the dependent variable by 0.8 percent or more.

Variables Not Selected

Preceding discussion focused on those variables that were selected as important in explaining variation in total debt of individual operators. A priori most, if not all, of the variables would be expected to be important in explaining variations in debt. However, a significant number of variables were not important. It is useful to evaluate the reasons why these variables were not selected as important.

The variables not selected as important were subjectively divided into three categories. First, there are a group of variables which could have explained a relatively high percent of the total variation, but did not enter because they were highly correlated with other variables that were selected. For example, operating expenses paid in cash (V_{25}) is highly correlated with expenditures paid by operator (V_{24}). However, V_{25} did not enter because after the effects of V_{24} were taken into account, it was no longer important. Likewise, total value of land and buildings owned (V_4) is highly correlated with total acres owned by operators (V_3). After the effects of V_4 were taken into account, V_3 was no longer important. If variables V_{24} and V_4 were excluded from the analysis it is likely that V_{25} and V_3 would have replaced them with only a small drop in variation explained.

A second group of variables not selected as important were those which could not explain the minimum level of variation in the dependent variable (.8) even if they were the only variables included in the analysis. For example operator's age (V_{33}) only explained .3 percent of the total variation in the dependent variable. Other variables in this category were race of operator (V_{37}) and expenditures paid by contractor (V_{23}). These variables are relatively unimportant in explaining variations in debt of individual operators in the U.S.

A third group of variables not selected were those which could explain only a relatively small amount of variation in total debt, above the minimum amount of variation (.8) but were not selected as other more important variables were selected. For example, type of business organization (V_{32}) could only explain 3.3 percent of the variation in the dependent variable. Other variables in this category included tenure of operator (V_{35}), days worked off farm (V_{36}), total off farm and other income (V_{29}), and geographic region (V_1).

Summary and Implications

In Census of Agriculture publications and other sources, the amount of individual farm operator debt has been cross-classified with a wide variety of variables. These cross-classifications are intended to assist in revealing how the amount of debt is affected by the variable in question. Unfortunately interpretation of these cross-classifications is difficult and may be misleading because of correlation and interaction among the variables. To overcome problems of this nature more sophisticated techniques are needed. This study examined one possible technique, AID.

Results of the study suggest that size and expenditure variables are of major importance in explaining variations in the amount of debt held by farm operators. After the effects of size and expenditure variables are accounted for, little additional variation in debt level could be explained by the "other" variables included in this study. These "other" variables include demographic, geographic, and structural variables. This suggests that cross-classifications between debt and size or expenditure variables would be more useful in understanding variations in debt than would cross-classifications between debt and geographic regions or between debt and demographic variables.

Results of this study also suggest the potential application of the AID technique for lenders serving a large number of borrowers. Information from loan applications could be analyzed to determine homogenous groups of borrowers. Results of the AID program would then provide a norm (average debt) for each group as well as a measure of deviations from the norm (standard deviations). Such information should prove useful to lenders in evaluating loan applicants.

References

Donaldson, Timothy R. "Factors Affecting Individual Farm Operators Level and Type of Farm Debt" unpublished M.S. Thesis, University of Illinois, Urbana, 1976.

Herr, William, "Understanding Changes in Non-Real Estate Farm Debt," Agricultural Finance Review, Vol. 28, November 1967, pp. 23-32.

Lins, Davis, "Determinants of Net Changes in Farm Real Estate Debt," Agricultural Economics Research, USDA, ERS, Vol. 24, January 1972.

Sonquist, J. A., E. L. Baker and J. N. Morgan, Searching for Structure, Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, University of Michigan, 1973.