Investment Behavior and Energy Conservation

Eddy L. LaDue, Lynn H. Miller, and Joseph H. Kwiatkowski

Binary logit and bivariate probit models were used to investigate the investment behavior of farmers relative to two energy-conserving assets, heat-recovery systems and precoolers. The bivariate probit procedure was useful in correcting for self-selectivity bias. Holdout samples and cross-validation procedures were used to develop true model statistics. Farm size, educational level of the operator, and the type of milking system in use were the important factors influencing investment behavior.

As the economic costs of electrical-energy generation rise and the environmental costs become more evident, energy conservation becomes a clear alternative to building more nuclear or sulphur-emitting plants. The first step to conservation is development of appropriate technologies. However, after such technologies are developed, adoption may be very slow when the technology is embodied in durable capital items such as structures or machines. An understanding of the investment behavior of those who could be expected to use energy-conserving equipment or structures would be of value in predicting or attempting to modify rates of adoption and, thus, energy use.

Attempts to quantify the behavioral relationships connected with investment at the firm (farm) level have generally focused on large field crop and grain storage equipment. Using a probit model, Hill and Kau found farm size, farm type, tenancy, operator age, and specific corn crop variables significant in determining investment in grain dryers in Illinois. Similarly, farm size, tenancy, corn production, and corn use variables were important determinants of grain bin investment in a tobit study by Dixon, Hill, and Saffell. A more recent multivariate analysis of tractor and combine investments found soil type, value of machinery inventory, operator age, and education to influence machinery investment decision making (Johnson, Brown, and O'Grady). In a simulated investment environment, Gustafson, Barry, and Sonka found that structural characteristics of the farm, including tenancy, leverage, and age of the existing machinery, complement to influence machinery investment.

The objectives of this research are to investigate investment behavior relative to two items of energy-conserving equipment that are available to dairy farm businesses. These equipment items differ from previously studied farm investments in that they are generally more modest in cost, involve adoption of relatively new technology, and provide sufficient energy conservation that behavior-modification programs might be of value to electric companies or society.

In the following sections of this paper we (1) provide an overview of the basic research approach, (2) explain the data set used, (3) discuss the development of and results obtained from a series of binary logit models of investment in heat-recovery systems, (4) review the rationale, design, and results of a series of bivariate probit models of investment in precoolers, and (5) provide some conclusions.

Research Approach

The micro-level investment behavior models reported to date generally discuss only a few of the variables identified as influencing investment by the more descriptive literature. This may be due to lack of data, a narrow view of investment behavior, or tight control over the theoretical model design. More realistically, it likely reflects strict adherence to classical analysis procedures whereby one carefully designs a theoretical economic model and then fits the data to the model. If that is the case, one suspects that much "analysis" that would be useful to other researchers is left on the cutting-room floor.

The authors are professor and former research support specialists, respectively, in the Department of Agricultural Economics, Cornell University.

This research was supported by the Niagara Mohawk Power Corporation under a contract with Cornell University and by Cornell University Agricultural Experiment Station Hatch Project 121-413, U.S. Department of Agriculture.

Five researchers finding that something is not important may be just as useful to decision makers as a similar degree of agreement in the affirmative. The basic research approach used in this analysis is to recognize that a rational economic argument can be made that a large number of variables (see Brase and LaDue) may influence investment behavior. The question to be answered is which variables are significant determinants of investment behavior. Thus, for each investment item, a series of similar models (binary logit or bivariate probit), but containing different sets of explanatory variables, are compared. Selection of a "best model" by comparing several models invokes the optimism principle (Picard and Cook), resulting in potentially biased coefficients and goodness-of-fit statistical measures. Practically, the model appears to classify farmers better than the data warrant. To address the problem, either the sample was split into an estimating sample, which was used to test alternate model specifications, and a holdout sample, which was used to determine (validate) the statistical properties of the final model, or where this was impossible, a cross-validation analysis was conducted to determine the true error rate.

The Data

The data used in this study were collected as part of a survey of a random sample of upstate New York farm businesses (Kelleher and Bills). Counties on Long Island and adjacent to New York City were excluded. A personal interview was used to obtain information on production and management practices, energy use, technology adoption, and investment behavior. Reasonably complete information on investment-related variables was obtained on 756 farms. Data were collected during the spring of 1987 and are basically cross-sectional observations.

Farms were defined as places selling $10,000 or more of agricultural products during the 1986 calendar year. Seventy-six percent of the farms surveyed were dairy farms. Representation of other farm types included other livestock, 6%; cash crop, 6%; fruit, 4%; vegetable, 2%; horticulture, 2%; and miscellaneous, 4%. The 756 farms used in this analysis had average total farm assets of $423,000 and average annual gross receipts of $127,000.

Heat-Recovery Systems

A heat-recovery system uses the heat removed from the milk at the bulk tank to preheat water going to the water heater. Heated refrigerant from the bulk tank is used to heat the water, which cools the refrigerant before it is cycled back to the bulk tank. Since dairy farms must cool all milk from body temperature to 32°-4Q°F and use large amounts of hot water in the milking and cleaning processes, substantially lower energy expenditures can be experienced when using a heat-recovery system.

The Model

A binary logit model is used where the probability of investing in a heat-recovery system *(Y)* is estimated as

$$(1) \qquad \hat{Y}_i = 1 - \left[\frac{1}{1 + \exp(a + b'X_i)}\right],$$

where r/ is the predicted probability that farmer / will invest in a heat-recovery system given the values for X,; *a* is a constant; exp is the base of the natural logarithm; *b'* denotes the vector of regression parameters; and X, denotes the values of the factors related to investment in a heat-recovery system for farmer i. To estimate this model, it is reformulated as

$$(2) \qquad Y_i^* = a + (b'X_i),$$

where Y_f is the log of the odds ratio of investing in a heat-recovery system (i.e., Y^* =

$$\ln\left(\frac{Y_i}{1 - Y_i}\right)).$$

The model was estimated using the supplemental LOGIST procedure from the Statistical Analysis Systems Institute (Harrell). The dependent variable was one for farms with a heat-recovery system and zero for those without such a system. Only data from dairy farms were used. A holdout sample was selected randomly from the total sample and set aside for development of model statistics for the final model selected. The estimating sample included 261 farms; the holdout sample included 267 farms. Sample sizes differ because of the random process used for assigning farms.

Since 38.6% of the farms in the sample had invested in a heat-recovery system, this sample prior probability of investment was used as the cutoff point for classification of farmers. Farms for which the probability of investment in a heat-recovery system exceeded the cutoff point were classified as investors. Those with lower probabilities were classified as not investing in a heat-recovery system. This is consistent with Beaver, who states that the optimal cutoff point is one that will minimize the percentage of incorrect classifications, and Maddala, who points out that, following Bayes theorem, if the costs of misclassifying are equal for both types of error, then the optimal cutoff point will be the prior population probability of being in a class. For this study, the costs of committing type I or type II error are equal. Although the population probability is not known, it is established from the sample probabilities.

The variables included in the models were those identified in the literature as influencing agricultural investment. The initial model contained eight economic variables. Average number of cows on the farm in 1986 was used as an indicator of size. Economies of scale imply that a minimum number of cows are needed to justify investment in a heat-recovery system, resulting in a higher probability of investment with larger size. However, as herd size increases beyond some point, it appears likely that the added economic incentive would increase the probability of investment in a heat-recovery system, only at a decreasing rate. Thus, cow numbers squared is added to allow for the expected curvilinear relationship.

The four geographic regions of New York State represent differences in soil and climate resources, input costs, and milk prices. Region 1 has a variable resource base, somewhat above-average input costs, above-average milk prices, and considerable urban pressure. Region 2 is largely hill and valley soils. Region 3 has the best soil and climate resources. Region 4 has modest soil resources, colder temperatures, lower input costs, lower milk prices, and few alternatives to dairy farming. Although those factors could have an influence on heat-recovery adoption, the expected sign of the included variables is indeterminate. Dummy variables were included for regions 1,2, and 4.

The existence of a parlor or pipeline milking system indicates a willingness to adopt milking technology. Also, parlors and pipelines are often more recent investments than a bucket or bucket/transfer system, implying a higher probability of mi lking-system investment since heat-recovery systems have become available. Both of these factors would encourage heat-recovery investment. A dummy structure was used to represent milking systems. Parlor and pipeline variables were included; the bucket or bucket/transfer system was excluded.

Education has been found to be positively correlated with investment in new technology (Funk). Education facilitates the evaluation process as well as management of the new asset. Education is measured as years of formal education.

The management index was constructed from respondents' answers to questions about use of farm records, input buying strategies, marketing procedures, personnel-management practices, and short-term goal-setting behavior (LaDue and Kwiatkowski). The index has four levels, with the best combination of management practices coded four. Given the profitability of a heat-recovery system for most farm situations, better managers are expected to be more likely to adopt.

Farmers with higher income expectations would be expected to have more funds available for investment and greater optimism for the future, thus resulting in greater investment. In addition to data on their 1986 cash income, farmers were asked to indicate how they expected income to change by 1990. These data were used to estimate 1990 expectations as (1) lower, (2) the same as, or (3) higher than actual 1986 incomes. Zero-one variables were included for lower (lesser) 1990 expectations and higher (greater) expected 1990 income.

Interest rate has long been considered an important factor influencing investment. Presumably lower interest rates make more investments financially feasible and, thus, result in greater investment. The rate used in the analysis was the predominant rate paid on the investment for those who invested. For noninvestors, the rate used was the average rate paid by investing farmers who had the same primary credit source.

Age has frequently been found to be related to investment. Younger farmers invest as they are trying to increase their level of income. Older farmers who have reached a reasonable income and enterprise size tend to reduce investment and, thus, disinvest as they near or reach retirement age.

The Results

The initial results are labeled Model 1 in Table 1. The variables that were not significant at the .05 level are age, interest rate, and management ability. One of the three regional dummies was also insignificant, although the other two were significant. The sign on the dummy variable for those expecting future (1990) income to be greater than 1986 income was unexpected.

The results achieved with age could be explained in two ways. First, age may represent a number of correlated variables such as education and size, and when they are included in the model, age becomes unimportant. Alternately, the low dollar investment required for a heat-recovery system and the high profitability of the investment may make it a high priority even for the young with few resources who are limited to modest total investment and the old who are reducing total business size.

The apparent unimportance of the interest rate

Table 1. Comparison of Heat Recovery Models Estimating Sample

Variables	Model 1	Model 5	Model 6	Model 7
Model Coefficients and *P* Values[a]				
Intercept	-7.221	-7.757	-6.549	-1.938
	(.00)	(.00)	(.00)	(.00)
Cows	0.01	0.011	0.008	0.016
	(-02)	(.01)	(.04)	(-00)
$Cows^2$	-0.00001	-0.00001	-0.00001	-0.00001
	(-08)	(.07)	(.14)	(-00)
Pipeline	2.365	2.329	2.442	2.552
	(.00)	(.00)	(.00)	(.00)
Parlor	3.210	3.010	2.968	3.115
	(.00)	(-00)	(-00)	(.00)
Education	0.318	0.260	0.226	
	(.00)	(.00)	(.00)	
Region 1	0.491	0.266		
	(-32)	(.57)		
Region 2	1.351	1.275		
	(.00)	(.00)		
Region 4	1.100	0.811		
	(-03)	(.09)		
Greater 1990 Inc.	-0.879			
	(.02)			
Lesser 1990 Inc.	-1.414			
	(.02)			
Management	0.142			
	(.43)			
Interest rate	-0.117			
	(.20)			
Age	0.001			
	(.93)			
Model Statistics				
Chi square	90.6	83.7	70.5	64.1
P value	0.00	0.00	0.00	0.00
R	.45	.45	.42	.41
C statistic	0.84	0.83	0.80	0.79
Correct Classification Percentages				
Total	77.5	73.7	72.3	70.9
With heat recovery	79.1	73.7	72.6	66.3
Without heat recovery	76.6	73.8	72.0	73.5
Class efficiency	77.6	73.8	72.2	70.7

[a] *P* values are in parentheses under the coefficient. *P* value indicates the probability that the coefficient is zero.

may also stem from the particular characteristics of this investment. The modest investment required likely implies that many fanners could make the investment from equity rather than borrowed capital. Also, the system may sufficiently add to profitability so that it is a good investment for a wide range of interest rates. An alternate explanation would be measurement error in the interest-rate variable. The actual rate that a fanner might pay on a heat-recovery system investment may differ from the general investment rate used in this analysis.

The reason for the insignificance of the management variable is less clear. Possibly, the investment is simple enough that profitability is easily identified from herd size even for those with limited management skills. Alternately, education may sufficiently represent managerial capacity for this investment.

The inconsistent sign achieved with the income-expectations variables led to an investigation of other income variables. Four alternates were considered: (1) a composite net farm income variable with weights of 25% for 1980, 50% for 1985, and 25% for 1986; (2) 1986 income; (3) 1990 expected income measured in dollars; and (4) 1980 income. None of these alternatives improved model results. Apparently, the profitability of this particular investment can be identified with sufficient clarity that farmers are little influenced by general expectations about future profitability or income levels.

All coefficients in Model 5 have an acceptable sign and are either significant at the .10 level or tied to variables that are significant. For the total model, the adjusted pseudo R is .45, which is good for cross-sectional farm-level data. The C statistic of 0.83 is also acceptable for this type of study.[1]

The model correctly classifies over 73% of all farmers. The correct classification rate for those with heat recovery is similar to the rate for those without heat recovery. This compares quite favorably to the conditional probability rate of 52.6%.[2] The conditional probability rate is the rate of correct classifications expected assuming one only knew the proportion of the population that had heat-recovery systems.

To examine the possibility that some other variables frequently identified as influencing investment behavior may be important in heat-recovery investment, variations of Model 5 were run incorporating these variables one at a time. The variables investigated were risk tolerance, form of business or organization, primary goals, and urban proximity. None of these variables were found to be important to investment in heat-recovery systems.

Given the high P values for two of the three regional dummy variables, the importance of region in the equation could be questioned. In Model 6, region is dropped. This resulted in a lower coefficient for herd size and a higher intercept. Overall, model statistics deteriorated somewhat, although the reduction in classification ability was modest.

The only variable in Model 6 that is farmer, rather than farm, related is education. Thus, a parsimonious predictive model developed by dropping education from Model 6 results in Model 7. This model could be used to predict heat-recovery investment without knowledge of farm-operator characteristics and with only two pieces of information on the farm itself. The loss in significance of model statistics and correct classification percentages from dropping the education variable is modest.

Based on the assessment that Model 6 is the "best" economic model, in that it has acceptable model statistics with relatively few problems, and that Model 7 may be useful as a predictive model for situations without information on the farmer, these two models were refit using the holdout sample (Table 2). The holdout sample statistics represent the true statistical characteristics of the model. As expected, the model statistics and correct classification percentages were somewhat lower for the holdout sample than were observed with the estimating sample. However, the statistics are still quite acceptable for cross-sectional farm data.

The only surprises were the lower value obtained for the education coefficient and that the education variable becomes insignificant. This appears to imply that either the two samples are significantly different in education characteristics or that the education variable is somewhat unstable in its effect.

Clearly the most important determinant of investment in a heat-recovery system is the type of milking technology employed. The more capital intensive the technology, the more likely the investment in a heat-recovery system.

Projected Probabilities

Using the probability form of the equation, the estimated equation can be used to calculate the probability of investment in a heat-recovery system for farms with different characteristics. To obtain the model coefficient that should have the highest likelihood of being accurate, the final forms of the equations were fit using the entire sample (estimating plus holdout samples). These coefficients were used to calculate the probabilities of investment.[3] When education is held at the average 12.55 years, the probability of a bucket or transfer milking system farm investing in a heat-recovery system is less than 10%. The likelihood of investment by pipeline owners ranges from about 35% for those with small herds to 70% for those with large herds. About half of the parlor owners with small herds invest in heat recovery compared to 85% for those with large herds.

The level of education significantly increased the probability of investment. With herd sizes for each milking system held at their means—62.7 cows for

[1] With a binary model, the C statistic is equivalent to the area under the receiver operating characteristic curve (Hanley and McNeil). Thus, the statistic has a range of 0.5 to 1.0, with 0.5 indicating no apparent discriminatory power and 1.0 indicating perfect discriminatory power. In this case, the C statistic represents the probability that a randomly chosen farm with a heat-recovery system will be correctly rated by being given a higher probability than a randomly drawn farm without a heat-recovery system.

[2] The naive conditional probability of a farm is calculated using the prior probability of the farm being in the two different groups. For this model, the prior probability of having a heat-recovery system is 38.6%. Thus, the conditional prior probability of correctly classifying a farm given this knowledge is (.386)(.386) + (.614)(.614) = 52.6%. To calculate the model's efficiency, the model's correct classification rates for each group are substituted. Thus, the efficiency is (.386)(.663) + (.614)(.735) = .707.

[1] Only the coefficients generated by this process are used. The statistics are presented for information only.

Table 2. Validation and Heat Recovery Model Values

	Holdout Sample		All Observations	
Variables	Model 6	Model?	Model 6	Model 7
Model Coefficients and *P* Values[3]				
Intercept	-3.972	-3.338	-5.099	-3.459
	(.00)	(-00)	(.00)	(.00)
Cows	0.011	0.012	0.009	0.009
	(.01)	(.01)	(.00)	(.00)
Cows^2	0.00001	0.00001	0.00001	0.00001
	(.03)	(.03)	(-01)	(.02)
Pipeline	2.196	2.224	2.333	2.360
	(.00)	(.00)	(.00)	(.00)
Parlor	2.496	3.556	2.760	2.838
Education	(.00) 0.057	(.00)	(-00) 0.134	(.00)
	(.49)		(.01)	
Model Statistics				
Chi square	63.2	66.3	128.9	123.0
P value	0.00	0.00	0.00	0.00
R	.39	.40	.41	.41
C statistic	0.77	0.77	0.77	0.78
Correct Classification Percentages				
Total	66.9	68.5	67.8	69.3
With heat recovery	84.8	84.8	72.9	71.0
Without heat recovery	53.6	56.8	64.4	68.2
Class efficiency	65.6	67.6	67.7	69.3

* *P* values are in parentheses under the coefficient. *P* value indicates the probability that the coefficient is zero.

pipeline systems and 127.3 cows for parlor systems—only 30% of the pipeline owners with a tenth-grade education could be expected to have a heat-recovery system compared to nearly 50% for those with a college education. Similarly, education increased the probability of investment by parlor owners from about 50% to over 70%.

The combined effect of herd size and education level explains a large part of the variation in heat-recovery-system investment. Milking-parlor owners with little education and a small herd only had about a 35% chance of investing in a heat-recovery system while those with college degrees and large herds had a 90% chance of owning a heat-recovery system.

Precoolers

A precooler uses cold well water to cool milk while it is being piped from the milking operation to the bulk tank. The milk passes through small tubes or channels that are surrounded by a counterflow of cold water. This process reduces the energy costs of cooling milk by reducing the temperature of the milk before it gets to the bulk-tank cooling system. The warmed water resulting from this process is frequently used for washing or animal consumption.

The Model

As currently designed, this technology is possible only on farms with a milking parlor or pipeline. By selecting a milking system, farmers may simultaneously eliminate the possibility of precooler ownership. Further, practically no one familiar with dairy farm technology would suggest that parlor and pipeline ownership are randomly distributed among farms. Since we can only observe precooler ownership with farmers who have the appropriate milking system, we have self-selectivity bias (Maddala).

To adjust for this bias, the probability of having a parlor or pipeline must be incorporated in model design. This is superior to estimating a model using only the farms with a parlor or pipeline since data inherent in the self-selection process would be omitted (Heckman). A bivariate probit model is used to simultaneously estimate two equations, one for ownership of a parlor or pipeline system and the second for investment in a precooler.

The two equations are specified as[4]

[4] See VanDeVen and VanPraag for a similar application.

$$(3) \qquad Y^*_{i1} = b_1'X_{i1} + e_1,$$

$$(4) \qquad Y^*_{i2} = b_2'X_{i2} + e_2,$$

where for the /th observation (farmer) Y\ is the estimated probit for investing in a precooler; Y_2 is the estimated probit for operating a milking-parlor or around-the-barn pipeline milking system; b\' denotes the vector of regression parameters for the investment in a precooler; b_2' denotes the vector of regression parameters for the operation of a milking-parlor or around-the-barn pipeline milking system; X! denotes the vector of factors related to the investment in a precooler; and X_2 denotes the vector of factors related to the operation of a milking-parlor or around-the-barn pipeline milking system.

Since precoolers can only occur on farms with parlors or pipelines, ($X_{(1}$, X,i) is observed only when Y_a = 1. It is assumed that: e_j ~ WU) with corr *(ei, e_2) = q*. An individual farmer is expected to invest in a precooler (parlor or pipeline) if $Y_n > 0$ ($y^*_{,2} > 0$); that is, Y_n = KYn = 1). An individual farmer is expected not to invest if r*.-i < 0 (r^*_a < 0); that is, Y_n = 0 (Y_{i2} - 0).

The model was estimated using the bivariate probit option of LIMDEP (Greene). The dependent variable for the milking-system equation was one for farms with a parlor or pipeline and zero for farms without such investment, and for the precooler equation was one for farms with a precooler and zero for those without. Attempts to fit this model to the estimating sample remaining after setting aside a holdout sample of approximately 50% were unsuccessful. Nearly the entire sample was required for the program to obtain a solution. Thus, the model was estimated using the entire data set. The data requirements for estimation also limited the number of variables to be included in the base model.

The variables included in the milking-system equation to determine the probability of a farm having a pipeline or milking parlor are conceptually similar to those important in heat-recovery or precooler investment. Farm size, as measured by average number of cows in 1986, is particularly important to milking-parlor investment. Because of the large investment required for a parlor, it is a profitable investment only for large farms. Pipeline systems are usually installed to increase the number of cows that can be handled per milking. The average number of cows per farm participating in the Cornell Dairy Farm Business Summary in 1987 (Smith, Knoblauch, and Putnam) was 48 for bucket or transfer systems, 71 for pipeline systems, and 157 for milking-parlor systems.

Age of the operator has been shown to have a negative influence on the probability of large machinery investment (Hill and Kau). Age of the operator is, thus, expected to have a negative effect on the selection of a parlor or pipeline milking system.

Those with higher levels of education are expected to be better able to evaluate and manage the complexities of such a major technologically laden investment. Therefore, higher-educated farmers are expected to more readily adopt parlor or pipeline milking systems. Regional dummies are included in the equation to represent the different soil, climatic, and cost differences of the different regions of the state.

A composite cash income is included in this equation as a combination of income expectations and a cash-flow variable. The variable is constructed as a sum of 25% of 1980, 50% of 1985, and 25% of 1986 net cash farm income. Since the milking systems were constructed prior to 1986, income over the 1980-85 period is likely more indicative of expectations at the time the investment was made. Those with more cash income for making payments on loans, or for making direct investments, and those with higher income expectations are expected to be more likely to invest in a modern milking system.

The precooler equation is specified exactly as the heat-recovery-system model except that the milking-system variables are excluded. Because the precooler performs a similar function to that of the heat-recovery system, the logic for inclusion of the individual variables is similar.

The Results

The results from estimating the hypothesized relationships are labeled Model 1 in Table 3. The overall model is significant in that the hypothesis that all the model coefficients are zero is rejected at the .05 level. The high level of correlation (.81) indicates that the (residuals of the) two equations are related and that fitting an equation for precooler investment using only those farmers who had a parlor or pipeline milking system would provide different results. Some factors that would appear to be influencing precooler investment would in truth be related to milking-system adoption rather than precooler adoption per se.

As observed with the heat-recovery model, age appears to have little relationship to this type of investment when other variables are included. Although management ability appears to be important in the milking-system equation, it contributes little to the decision to invest in a precooler. As was observed for the heat-recovery system, it may

Table 3. Comparison of Bivariate Probit

Intercept	-2.940	-2.758
	(.00)'	(.00)
Cows	0.807	0.908
	(.00)	(.00)
Cows^2	-0.00070	-0.00083
	(.02)	(.00)
Education	0.109	0.102
	(.02)	(-01)
Cash income	-0.027	
	(.86)	
Region 1	-0.440	
	(.06)	
Region 2	-0.093	
	(-65)	
Region 4	-0.151	
	(.50)	
Management	- 0.008	
	(.92)	
Age	0.007	
	(-34)	
Milking-System Equation		
Intercept	-1.758	-1.769
	(.02)	(.00)
Cows	2.228	2.263
	(.00)	(.00)
Management	0.258	0.267
	(.00)	(.00)
Region 1	0.741	0.660
	(.00)	(.00)
Region 2	0.525	0.515
	(-01)	(.01)
Region 4	0.468	0.452
	(.04)	(.04)
Cash income	0.515	0.515
	(.14)	(.14)
Education	0.01 1	
	(.81)	
Age	0.741	
	(.70)	
Model Statistics		
Log likelihood	-341.1	-343.8
Correlation	-.811	-.752
Significance	0.00	0.00

[a] *P* values are in parentheses under the coefficient. *P* value indicates the probability that the coefficient is zero.

be that the precooler is a small enough investment or that the economics of investment are clear enough that management ability is not important in this investment decision. Alternately, education may sufficiently represent management ability for this decision. Investment is apparently related to managerial capacities that are better represented by education than the particular set of management practices contained in the management index.

in the milking-system equation, however, education is consistently insignificant, but management is significant. Farmers with better management ability appear more likely to believe they can correctly evaluate and manage such a system for their farm and to have the funds or credit capacity required to make the large investment. Education adds little to the assessment of managerial capacities indicated by the managerial index.

Geographic region appears to be important in the decision to invest in a milking system but not in the precooler decision. Regional cost, soil, and climate differences are apparently important in the decision to invest in a milking parlor, but once that decision (a parlor or pipeline system) is made, these factors are unimportant to the precooler decision. Since the precooler uses quantities of water, regional differences in water availability could influence investment. However, either this is not the case or areas with limited water are not synonymous with the geographic regions selected.

Cash income also appears to be important in the milking-system decision but not the precooler decision. Given the relatively small dollar outlay required to install a precooler, this result .is quite logical.

The final model, Model 6, includes all of the variables of the initial model except age in one or both of the equations. All coefficients are significant at the .05 level except cash income. To test other variables that the literature has identified as important to investment behavior, the following were added to Model 6 one at a time: (1) form of business ownership, in the form of dummies for sole proprietorship, partnership, or corporations; (2) distance to a city with a population of 20,000 persons or more; (3) degree of risk tolerance in the form of dummies for more risk averse, more risk tolerance, or average risk tolerance; and (4) goals for the next ten years specified as sell the farm, pass farm to next generation, expand size or income, or improve family living. None of these variables had significant coefficients or improved the significance of the overall model.

The correct classification percentages for Model 6 are shown in the top half of Table 4. Since 22.4% of the sample farms had precoolers, that sample prior probability was used on the cutoff point for classification of farmers. Classification percentages are not calculated for the other models because the classification proportions are expected to move with the statistical properties of the model. In spite of the rather favorable statistical properties of the model, its classification ability is about the same as the conditional probability rate.

Consistent with the findings of the heat-recovery

Table 4. Bivariate Precooler Model 6 Correct Classification Percentages

Total sample	65.8
With precooler	85.4
Without precooler	60.2
Conditional model rate	65.2"
Model classification efficiency	65.8

[a] Based on sample probability of precooler investment of 22.4%, (22.4 x 22.4 + 77.6 x 77.6).

investment analysis, the most important factors influencing investment in a precooler are herd size and education. However, investment in the milking system required for use of a precooler is also influenced by management ability, geographic region, and cash income.

A tenfold cross-validation analysis (Efron; Frydman, Altman, and Kao) was used to estimate the true error rate (correct classification rate) for the final model (Model 6). The SAS RANUNI random number generator was used to randomly assign each observation to one of the ten groups. Probit values were calculated from the precooler equation generated from each estimating sample. Probit values were converted to probabilities using the PHI cumulative standard normal distribution function from LIMDEP (Greene). A cutoff point equal to the prior probability of precooler investment (22.4) was used in developing predictions. Calculations included only those farms with a parlor or pipeline system; bucket-system farms were omitted (for more detail see LaDue, Miller, and Kwiatkowski).

The results of the cross-validation analysis (Table 5) indicate that the true error rate of the model is approximately equal to or slightly greater than the conditional model rate. Although the model does not improve our probability of correctly classifying investment, it does provide some indication of the factors influencing investment. A predictive model with a lower error rate (higher correct classification rate) could be developed by modifying the cutoff probability used for classification. For

Table 5. Bivariate Precooler Model 6 Summary of Cross-validation Results

Sample or Measure	Average Correct Classification Percentages: Estimating Samples	Holdout Samples
Total sample	64.5	61.3
Farms with precooler	86.5	76.5
Farms without precooler	58.2	56.8
Model classification efficiency	63.6	61.0

example, use of a .5 probability cutoff as is used by the LIMDEP model (Greene) significantly increases the correct percentages for those without a precooler (and reduces it for those with a precooler) and, thus, raises the correct classification rate.

Projected Probabilities

For this model, the probability that a farmer will invest in a precooler is indicated by

$$\hat{Y}_i = 1 - F(-bi'X_{i1}), \tag{5}$$

where *Yj* is the probability that farmer *i* will invest in a precooler; and *F* is the standard normal cumulative distribution function, #(0,1).

Calculating the investment probabilities using Model 6 coefficients (Table 3) indicates that only about 15% of the farms with small herd sizes invested in precoolers. However, farms with 400 to 500 cows had an 80% to 90% probability of precooler investment. Similarly, only about 12% of the farmers with only a tenth grade or less education owned a precooler. The probability of investment increased as education increased so that those with some college beyond the bachelor's level had about a 35% likelihood of investment.

Conclusions

The bivariate probit model can provide a useful tool for evaluating systems where self-selectivity bias exists. However, estimating the two binary dependent variable equations requires a relatively large number of observations, particularly when binary or dummy independent variables are included.

Investment in heat-recovery systems and precoolers is primarily determined by a small number of variables. The comparison of binary logit models indicated that only farm size, the existence of a parlor or pipeline system, and education were important in determining heat-recovery-system investment. Similarly, although the comparison of bivariate probit models found farm size, management ability, geographic region, and level of cash income important in milking-system investment, only herd size and education were important in precooler investment.

The particular character of the variables that influence investment in these energy-saving devices should facilitate efforts to improve adoption rates and, thus, improve energy conservation. They are generally not disclosure-sensitive in nature and farmers would have little difficulty providing their values. The most fertile ground for increasing adoption rates would be farmers with large herds and pipeline or parlor milking systems. Initial efforts likely should focus on the higher educated, but educational programs designed to improve farmers' understanding of the technology and their ability to handle it should also be initiated.

References

Beaver, W. H. "Financial Ratios as Predictors of Failure." *Journal of Accounting Research, Supplement* (1966):71-127.

Brase, B. T., andE. L. LaDue. "Fanner Investment Behavior: A Review of Literature." Department of Agricultural Economics Research Bulletin 89-5. Cornell University, March 1988.

Dixon, B. L., L. D. Hill, and M.S. Saffell. "The Relationship of Farmers' Intentions to Purchase and Their Actual Purchases of Metal Bin Grain Storages." *North Central Journal of Agricultural Economics 2,* no. 1 (1980):61~68.

Efron, B. "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association* 78, no. 382 (June 1983).

Frydman, H., E. I. Altman, and D. L. Kao. "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress." *The Journal of Finance* 40, no. 1 (March 1985):269-90.

Funk, T. F. "Farmer Buying Behavior: An Integrated Review of Literature." Working Paper AE-72-16. University of Guelph, 1972.

Greene, W. H. *UMDEP, Version 5.* New York: William H. Greene, 1988.

Gustafson, C. R., P. J. Barry, and S. T. Sonka. "Machinery Investment Decisions by Cash Grain Farmers in Illinois." Paper presented to NC-161 Regional Research Project, Evaluating Financial Markets for Agriculture, St. Paul, MN, 7-8 October 1986.

Hanley, J. A., and B. J. McNeil. "The Meaning and Use of the Area under the Receiver Operating Characteristics Curve." *Diagnostic Radiology* 143, no. 1 (I982):29-36.

Harrell, F. E., Jr. "The Legist Procedure." In *SUGI Supplemental Library User's Guide, Version* 5 *Edition,* chapter 23. Gary, NC: SAS Institute, Inc., 1986.

Heckman, J. J. "Sample Selection Bias as a Specification Error." *Econometrica* 47, no. 1 (January 1979):153-61.

Hill, L. D., and P. Kau. "Application of Multivariate Probit to a Threshold Model of Grain Dryer Purchasing Decisions." *American Journal of Agricultural Economics* 65, no. 3 (1973):553-57.

Johnson, T. G., W. J. Brown, and K. O'Grady. "A Multivariate Analysis of Factors Influencing Farm Machinery Purchase Decisions." *Western Journal of Agricultural Economics* 10, no. 2 (1985):294-306.

Kelleher, M. J., and N. L. Bills. "Statistical Summary of the 1987 Farm Management and Energy Survey." Department of Agricultural Economics Research Bulletin 89-3. Cornell University, February 1989.

LaDue, E. L., and J. H. Kwiatkowski. "An Analysis of the Investment Related Characteristics of New York Farmers." Department of Agricultural Economics Research Bulletin 89-18. Cornell University, September 1989.

LaDue, E. L., L. H. Miller, and J. H. Kwiatkowski. "An Analysis of Alternate Micro Level Models of Investment Behavior." Department of Agricultural Economics Working Paper 89-11. Cornell University, December 1989.

Maddala, G. S. "Limited-Dependent and Qualitative Variables in Econometrics." Econometric Society Monographs, no. 3. Cambridge: Cambridge University Press, 1983.

Picard, R. R., and R. D. Cook. "Cross-Validation of Regression Models." *Journal of the American Statistical Association* 79, no. 387 (1984):575-83.

Smith, S. F., W. A. Knoblauch, and L. D. Putnam. "Dairy Farm Business Summary, New York." Department of Agricultural Economics Research Bulletin 88-8. Cornell University, July 1988 and prior issues.

VanDeVen, W. P. P. M., and M. S. B. VanPraag. "The Demand for Deductibles in Private Health Insurance, A Probit Model with Sample Selection." *Journal of Econometrics* 17(1981):229-52.