Impact Analysis of Auburn Related Services on Farmers in Alabama

by

Le Wu

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 4, 2012

Keywords: Impact Analysis; Probit Model; Binary Logit Models; Ordinal Logit Models.

Approved by

Denis Nadolnyak, Co-Chair, Assistant Professor of Agricultural Economics and Rural Sociology
Yaoqi Zhang, Co-Chair, Associate Professor of Forestry and Wildlife Sciences
Valentina Hartarska, Associate Professor of Agricultural Economics and Rural Sociology

Abstract

This thesis studies what factors influence farmers' decisions to use services provided by Auburn University and how these programs affected the performance of farmers in Alabama. The study is based on survey data collected in Fall 2010. I apply classical statistical tools such as two samples test statistics, discrete choice models and generalized linear models to analyze the relation between AU related services and the performance of farmers in Alabama. Finally, I determine how total gross value of sales, value of total assets, education, age, return of investments, years of farming experience and years of the current farming operation affect choices of farmers to use specific AU related services. The results show that farmers benefit from some services such as AU research stations, County or Regional extension agents, AU professors and AU website. Farmers in User-group are more likely to perform better than those in Nonuser-group. Farmers using these Auburn University services had more total gross value of sales than those in Nonuser groups. Using such services as AU research stations, County or Regional extension agents, AU website and Climate Forecasts at agroclimate.org, farmers had more chance to earn an increase in return of investments than those in Nonuser groups.

Acknowledgments

I would like to sincerely thank two co-chairs, Dr. Denis Nadolnyak and Dr.Yaoqi Zhang. Without their support, excellent guidance for doing research, patience, and caring, I would never complete the thesis. I would also like to thank the Masters Program Coordinator and my committee member, Dr. Valentina Hartarska for guiding my research. Finally, I would also like to thank my parents and my husband, Aijun Zhang for their continued support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction, Motivation and Literature Overview

This current thesis studies the impact of agricultural experimental station and services provided by Auburn University(AU) on farmers in Alabama. The purpose of the thesis is to investigate what factors influence farmers' decisions to use services provided by Auburn University and how these programs affected the performance of farmers in Alabama.

Agricultural extension services play critical roles in assisting farmers such as information on latest farming technology and management knowledge. The origins of agricultural extension services can be traced to thousands years ago. Records of agricultural practices have been found from ancient Egypt and China back to 3,000 years. It is not known where or when the first extension activities started. The extension services have been evolved thousands years, although modern forms occurred largely during the last two centuries. Since the middle of the 19th century, extension service attracted the attention of government officials in Europe, for example, Ireland, Germany, Denmark, Netherlands, Italy, and France and started to spread there. Even in the nearest 21 century, agricultural extension services play important roles, in particular for developing countries.

Some extension service programs are supported by non-governmental organizations or the national governments, especially in those developing countries of Asia and Africa. Most developing countries established formal agricultural extension programs. There are many examples about the research on impact of agricultural extension on farmers. The following is not a complete list. In [30], Thirtle, Lin, and Piesse (2003) studied how the impact of research led agriculture productivity growth on poverty reduction in Africa, Asia and Latin America. During the 1970s and 1980s, the training and visit (TV)system was established by the World Bank. According to Birkhaeuser, Evenson and Feder(1991), at that time more than 40 countries had adopted this approach. TV was originally tested in Turkey in the late 1960s (Swanson, Bentz and Sofranko 1998). Anderson and his collaborators ([3], [4], [5]) studied some agricultural advisory services and TV extension programs in Asia and Aferica. In [12], Feder and Just, et al. (1985) and in [13], Feder, Gershon, and Roger Slade (1986), studied the (TV)system in India. The case study in India showed that the incremental investment in TV extension was likely to generate at least a 15 to 20 percent rate of return. In [11], Evenson, and Germano(1998) studied the effects of agricultural extension on farm yields in Kenya. In [32], Wu, Praneetvatakul, Waibel and Wang (2005) studied the impact of farmer field schools on yield, pesticide cost and gross margin in Shandong Province, P.R. China. In Argentina([26]), Pedro, Maffioli and Ubfal(2008) studied the impact of agricultural extension services for the particular case of grape production. In Nepal [27], Ransom, Paudyal, and Adhikari (2003) studied adoption of improved maize varieties in the Hills of Nepal and Jamison and Moock(1984) [18] studied how farmer education affect farm efficiency. In the Peruvian Andes [14], the authors studied the impact of farmer field schools on knowledge and productivity. In [25], Owens, Hoddinott and Kinsey (2003) studied the impact of agricultural extension on farm production in resettlement areas of Zimbabwe.

Universities play important roles in improving the productivity of farmers by providing technology and management knowledge to farmers. Extension services attracted by universities starting from universities of Oxford and Cambridge around 1850. The success of such efforts in Britain influenced universities in United States. During the first two decades of last century, in United States, a lot land-grant colleges concerned with serving the needs of farmers. The Hatch Act of 1887 gave federal land grants to states in order to create a series of agricultural experiment stations, which were usually connected with land-grant state colleges and universities.

Auburn University, one of Alabama's three land-grant universities, is the headquarter of Alabama Agricultural Experiment Station (AAES). For more than 125 years, Auburn University provided the services and support for the Alabama farmers. Auburn University sponsored some programs for farmers in the area of Alabama. Farmers may seek help from these programs by ways of AU research stations, county or regional extension agents, AU professors, information on a website provided by AU or other related to AU.

To study the impact of Auburn Services, I adopt impact analysis method. There is a large of literature which is dealing with the impact of agricultural extension services on farmers. Impact analysis has been widely used in agricultural economics and many other areas. In software engineering, impact analysis is defined by Bohner and Arnold(1996)[7] as "the determination of potential effects to a subject system resulting from a change". In Wikipedia, economic impact analysis(EIA) is defined as "EIA examines the effect of a policy, program, project, activity or event on the economy of a given area". The area can range from a local neighborhood to the global world. In the thesis, I consider only a local area, Alabama in the southern United States.

A range of approaches to impact analysis have been promoted over the years. Performance evaluation is an important process in impact analysis, particularly to agricultural extension on farms. Ekboir(2003)[10] studied why impact analysis should not be used for research evaluation and what the alternatives are. Araji, Sim, and Gardner (1978)[2] applied ex ante approach to investigate returns to agricultural research and extension. Based on a sample of Wisconsin farmers, Chavas and Aliber(1993)[8] employed a nonparametric approach to analyze economic efficiency in agriculture. In [9] by a nonneutral stochastic frontier approach Dinar, Karagiannis and Tzouvelekas(2007) evaluated the impact of agricultural extension on farms performance in Crete.

A lot of research found that through extension service programs, the majority of farmers could increase their income and yields. In [22], Lohr and Park applied a nonlinear logit model to study the performance effectiveness ratings about cooperative extension and organic farmers. They found that part-time newer adopters of organic farming methods are more likely to rate extension service providers as effective providers. The authors in [17] examine an agricultural technology assistance program implemented by Non-governmental organizations in Bosnia-Herzegovina (BiH) which significantly increased agricultural incomes for participants. Through agriculture extension program, the majority of sustainable farmers thought that their yields had increased or remained the same.

About the data, to analyze the impact of extension services on farmer, some authors worked on panel data. They investigate the effects of income over multiple time periods for the same individual farmer and compare the change of income of the farmer before and after they seek help from these agricultural extension services. While others, due to the cost of survey, the data could not obtained through a long time. They collected cross-sectional data by observing the individual farmer at the same point of time, or without regard to differences in time. In this thesis, due to the cost reason, I use cross-sectional. Our data is based on the project of Dr. Hartarska and Dr. Nadolnyak. In this thesis, I only use part of data.

To measure the impact of Auburn sponsored programs on Alabama farmers, I employ standard statistical techniques. Among these techniques, the main technique in this thesis is multivariate statistical regression, including Hoteling T-Square, the categorial variable regression and regression with dummy variables. I apply SAS programming to analyze the results. Refer to [15] and [29] for the details about these statistical techniques or tools.

Evaluating the performance is a starting point for impact analysis. It is a great challenge to measure the performance. Some outcomes are affected by a program and measurable. For example, effects of income and the like. However, some outcomes are difficult to measure or unmeasurable. For example, effects of ability and tacit knowledge. In this thesis, I use return of investments and gross value of sales to measure the performance.

Among univariate statistics, Two Sample Hotelling T-square Test undertakes tests of the differences between the (multivariate) means of two populations. Two populations may correspond to two different groups. For example, I compare the income difference between two populations of farms. The first population is the population of farmers who used one of the Auburn services and the second population is the population of who did not use it. For both population of farmers, four measurements were taken (See Chapter 4 for details).

Two Sample Hotelling T-square Test is used to determine whether the means of two groups are equal. Our results indicate that there does exist difference between the means of two groups.

Next, I investigate what factors affect choices of farmers on Auburn services. Such problems belong to discrete choices models. Among those, binary logit models can be used to examine choices of Auburn services by farmers. The results indicate that many factors, such as education, age and return of investments may affect the choices of farmers. The results will be shown in Chapter 4.

To investigate the impact of choice of Auburn related services on return of investments, I apply ordinal logit models instead of binary logit models. I interchange the ordinal variable Roachange (return of investments) with the service variable (Choice of Auburn related services) in the above model. The results will be shown in the Chapter 5.

Finally, we are interested in the impact of Auburn services and the performances of farmers in Alabama. Statistical regression methods with dummy variables will play critical roles in chapter 6. For example, if the farmer is in group 1, we define the dummy to be 1 and otherwise 0. By regression, we want to better predict the dependent variable and evaluate the difference between these two groups.

The rest of the thesis is organized as follows. In chapter 2, I summarize the data. In chapter 3, I introduce the methodology that I apply. In chapter 4, I use Two-sample Test Statistics to investigate whether the two means of two populations are the same and Binary Logit Models will be used to determine how the impact factors will influence choices of Auburn related services by farmers. In chapter 5, I investigate the impact of choice of the service by AU on the return of investments. In chapter 6, I use statistical regression involved dummy variables in explanatory variables to study the impact of choices of Auburn related services by farmers on gross value of sales. The thesis will end up with conclusions in chapter 7.

Chapter 2

The Data

In this chapter, I summarize the data. The present study is based on statistical analysis of a survey to a random sample of Alabama farmers. 389 completed questionnaires were read. The survey served as the basis for the ensuing analyses.

During these variables, some are continuous quantity variables, and others are discrete variables, particularly categorical variables. Examples for continuous variables are Started and so on. Description of continuous variables used in regression analysis of the thesis are shown in Table 2.1.

Table 2.1: Description of Continuous Variables

Variable Name	Description of the Continuous Variables
Started	Years since the current farming operation was established
MGrossvalue	Total gross value of 2009 sales at the mean
Farming Experience	Years of experience the respondent had
MBegass	The value of the total assets at the mean
Age	Age
Roa	Percent changed of return of investments

The categorical variable here refers to as a binary, ordinal, nominal or event count variable. A set of data is said to be categorical if the values of observations belonging to it can be sorted according to category. Each value is chosen from a set of disjoint categories. The following are some examples.

(i) Binary data;

A binary data is one that takes the values 0 or 1 to indicate whether a record belongs to a category or not. Typical example is Sex (M/F). In Chapter 2, our response variable, whether the farmer used certain Auburn related service (Y/N) will be of such type. I define five service variables related to Auburn University in Table 2.2.

Table 2.2: Description of Binary Variables

Variable Name	Description of the Category Variables
Aures	1 if the respondent used service from AU research stations; 0 otherwise
Extagent	1 if the respondent reported County or Regional extension agents useful; 0 otherwise
Auprof	1 if the respondent sought help from AU professors ; 0 otherwise
Auweb	1 is the respondent used AU website; 0 otherwise
Agriclin	1 if the respondent used Climate Forecasts at Agroclimate.org; 0 otherwise.
Sex	Gender(1:Female; 0:Male)
Marital Status	(1:Married; 0:Single/Divorced/Widowed)

(ii) Nominal data;

Nominal data are categorical data where the order of the categories is arbitrary. A good example in our questionnaire is race/ethnicity, which has values A=Black or African American, B=Asian or Pacific Islander, C=Other, D=White or Caucasian, E=American Indian or Alaska Native and F=Spanish, Hispanic or Latino Origin. Note that the order of the categories is arbitrary.

Table 2.3: Description of Nominal Variables

Variable Name	Description of the Nominal Variables
Race	race/ethnicity A. Black or African American B. Asian or Pacific Islander C. Other(specify) D. White or Caucasian E. American Indian or Alaska Native F. Spanish,Hispanic or Latino Origin

(iii) Ordinal data;

Ordinal data are categorical data where there is a logical ordering to the categories. A good example is the Likert scale that you see on many surveys: 1=Strongly disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly agree. In our questionnaire, one question is "If you were to need additional financing now, how difficult would it be to use the following sources?" The options for loans from friends and family are 1=Not difficult; 2=Somewhat difficult;

3=Quite difficult; 4=Very difficult;5=No access. See Table 2.4 for ordinal data used in this thesis.

Table 2.4: Description of Ordinal Variables

Variable Name	Description of the Ordinal Variables
Grossvalue	Total gross value of 2009 sales
	"1" <5,000
	"2" 5,000-9,999
	"3" 10,000-24,999
	"4" 25,000-49,999
	"5" 50,000-99,999
	"6" 100,000-249,999
	"7" 250,000-499,999
	"8" 500,000-999,999
	"9" 1,000,000 or more
	"10" None
Begass	The value of the total assets
	"1"-"10" the same as Grossvalue above.
Educ	The highest level of formal education the respondent completed
	1. Some high school or less
	2. Graduated high school
	3. Some college or technical school
	4. College graduate
	5. Some graduate school
	6. Masters degree or higher
Roachange	Return of investments
	1 if return of investments decreased;
	2 if return of investments unchanged;
	3 if return of investments increased;

Table 2.5 shows the summary statistics of continuous variables, including the means, minimum and maximum. According to Bureau of Labor Statistics, about forty percent of farmers in this country are 55 years old or older and the average age of farmers continues to rise. In our data, the average age of farmers is 52.803. The total gross value of 2009 sales averaged $\$38,871$. The value of total assets averaged $\$179,488$. Most farmers has over 10 years of experience in farming and they had around 7 years in the current farming. The mean of Roa is 1.382, which means farmers obtained 1.382% increased return of investments.

Table 2.5: Means of Continuous Variables

Variable	Mean	Std Dev	Minimum	Maximum
Started	7.709	8.548	0	57.000
MGrossvalue	38,871	111,190	0	1000,000
MBegass	179,488	258,970	2,500	1000,000
Farming Experience	10.521	13.159	0	65.000
Age	52.803	12.587	23.000	87.000
Roa	1.382	10.235	-80.000	99.000

For categorical variables, means do not make any sense. I apply frequency analysis and show the percentage statistics. Tables 2.6-2.12 show the percentage of respondents using Auburn services. The results indicate that only 32.13% of respondents used Auburn research extensions. By observation of Table 2.6, the percentage of respondents using Auburn research extensions is higher for the respondents who are younger or have higher education. The ratio of the percentage of farmers who used Auburn research extensions for 50 or less to 51 and over is 38.32/27.48. Table 2.6 shows that the percentage of farmers who used Auburn research extensions is 17.39%, 25.27%, 32.28%, 32.05%, 50.00% and 44.64% corresponding to the education level high school or less, graduate high school, some college school, college graduate, some graduate school, master or higher respectively. The percentage is nearly monotonic in creasing with respect to the education level. Refer to Table 2.6 for more details, where Column N means the percentage of farmers who never used Auburn research extensions and Column Y means the percentage of farmers who used. There are three subgroups for Y, which are Y1: Not helpful, Y2: Somewhat and Y3: Very Helpful. Columns N and Y(Y1,Y2,Y3) in Tables 2.7-2.10 have the same meanings as in Table 2.6. Among the group of farmers who used certain Auburn related service, most agreed that Auburn related service has been somewhat helpful or very helpful.

Table 2.7 shows the percentage of respondents using County or Regional Extension Agents. Compared with the previous case with AU research extensions, the results indicate that County or Regional Extension Agents is more popular service among farmers. There are 56.04% of respondents who used County or Regional Extension Agents. By observation

Table 2.6: Percentage Statistics for AU Research Stations

AU Research stations	N (Nonuser%)	Y (User%)	Y1 (Not Helpful%)	Y2 (Somewhat%)	Y3 (Very Helpful%)
Full Sample	264	125	6	69	50
	(67.87)	(32.13)	(4.80)	(55.20)	(40.00)
Age of farmers					
50 or less	103	64	4	37	23
	(61.68)	(38.32)	(6.25)	(57.81)	(35.94)
51 and over	161	61	2	32	27
	(72.52)	(27.48)	(3.28)	(52.46)	(44.26)
Level of Education					
High school or less	19	4	0	2	2
	(82.61)	(17.39)	(0)	(50)	(50)
Graduated high school	68	23	1	12	10
	(74.73)	(25.27)	(4.35)	(52.17)	(43.48)
Some college school	86	41	1	24	16
	(67.72)	(32.28)	(2.44)	(58.54)	(39.02)
College graduate	53	25	3	13	9
	(67.95)	(32.05)	(12.00)	(52.00)	(36.00)
Some graduate school	7	7	0	6	1
	(50.00)	(50.00)	(0.00)	(85.71)	(14.29)
Masters degree or higher	31	25	1	12	12
	(55.36)	(44.64)	(4.00)	(48.00)	(48.00)

of Table 2.7, the percentage of respondents using County or Regional Extension Agents is almost the same between age 50 or less and 51 or above, 59.28/53.60. Similar to the previous case, better educated farmers are more likely to seek help from County or Regional Extension Agents. For example, the percentage for those who have Master or Ph.Ds is 73.21%.

Table 2.7: Percentage Statistics for County or Regional Extension Agents

Extension Agents	N (Nonuser%)	Y (User%)	Y1 (Not Helpful%)	Y2 (Somewhat%)	Y3 (Very Helpful%)
Full Sample	171	218	15	124	79
	(43.96)	(56.04)	(6.88)	(56.88)	(36.24)
Age of farmers					
50 or less	68	99	7	56	36
	(40.72)	(59.28)	(7.07)	(56.57)	(36.36)
51 and over	103	119	8	68	43
	(46.40)	(53.60)	(6.72)	(57.14)	(36.14)
Level of Education					
High school or less	12	11	0	8	3
	(52.17)	(47.83)	(0.00)	(72.73)	(27.27)
Graduated high school	45	46	3	25	18
	(49.45)	(50.55)	(6.52)	(54.35)	(39.13)
Some college school	62	65	5	38	22
	(48.82)	(51.18)	(7.69)	(58.46)	(33.85)
College graduate	36	42	3	25	14
	(46.15)	(53.85)	(7.14)	(59.52)	(33.34)
Some graduate school	1	13	1	10	2
	(7.14)	(92.86)	(7.69)	(76.92)	(15.39)
Masters degree or higher	15	41	3	18	20
	(26.79)	(73.21)	(7.32)	(43.90)	(48.78)

Table 2.8 shows the percentage of respondents who sought help from Auburn professors. The results indicate only 19.54% of respondents, that is, only 1/5 of farmers did so. By observation of Table 2.8, the percentage of respondents who sought help from Auburn professors is much higher for the respondents who are younger or have higher education. Among the group who sought help from AU professors, most people (over 90%) responded that this is somewhat helpful or very helpful.

Table 2.8: Percentage Statistics for AU Professors

AU Professors	N (Nonuser%)	Y (User%)	Y1 (Not Helpful%)	Y2 (Somewhat%)	Y3 (Very Helpful%)
Full Sample	313	76	4	45	27
	(80.46)	(19.54)	(5.26)	(59.21)	(35.53)
Age of farmers					
50 or less	125	42	4	25	13
	(74.85)	(25.15)	(9.52)	(59.52)	(30.96)
51 and over	188	34	0	20	14
	(84.69)	(15.31)	(0.00)	(58.82)	(41.18)
Level of Education					
High school or less	20	3	0	1	2
	(86.96)	(13.04)	(0.00)	(33.33)	(66.67)
Graduated high school	78	13	1	10	2
	(85.71)	(14.29)	(7.69)	(76.92)	(15.39)
Some college school	107	20	1	14	5
	(84.25)	(15.75)	(5.00)	(70.00)	(25.00)
College graduate	58	20	2	10	8
	(74.36)	(25.64)	(10.00)	(50.00)	(40.00)
Some graduate school	10	4	0	4	0
	(71.43)	(28.57)	(0.00)	(100.00)	(0.00)
Masters degree or higher	40	16	0	6	10
	(71.43)	(28.57)	(0)	(37.50)	(62.50)

Table 2.9 shows the percentage of respondents using AU website. The results indicate that only 28.02% of respondents used AU website. By observation of Table 2.9, The ratio of the percentage of farmers who used AU website for 50 or less to 51 and over is 37.13/21.17. Farmers who have some graduate school or Master or Ph.Ds are more likely to seek help from AU website.

Table 2.9: Percentage Statistics for AU Website

AU Website	N (Nonuser%)	Y (User%)	Y1 (Not Helpful%)	Y2 (Somewhat%)	Y3 (Very Helpful%)
Full Sample	280	109	5	60	44
	(71.98)	(28.02)	(4.59)	(55.05)	(40.36)
Age of farmers					
50 or less	105	62	5	34	23
	(62.87)	(37.13)	(8.07)	(54.84)	(37.09)
51 and over	175	47	0	26	21
	(78.83)	(21.17)	(0.00)	(55.32)	(44.68)
Level of Education					
High school or less	21	2	0	1	1
	(91.30)	(8.70)	(0.00)	(50.00)	(50.00)
Graduated high school	70	21	1	13	7
	(76.92)	(23.08)	(4.76)	(61.91)	(33.33)
Some college school	95	32	1	15	16
	(74.80)	(25.20)	(3.13)	(46.88)	(49.99)
College graduate	52	26	2	16	8
	(66.67)	(33.33)	(7.69)	(61.54)	(30.77)
Some graduate school	9	5	0	4	1
	(64.29)	(35.71)	(0.00)	(80.00)	(20.00)
Masters degree or higher	33	23	1	11	11
	(58.93)	(41.07)	(4.34)	(47.83)	(47.83)

Table 2.10 shows the percentage of respondents using agroclimate.org. The results indicate that only 12.34% of respondents used Agroclimate.org, which means only 1/8 of farmers seek help from agroclimate.org. Among these respondents, they are more likely to have higher education. Agroclimate.org, a Service of the Southeast Climate Consortium, is ranked number 10,653,353 in the world according to the Alexa Traffic Rank, which shows that this is not a well-known website in the world.

Table 2.10: Percentage Statistics for Climate Forecasts at Agroclimate.org

Agroclimate.org	N (Nonuser%)	Y (User%)	Y1 (Not Helpful%)	Y2 (Somewhat%)	Y3 (Very Helpful%)
Full Sample	341	48	10	31	7
	(87.66)	(12.34)	(20.83)	(64.58)	(14.59)
Age of farmers					
50 or less	141	26	8	15	3
	(84.43)	(15.57)	(30.77)	(57.69)	(11.54)
51 and over	200	22	2	16	4
	(90.09)	(9.91)	(9.09)	(72.73)	(18.18)
Level of Education					
High school or less	21	2	0	1	1
	(91.30)	(8.70)	(0)	(50)	(50)
Graduated high school	81	10	3	5	2
	(89.01)	(10.99)	(30.00)	(50.00)	(20.00)
Some college school	112	15	3	8	4
	(88.19)	(11.81)	(20.00)	(53.33)	(26.67)
College graduate	69	9	4	5	0
	(88.46)	(11.54)	(44.44)	(55.56)	(0.00)
Some graduate school	10	4	0	4	0
	(71.43)	(28.57)	(0.00)	(100.00)	(0.00)
Masters degree or higher	48	8	0	8	0
	(85.71)	(14.29)	(0.00)	(100.00)	(0.00)

Table 2.11 and Table 2.12 show Percentage Statistics for Service Variable vs Begass (the value of total assets) and Service Variable vs Grossvalue (total gross value of 2009 sales) respectively. By observation of Table 2.11, for those Non-user group of AU research stations (Aures =0) and the value of total assets less than 5,000 dollars (Begass=1), We have number 47 (70.15%), which means the frequency is 47 and the percent is 47/(47+20)=70.15% in the group when Begass=1. In Chapter 4 and 5, the Begass and Grossvalue will be used in the right hand side as explanatory variables. In such situation, we have to choose a reference category to define the dummy variable. In the thesis, I choose the group with the most observations for total gross value of 2009 sales and the value of total assets. By observation of Tables 2.11-2.12, we can see that the reference groups are the group with the value of total assets between 100,000 to 249,999 (Begass=6) for Begass and the group with total gross value of 2009 sales less than 5,000 dollars (Grossvalue=1) for Grossvalue respectively.

Table 2.11: Percentage Statistics for Service Variable vs Begass

Total Assets($)	<5K	[5K,10K)	[10K,25K)	[25K,50K)	[50K,100K)	[100K,250K)	[250K,0.5M)	[0.5M,1M)	>1M
AU research stations									
0	47	24	27	29	21	51	26	21	8
	(70.15)	(68.57)	(67.50)	(69.05)	(61.76)	(68.00)	(66.67)	(67.74)	(61.54)
1	20	11	13	13	13	24	13	10	5
	(29.85)	(31.43)	(32.50)	(30.95)	(38.24)	(32.00)	(33.33)	(32.26)	(38.46)
Extension agents									
0	41	13	15	18	14	30	14	11	7
	(61.19)	(37.14)	(37.50)	(42.86)	(41.18)	(40.00)	(35.90)	(35.48)	(53.85)
1	26	22	25	24	20	45	25	20	6
	(38.81)	(62.86)	(62.50)	(57.14)	(58.82)	(60.00)	(64.10)	(64.52)	(46.15)
AU professors									
0	57	28	35	36	25	59	29	25	7
	(85.07)	(80.00)	(87.50)	(85.71)	(73.53)	(78.67)	(74.36)	(80.65)	(53.85)
1	10	7	5	6	9	16	10	6	6
	(14.93)	(20.00)	(12.50)	(14.29)	(26.47)	(21.33)	(25.64)	(19.35)	(46.15)
AU website									
0	53	28	27	34	21	53	26	21	6
	(79.10)	(80.00)	(67.50)	(80.95)	(61.76)	(70.67)	(66.67)	(67.74)	(46.15)
1	14	7	13	18	13	22	13	10	7
	(20.90)	(20.00)	(32.50)	(19.05)	(38.24)	(29.33)	(33.33)	(32.26)	(53.85)
Agriclin									
0	60	30	33	39	30	61	35	27	13
	(89.55)	(85.71)	(82.50)	(92.86)	(88.24)	(81.33)	(89.74)	(87.10)	(100.00)
1	7	5	7	3	4	14	4	4	0
	(10.45)	(14.29)	(17.50)	(7.14)	(11.76)	(18.67)	(10.26)	(12.90)	(0.00)

Table 2.12: Percentage Statistics for Service Variable vs Grossvalue

Total Gross Value ($)	<5K	[5K,10K)	[10K,25K)	[25K,50K)	[50K,100K)	[100K,250K)	[250K,0.5M)	>1M	No Record
AU research stations									
0	139	37	13	8	11	17	2	1	36
	(74.33)	(75.51)	(44.83)	(47.06)	(73.33)	(47.22)	(33.33)	(33.33)	(76.60)
1	48	12	16	9	4	19	4	2	11
	(25.67)	(24.49)	(55.17)	(52.94)	(26.67)	(52.78)	(66.67)	(66.67)	(23.40)
Extension agents									
0	93	22	9	2	4	13	1	0	27
	(49.73)	(44.90)	(31.03)	(11.76)	(26.67)	(36.11)	(16.67)	(0.00)	(57.45)
1	94	27	20	15	11	23	5	3	20
	(50.27)	(55.10)	(68.97)	(88.24)	(73.33)	(63.89)	(83.33)	(100.00)	(42.55)
AU professors									
0	163	40	21	11	11	21	4	1	41
	(87.17)	(81.63)	(72.41)	(64.71)	(73.33)	(58.33)	(66.67)	(33.33)	(87.23)
1	24	9	8	6	4	15	2	2	6
	(12.83)	(18.37)	(27.593)	(35.29)	(26.67)	(41.67)	(33.33)	(66.67)	(12.77)
AU website									
0	138	39	18	11	11	20	1	2	40
	(73.80)	(79.59)	(62.07)	(64.71)	(73.33)	(55.56)	(16.67)	(66.67)	(85.11)
1	49	10	11	6	4	16	5	1	7
	(26.20)	(20.41)	(37.93)	(35.29)	(26.67)	(44.44)	(83.33)	(33.33)	(14.89)
Agriclin									
0	160	45	24	14	14	31	6	3	44
	(85.56)	(91.84)	(82.76)	(82.35)	(93.33)	(86.11)	(100.00)	(100.00)	(93.62)
1	27	4	5	3	1	5	0	0	3
	(14.44)	(8.16)	(17.65)	(11.90)	(6.67)	(13.89)	(0.00)	(0.00)	(6.38)

Chapter 3

Methodology

As I mentioned in the introduction, my goals in this thesis are to investigate what factors affect choices of farmers on services provided by Auburn University and the relation between the impact of Auburn related services and the performances of farmers in Alabama. Classical statistics methods will play important roles. The tools used are Two-sample Test Statistics, Discrete Choice Models and Statistical Regression with Dummy Variable.

3.1 Two-sample Test Statistics

In this section, I introduce how to investigate whether the two means of two populations are the same by Two-sample Test Statistics. There are some assumptions for this test to be valid. We assume that the observations are independent. The variances of the two samples may be assumed to be equal or unequal. The data from both groups are normally distributed.

Two Sample Hotelling T-square Test undertakes tests of the differences between the (multivariate) means of two populations. Two populations may correspond to two different groups. I compare the means of two populations of farmers. The first population is the population of farmers who used certain service from Auburn University (User group) and the second population is the population of who did not use this service (Non-user group). For both population of farmers, four measurements were taken, which are the following four variables: Started (the years of farming operation established), Farming Experience (years of experience the respondent had), Roa (percent changed for Return of Investments) and Age (the age of farmers). Here are some questions:

- Question 1: Does User group have the same means as Non-user group?
- Question 2: Does User group have larger mean for one of four measurements, for example,

the years of farming operation established than that in Non-user group?

• Question 3: Does User group have larger mean for one of four measurements than that in Non-user group by at least some pre-determined threshold amount?

First, are these means are significantly different? To test the difference of the means for the above four variables between those farmers who used Auburn related services(User group) and those who did not (Non-user group), we have the following hypothesis, both groups have equal means, against the alternative hypothesis that the means are not equal, that is, $H_0 : \mu_1 = \mu_2, H_A : \mu_1 \neq \mu_2$, where μ_1 is the mean vector of those who used Auburn Research Stations and services and μ_2 is the mean vector of those who did not.

To test this, I use Two-sample Hotelling's T-Square test statistic. Let x_1, x_2 denote the random variable with four components as above and let $\bar{x}_1, \bar{x}_2$ be the sample means in User group and Non-user group respectively. Then the Hotelling's T-Square can be defined by the following,

$$T^2 = (\bar{x}_1 - \bar{x}_2)'[(1/n_1 + 1/n_2)W]^{(-1)}(\bar{x}_1 - \bar{x}_2), \tag{3.1}$$

where n_1 is the number of observations in User group, n_2 is the number of observations in Non-user group and p is the number of components in the mean. W is the pooled sample covariance

$$W = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}.$$

where S_1 and S_2 are the sample variance from two groups, respectively. For large samples, this test statistic can be related to the F-distribution by with $(p, n_1 + n_2 - 1 - p)$ degrees of freedom. Next, define

$$F = \frac{n_1 + n_2 - p - 1}{(n1 + n2 - 2)p} T^2. \tag{3.2}$$

We would reject null hypothesis at level α if it exceeds the critical value from the F-table evaluated at α, that is, $F > F_{(p, n_1 + n_2 - p - 1)}(\alpha)$. Next step in our analysis is to answer Question 2 and 3. For one particular variable of four measurements, for example, the mean for total gross value of 2009 sales, we are interested in whether the mean in User group is larger than that in Non-user group by at least some pre-determined threshold amount. To this end, I construct the $(1-\alpha)\times 100\%$ Confidence Ellipse for the difference $\mu_1 - \mu_2$. Let $c^2 = \frac{(n1+n2-2)p}{n_1+n_2-p-1}F_{(p,n_1+n_2-p-1)}(\alpha)$ and then $(\bar{x}_1 - \bar{x}_2) \pm c\sqrt{[(1/n_1 + 1/n_2)W]}$ will cover $(\bar{x}_1 - \bar{x}_2)$ with $(1-\alpha)\times 100\%$ confidence.

3.2 Discrete Choice Models

Although in Chapter 2, the summary statistics suggest important relations between the choice of Auburn related services and the age and education and the two sample test statistics indicate difference of the means of some explanatory variables in two groups, the multivariate statistical techniques can be used to estimate the partial impact of each independent variable. The standard technique for modeling such individual choice behavior is the discrete choice model.

In this thesis, I study two types of choices: binary choice (User/ Non-user of AU related service) and ordinal choice (decreased/unchanged/increased return of investments). To deal with such models, I apply binary logit regression in Chapter 4 and ordinal logit regression in Chapter 5 respectively. To apply such theory, I compute odd ratios and marginal effects to measure the effects of each independent variable. In the end, I investigate the goodness of fit for logistic regression.

In discrete choice models, the dependent variable is a categorical variable, either binary or ordinal. In statistics, the binomial random variable X is the number of ”successes” in the fixed N trials. If there are only two outcomes, usually referred to as ”successes” and ”failure”, and assume trials are independent with the same ”successes” probability p, then X has a binomial distribution denoted by $B(N, p)$. In Chapter 4, our response variable (Service Variable) is binary type (See Table 2.2 for all five Service Variables), while in Chapter 5, our response variable (Roachange) is ordinal type (See Table 2.4). However, independent variables on the right-hand side (RHS) for both models are either continuous or categorical.

The basic technique for analysis of discrete choice models is logistic regression. A generalized linear model (GLM), or Ordinary Linear Square (OLS) is not appropriate for probability models. Although the OLS estimates are unbiased, the errors (residuals) from the linear probability model violate the homoskedasticity and normality of errors assumptions of OLS regression. These problems will lead to invalid standard errors and hypothesis tests. Another problem is that without any restriction for the response variable, we may get predictions either less than zero or greater than one. The probabilities of predictions outside the range (0,1) are meaningless. For a more thorough discussion about the linear probability model, refer to Long (1997)[23] or Aldrich and Nelson (1984) [1]. To overcome such drawbacks, the categorical dependent variable regression models (CDVMs) have drawn a lot attention. Daniel McFadden won the Nobel prize in 2000 for his pioneering work in developing the theoretical basis for discrete choice model. Regression models for discrete choice may take many forms, including Binary Logit, Ordinal Logit, Multinomial Logit, Multinomial Probit, Nested Logit and so on. Unlike the linear probability models (LPM), the CDVMs are not linear.

Logistic regression has several advantages over LPM. Obviously, it is a nonlinear model and it does not assume a linear relation between the independent variable and dependent variable, and so we can add some interaction terms or some power terms. Unlike LPM, we do not need the assumption of the equal variance and the normality of independent variables. But, even though logistic regression has more flexibility, it has its own disadvantages. The model requires much more data to achieve stable, meaningful results. Our data have 361 valid sample points, so logistic regression models are appropriate. Another disadvantage is that their results are less straightforward to interpret.

To get the estimates of parameters, I apply maximum likelihood estimation(MLE). Here I show MLE for the binary choice only. Let Y be the response, where they are 0s if the respondents do not use the service by Auburn University, 1s otherwise. Y is assumed to be binomially distributed. Suppose that Y_i is a Bernoulli random variable, we can write $P(Y_i = 1) = P_i$ and $P(Y_i = 0) = 1 - P_i$. For a binomial dependent variable, the binomial logistic model is appropriate. This approach will be used to explain farm and farmer characteristics that influence farmers' decisions to use AU related services. We have the following,

$$Log[P/(1-P)] = \beta_0 + \beta X, \tag{3.3}$$

where X=(Started, Farming Experience, Grossvalue, Begass, Age, Educ, Roachange), $logP$ is the logarithmic probability of using certain Auburn service at the farm level, $log(1-P)$ is the logarithmic probability of not using this Auburn service. In this thesis, I fit five models with service variable by Aures, Extagent, Auprof, Auweb and Agriclin respectively(See Tables 2.1-2.4 for description).

Suppose we have n observations. $f_i(Y_i)$ is the probability that $Y_i = 1$. The joint probability function $f(Y_1, ..., Y_n)$ is given by

$$f(Y_1, ..., Y_n) = \prod_1^n f_i(Y_i). \tag{3.4}$$

This is so called likelihood function. To simply the likelihood function, we obtain the so called log likelihood function(LLF) as the following

$$lnf(Y_1, ..., Y_n) = ln \prod_1^n f_i(Y_i). \tag{3.5}$$

It is easy to verify that

$$lnf(Y_1, ..., Y_n) = \sum_1^n (Y_i(\beta_0 + \beta X)) - \sum_1^n ln[1 + exp(\beta_0 + \beta X)]. \tag{3.6}$$

Finally, by maximizing the LLF, we can obtain the parameters β_0 and β.

Besides the estimation of parameters, other multivariate statistical techniques, such as odds ratios and marginal effects are calculated to examine the relationship between choices and other explanatory variables. Such approaches allow the partial impact of each explanatory variable to be estimated with all other covariates controlled. I also fit five models with the same independent variables but different service variable, which is Aures, Extagent, Auprof, Auweb or Agriclin respectively (See Tables 2.1-2.4 for their description). The results will be showed in Chapter 4.

In chapter 5, I would like to investigate the impact of choices of Auburn related services by farmers on their return of investments. I also use the similar techniques as the above. Then we have the following model, ordinal logit model,

$$Log[P1/(1-P1)] = \beta_{01} + \beta X, \tag{3.7}$$

and

$$Log[(P1+P2)/P3] = \beta_{02} + \beta X, \tag{3.8}$$

where X=(Started, Grossvalue, Farming Experience, Begass, Age, Educ, Service Variable), $P1$ ($P2$ or $P3$) is the probability of decrease (unchange or increase) in return of investments. Here Service Variable is corresponding to the choice of services of AU by farmers, that is, Aures, Extagent, Auprof, Auweb or Agriclin.

The ordinal logit regression is an extension of binary logit. For ordinal logit, we need one more assumption, the proportional odds assumption, that the relationship between any two pairs of outcome groups is statistically the same. This means that the coefficients that describe the relationship between, say, the decreased return of investments versus all unchanged or increased return of investments are the same as those that describe the relationship between the decreased or unchanged versus increased. Because the relationship between all pairs of groups is the same, we have the same coefficients for X. The techniques are similar and we use maximum likelihood to get the estimates of parameters. The results related to odds ratios and marginal effects will be showed in Chapter 5.

3.3 Regression with Dummy Variables

To investigate the impact of use of Auburn related services on performance of farmers in Alabama, I apply Regression with Dummy Variables. When scored as either a 0 or 1, dichotomies are often referred to as "dummy" variables. They indicate either the absence or presence of a characteristic or trait. Hence they function as a "dummy" for the variable in question. We can also use dichotomous variables in regression, not as response variable but as independent variables.

Consider the hypothesis that total gross value of sales depends on age, education, years since the current farming operation was established, years of experience, the value of total assets and some category variables, percent changed for return of investments, service variable (Aures, Extagent, Auprof, Auweb or Agriclin). To better predict the total gross value of sales (MGrossvalue), we could develop an equation of the form

$$Y = \beta_0 + \beta X + dD, \tag{3.9}$$

where: Y=MGrossvalue;

β_0 = Constant or intercept term;

X = (Started, Farming Experience, Begass, Age, Educ, Roa, Sex, Martial Status, Race);

β = Coefficient of X;

D = Service variable (Aures, Extagent, Auprof, Auweb, or Agriclin);

d = Coefficient of D.

To define dummy variable, at least one category must always be omitted which is with the value of zero and should be the the one best suited to be the reference value for all the other categories. In our case, we have only two categories. The omitted category becomes the reference category against which the effect of the other categories are assessed. We can interpret the results as the difference between each category and this omitted category.

If D=1, which means, the farmer used certain Auburn related services, we have

$$Y = (\beta_0 + d) + \beta X. \tag{3.10}$$

If D=0, which means, the farmer did not use this service, we have

$$Y = \beta_0 + \beta X. \tag{3.11}$$

These regression equations are graphed in the following figure.

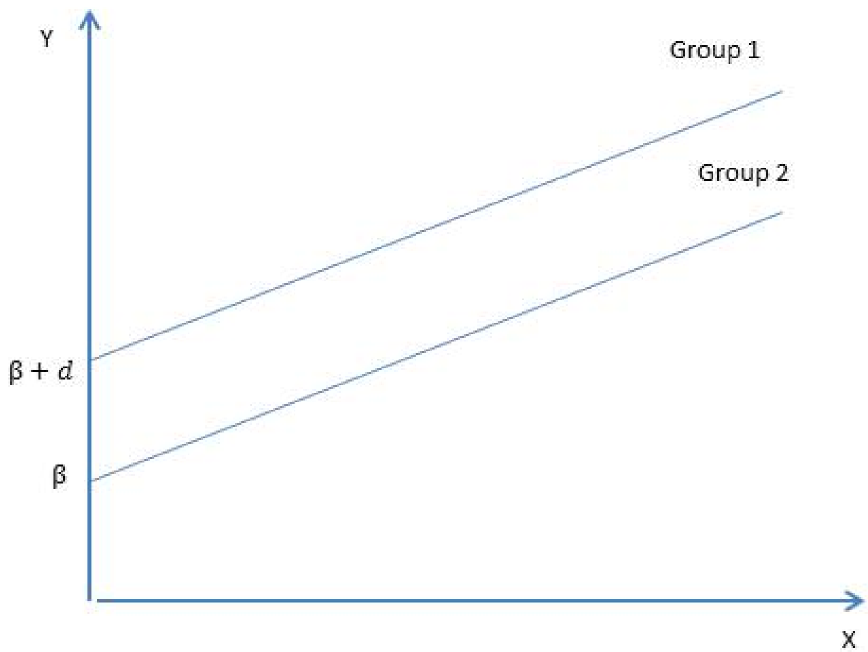

Figure 3.1: Parallel lines

d gives the difference in intercepts for the two regression lines. Because these regression lines are parallel, d also represents the constant separation between the lines, the expected total gross value of sales advantage accruing to farmers in User group when other variables are held constant. If farmers in User group were disadvantaged relative to those in Non-user group, then d would be negative. β_0 gives the intercept for which D =0. β is the common slope.

So far we only considered models without interaction. Effect of one independent variable on the dependent variable depends on the value of the other independent variable. Two variables are said to interact in determining a dependent variable if the partial effect of one depends on the value of the other. We have three types of interactions. Interaction between a quantitative and a qualitative variable means that the regression surfaces are not parallel. Interaction between two qualitative variables means that the effect of one of the variables depends on the value of the other variable. Interaction between two quantitative variables is a bit harder to interpret.

I consider some interaction terms, for example, Started and Farming Experience. The model will have the form

$$Y = \beta_0 + \beta X + dD + \gamma Z, \tag{3.12}$$

where Z = some interaction terms;

γ= the coefficients of Z. I determine the exact variables Z in Chapter 6. Such regression equations are graphed in the following figure.

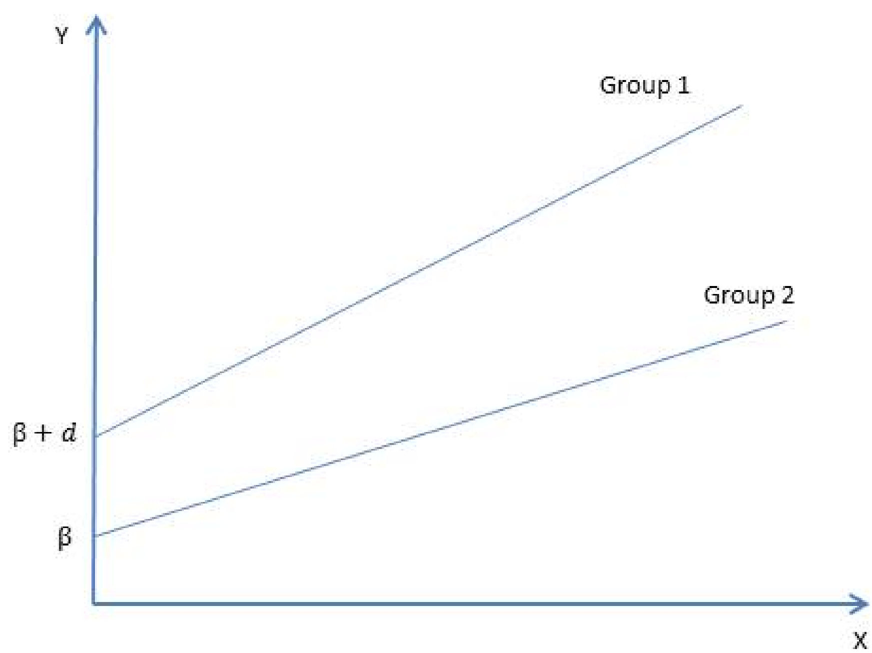

Figure 3.2: Predictors with Interaction: Non-Parallel lines

The results will be shown in chapter 6.

Chapter 4

Factors Affecting Use of Auburn University Services

In this chapter, I investigate impact factors on choices of Auburn related services. The Alabama farmers are classified as either User group or Non-user group by whether farmers used or not one of five services (Aures, Extagent, Auprof, Auweb or Agriclin) by Auburn University. First, I use Two-sample Test Statistics to investigate whether the two means of two populations are the same. Then, to investigate further the impact factors on choices of Auburn related services, I employ binary response logit regression. Specifically, I use the following logit model

$$Log[P/(1-P)] = \beta_0 + \beta X, \tag{4.1}$$

where X=(Started, Farming Experience, Grossvalue, Begass, Age, Educ, Roachange), $logP$ is the logarithmic probability of using certain Auburn service at the farm level, $log(1-P)$ is the logarithmic probability of not using this Auburn service. In this thesis, I fit five models with the response variables Y by Aures, Extagent, Auprof, Auweb and Agriclin respectively(See Tables 2.1-2.4 for description).

I perform a sequence of analyses. I compute the odds ratios and marginal effects to measure the effects of each independent variable and investigate the goodness of fit for logistic regression.

4.1 Two Samples: Difference between Two Means

First, I divide the whole population by user and non-user of certain AU related service into two groups. We are interested in four variables Started, Farming Experience (ExpFarm), Age and Roa. The following table is the means in two groups by different service variable.

Table 4.1: Means of Continuous Variables (Std in Parentheses)

Variable	**Total**	**AU res.**		**Extension agents**		**AU professors**		**AU websites**		**Agriclin**	
		User	Nonuser	User	Nonuser	User	Nonuser	User	Nonuser	User	Nonuser
Started	7.7 (8.55)	8.1 (9.33)	7.5 (8.15)	7.9 (8.95)	7.4 (8.01)	9.8* (12.22)	7.2 (7.27)	8.0 (10.34)	7.6 (7.73)	7.6 (7.54)	7.7 (8.7)
ROI(%)	1.38 (10.24)	1.46 (6.78)	1.35 (11.57)	1.72 (10.42)	0.94 (10.01)	2.89* (6.33)	1.00 (10.98)	2.06 (5.79)	1.11 (11.54)	1.56 (5.66)	1.36 (10.76)
ExpFarm	10.5 (13.16)	13.1*** (14.04)	9.2 (12.53)	10.8 (13.72)	10.1 (12.41)	12.0 (13.11)	10.1 (13.17)	12.1 (13.06)	9.9 (13.17)	13.0 (12.63)	10.1 (13.22)
Age	52.8 (12.59)	50.7** (12.99)	53.8 (12.27)	52.2 (12.28)	53.5 (12.97)	49.9** (12.74)	53.6 (12.46)	49.4*** (12.2)	54.2 (12.5)	49.6* (12.10)	53.3 (12.61)

Statistically significant difference in means *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Are these means are significantly different? To answer this question, we have the following hypothesis, both groups have equal means, against the alternative hypothesis that the means are not equal, that is, $H_0 : \mu_1 = \mu_2, H_A : \mu_1 \neq \mu_2$, where μ_1 is the mean vector of those who used Auburn related services and μ_2 is the mean vector of those who did not.

For example, we have two groups divided by Aures. With $\alpha = 0.05, F_{4,356} = 2.397$, and $F = 3.196 > F_{4,356} = 2.397$, we can reject H_0, that is, the mean vector of those who used Auburn related services is different with the mean vector of those who did not.

The following table is the test results for all four variables for all five pairs of groups.

Table 4.2: Hoteling T-square Test Results

	Aures	Extagent	Auprof	Auweb	Agriclin
T2	12.891	2.147	15.318	14.622	5.218
F value	3.196	0.532	3.797	3.625	1.289
F critical value (4,359)	2.397	2.397	2.397	2.397	2.3970
Test results	reject	fail to reject	reject	reject	fail to reject

Based on the results, we can conclude that the means of User group are different with those of Non-user group of using service from Auburn research stations, Auburn professor and Auburn websites. Next step in our analysis is to determine the mean of which variable has difference. We set $\alpha = 0.05$. For simplicity, we denote μ_{1i} and μ_{2i} be the i-th component of μ_1 and μ_2 that is the means of population in User group and Non-user group respectively. 0 is in all confidence intervals except for (Age, Auweb). If 0 is in the interval, then we can not distinguish the difference of means in two groups. On the other hand, if 0 is not in the interval, say, for (Age, Auweb), we are 95% confident to conclude that the age of the respondent in User group is at least 0.31 year younger than those in Non-user group. The following table shows the 95% simultaneous confidence interval for the mean difference.

Table 4.3: 95% Simultaneous Confidence Interval for Difference between Two Means

Mean Difference	Aures	Extagent	Auprof	Auweb	Agriclin
Started	-2.43, 3.53	-2.32, 3.33	-0.86, 6.06	-2.64, 3.56	-4.43, 4.12
Farming Experience	-0.67, 8.42	-3.60, 5.10	-3.45, 7.27	-2.56, 6.97	-3.74, 9.44
Age	-7.51, 1.20	-5.48, 2.84	-8.80, 1.40	-9.31, -0.31	-9.96, 2.63
Percent changed of ROI	-3.46, 3.68	-2.60, 4.17	-2.28, 6.05	-2.77, 4.66	-4.84, 5.26

We raise the level to 0.1, we have the following table the 90% simultaneous confidence interval for the pairs of mean components. 0 is not in confidence intervals for pair (Age, Auweb). We are 90% confident to conclude that, the age of the respondent in User group is at least 0.74 year younger than those in Non-user group.

Table 4.4: 90% Simultaneous Confidence Interval for Difference between Two Means

Mean Difference	Aures	Extagent	Auprof	Auweb	Agriclin
Started	-2.15, 3.24	-2.05, 3.06	-0.53, 5.73	-2.34, 3.27	-4.02, 3.71
Farming Experience	-0.24, 7.98	-3.19, 4.68	-2.94, 6.75	-2.10, 6.51	-3.11, 8.81
Age	-7.09, 0.79	-5.08, 2.44	-8.32, 0.91	-8.87, -0.74	-9.35, 2.03
Percent changed of ROI	-3.12, 3.34	-2.28, 3.84	-1.88, 5.65	-2.41, 4.30	-4.36, 4.77

We raise the level to 0.2, and we have the following table the 80% simultaneous confidence interval for the pairs of mean components. In this case, we have two pairs (Farming Experience, Aures), (Age, Auweb) which have 0 out of the confidence intervals. We are 80% confident to conclude that,

(i)for Aures, the years of experience the respondent had in User group is at least 0.27 year older than those in Non-user group;

(ii) for Auweb, the age of the respondent in User group is at least 1.24 year younger than those in Non-user group.

Table 4.5: 80% Simultaneous Confidence Interval for Difference between Two Means

Mean Difference	Aures	Extagent	Auprof	Auweb	Agriclin
Started	-1.81, 2.91	-1.74, 2.74	-0.15, 5.34	-2.00, 2.92	-3.54, 3.23
Farming Experience	0.27,7.47	-2.70, 4.20	-2.34, 6.15	-1.57, 5.98	-2.37, 8.07
Age	-6.61, 0.30	-4.61,1.98	-7.74, 0.34	-8.37, -1.24	-8.65, 1.32
Percent changed of ROI	-2.72, 2.94	-1.90, 3.47	-1.42, 5.19	-2.00 3.89	-3.79, 4.21

4.2 Odds Ratios

The odds of event are defined as the ratio of the probability of event(success) over the probability of nonevent(failure). Odds ratios provide a method of describing the strength of the partial relationship between an individual predictor and the predicted event. About the odds ratio, we have the following properties. The odds increase as the probability increases or vice versa. Since probability ranges from 0 to 1, odds range from 0 to positive infinity. Since the log is a strictly increasing function, the greater the odds, the greater the log of odds and vice versa.

Consider the general logistic regression model,

$$logit(P) = log(P/(1-P)) = \beta_0 + \beta X,$$

where the logistic regression of Y on $X = (x_1, ..., x_k)$ estimates parameter values for $\beta_0, \beta = (\beta_1, ..., \beta_k)$ for some positive integer k.

Note that the equation above is equivalent to

$$P = exp(\beta_0 + \beta X)/(1 + exp(\beta_0 + \beta X)).$$

Assume that x_i is changed to $x_i + 1$ and the others hold constant for $i = 1, ..., k$. Then the coefficient β_i for x_i is the difference in the log odds, which implies that the quotient of odds is e^{β_i}. In other words, for a one-unit increase in x_i, the expected change in odds is $e^{\beta_i} - 1$ times.

For example, for the Age variable in the Model with service variable Auburn research stations, $e^{-0.00863} = 0.991$. For a one-unit increase in farmer's age, the expected change in log odds is 0.991. It is not so significant since odds ratios decrease by only 0.9%. Anyway, younger farmers are slightly more likely to seek service from the AU Research Stations.

For the education level, I choose the group with farms of some high school or less(Educ=1) as the reference category. For the response variable variable with Aures, exponentiating the Educ coefficients I obtain odds ratios of 3.321, 5.480, 5.626, 18.796 and 11.967. The odds of probability of use of Auburn research stations to that of Non-use increase by 232%, 448%, 463%, 1780% and 1097% as I move the education level from some high school or less(Educ=1) to graduated high school(Educ=2), some college school(Educ=3), college school(Educ=4), some graduate school(Educ=5) and masters degree or higher(Educ=6) respectively.

For the value of total assets (Begass) and total gross value of 2009 sales (Grossvalue), our reference categories are chosen as the group with the most observations: with total assets of farmers between \$100,000 and \$249,999 (Begass=6), with total gross value of sales of farmers less than \$5,000 (Grossvalue=1). For return of investments (Roachange), we choose increased return of investments (Roachange=3) as reference category.

4.3 Marginal Effects

According to the parameter estimate, the coefficients of the model are directly related to the probability of the ratio of farmers who used the services and those who did not. We can view these parameters as the main effects of the corresponding factors. Recall that

$$p = \frac{exp(\beta_0 + \beta X)}{1 + exp(\beta_0 + \beta X)}.$$

It is not easy to express the effect on the probability of increasing a predictor by one unit while holding the other variables constant. To avoid redundancies I use dummies and reference base for categorical variables. The intercept is the value of the ratio when the value of all independent variables are zero. Therefore, a positive coefficient of continuous variable indicates that as the corresponding variable goes up one unit, it will increase log odds ratios, while a negative regression coefficient means that the corresponding variable will decrease the log odds ratios. A large regression coefficient means that the impact factor strongly influences the ratio, while a near-zero regression coefficient means that that impact factor has little influence on the ratio. Since log function is monotone increasing, odds ratios will move along the same direction as log odds ratio.

As we see, the coefficients from these models are difficult to interpret because they measure the change in the unobservable associated with a change in one of the explanatory variables. A more useful measure is what we call the marginal effects. We only focus on the probability models, which are nonlinear models. Different from the LPM, the parameters can not directly be interpreted as marginal effects on the dependent variable.

For continuous independent variables, marginal effects measure the expected instantaneous change in the dependent variable as a function of a change in a certain explanatory variable while keeping all the other covariates constant. They are obtained by computing the derivative of the conditional mean function with respect to x. Recall that

$$P = exp(\beta_0 + \beta X)/(1 + exp(\beta_0 + \beta X)).$$

Let ME_k denote the marginal effect with respect to x_k, the k-th independent variable of X. Then we have

$$ME_k = \frac{\partial P}{\partial x_k} = \beta_k P(1 - P).$$

Marginal effect with respect to x_k depends on the coefficient β_k and the probability P of using certain AU related service. If β_k is positive, then its marginal effect is positive, which means that the probability of using certain AU related service will increase as x_j increases.

For categorical variables, the effects of discrete changes are computed, that is, the marginal effects for categorical variables show how P is predicted to change as X_k changes from itself to the reference category holding all other predictors equal. For example, suppose we would like to compute the marginal effect for decreased return of investments (Roachange=1). Recall that our reference group is the group with increased return of investments (Roachange=3). We consider the model associated with binary choice for Auburn research stations(Aures). The corresponding marginal effect is given by

$$ME(Roachange(1)) = P(Aures = 1|X; Roachange = 1) - P(Aures = 1|X; Roachange = 3).$$

There are two ways to view the "average" or "overall" marginal effects for the predictors: marginal effects at the mean(MMEs) and averaged marginal effects(AMEs). MMEs are computed by setting Xs at their means. The use of means to compute marginal effect is criticized because no real one may actually have values at means and for categorical variable like Sex (0, Male;1 Female), a value like 0.42 makes no sense. Even though MMEs is popular, but many think that AMEs are superior. With AMEs, a marinal effect is computed for each observation, and then average all individual marginal effects. I use AMEs in this thesis. More details refer to Greene 1997 (p. 876).

I obtained the information about the average marginal effects(AME) to our five models in Table 4.7-4.11. I mark statistically significant variables with *** for 1% level of significance, ** for 5% and * for 10% for all tables throughout this thesis. No one is in the group with total gross value of sales between $500,000 and $999,999 (Grossvalue=8), so I omitted it.

I consider all seven factors in the following. Odds ratios, parameter estimates, standard errors and average marginal effects are shown in Tables 4.6-4.11.

Return of Investments of the Farmers(Roachange)

These tables indicate that the return of investments of farmers is an important factor of using certain service of Auburn. All coefficients of Roachange are negative. The negative parameter estimates show a negative effect for the odds P/(1-P). The odds P/(1-P) decreases which implies the probability P decreases. In other words, when farmers obtain a decreased or unchanged return of investments, they will be less likely to seek help from this service compared to farmers with an increased return of investments. In the current case, for Roachange and AU research stations, their average marginal effects -0.168 and -0.199 tell us that, the predicted probabilities of using this service are 0.168 and 0.199 less for the individual with decreased return of investments than for one who has increased return of investments.

Education of the Farmers(Educ)

Tables 4.6-4.11 indicate that the education level of farmers plays an active factor in using the services of Auburn. The odds of User to Non-user of Auburn research stations are monotone increasing as the education level from high school or less to some graduate school and then decrease for Masters degree or higher. As we see, compared to the farms with some high school or less (Educ=1), all odds are greater than 1 for Auburn research stations. For the other four models, we see the similar stories, which show monotone increasing odds as the education level move from low to some graduate school for all four services and then decrease to Masters degree or higher for all four services except Auburn professors, and all odds are great than 1. The increasing parameter estimates of the education with respect to education level show a monotonic effect on using the services of Auburn. Since all odds are great than 1 and monotone increasing effect, we can conclude that better educated farms are more likely to choose the services of Auburn University.

Age of the Farmers(Age)

All coefficients of Age are negative except for model using County or Regional Extension Agents. The negative coefficient of the variable Age indicates that older farmers have lower probability of using the Auburn services. However, older farmers are more likely to use County or Regional Extension Agents because its coefficient of Age is positive.

In Table 4.7, the coefficient of Age indicates that the logit of the probability of using AU Research Stations decreases 0.0015778 when the age goes up one year. Exponentiating the value of its coefficient -0.00863 we note that the odds of using AU Research Stations are multiplied by 0.991, that is, decrease 0.9% for every year of age (See Table 4.6). Note that $exp(1 + x) \approx 1 + x$ when x is small. Thus, for small x, we can use x as an approximation of the percent change in the odds associated with a unit change in the predictor. Here the coefficient of Age is 0.9% and the approximation is 0.863%. It shows a negative effect but it is small in magnitude. We can have the similar interpretations for parameter estimates, odds ratio and marginal effects for the other four cases to the case with Aures. But we can see that all numbers are near 0 no matter they are negative or positive, which means age is not a significant factor to the odds ratio or probability of using the AU related service.

Years of Experience of the Farmers(Farming Experience)

The coefficients of Years of experience of farmers are so close to 0. All of the odds ratios are close to 1 and a little greater than 1. All have positive average marginal effects(AMEs) but all numbers are less than 0.005. We can conclude that Years of Experience of the Farmers(Farming Experience) has little effect on use of Auburn services.

Years of the Current Farming Operation of the Farmers(Started)

We have similar results for Years of the Current Farming Operation of the Farmers(Started) to those for Years of Experience of the Farmers(Farming Experience). The coefficients of Years of experience of farmers are so close to 0. All of the odds ratios are close to 1 and a little greater than 1. All have positive average marginal effects(AMEs) but all numbers are less than 0.007. We can conclude that Years of Experience of the Farmers(Farming Experience) has little effect on use of Auburn services.

Total Gross Value of Sales of the Farmers(Grossvalue)

The number of observations with Total Gross Value of Sales of the Farmers less than $5,000 (Grossvalue=1) is the largest and I choose it as the reference category. If Grossvalue=10, which means that the respondent did not choose any answers and we put all of these farmers in one group. Here I consider the case with binary choice of using Auburn research stations and for other four, we can follow the similar interpretations. For Grossvalue=2 and 10, the coefficients are negative, which means that these groups with Total Gross Value of Sales of the Farmers ($< 10,000$) will lead to a lower probability of using the services for Aures. For Grossvalue more than $10,000, that is Grossvalue=2,3,...,9, the coefficients are positive, which means that these groups with Total Gross Value of Sales of the Farmers will lead to a higher probability of using the services for Aures compared to the group with Total Gross Value of Sales less than $5,000. We can see in average marginal effects(AMEs) the average change of predicted probabilities from the corresponding group to the reference group with Total Gross Value of Sales less than $5,000.

The Value of the Total Assets of the Farmers(Begass)

The number of observations with Total Assets of farmers between $100,000 and $249,999 (Begass=6) is the largest and I choose it as the reference category. We consider the case of binary choice of using service of Auburn research stations. For the odds, I found that odds with Total Assets of farmers less than $5,000(1 vs 6), Total Assets of farmers between $10,000 and $24,999 (3 vs 6) and Total Assets of farmers between $50,000 and $99,999(5 vs 6) are greater than 1, which means a positive effect on the odds when I move from the reference group to the current group. The coefficients of other groups with Total Assets of farmers are negative, which means that these Total Assets of farmers will lead to a lower probability of using service of Auburn research stations. Refer to the average marginal effects (AMEs) for the change of predicted probabilities of using this service from the reference group to the current group. The other four cases of binary choice of using Auburn services can follow by a similar way.

4.4 Goodness of Fit Statistics

In this section, I investigate the goodness of fit for logistic regression. Among the most common measure of Goodness of Fit are percent correctly predicted and -2 Log L statistics.

I introduce the percent correctly predicted first. To obtain the percent correctly predicted, we need compute the estimated probability for each observation. For instance, our model is estimated by $P(Y=1|X) = exp(\hat{\beta}_0 + X\hat{\beta})/[1 + exp(\hat{\beta}_0 + X\hat{\beta})]$. If the probability of occurrence of the predicted event is λ or higher, we predict that the event will occur. If the $P(Y=1|X) > \lambda$, where λ is the cutoff level, we set $\hat{y} = 1$ and zero otherwise. So we can classify the sample into four groups by pairs $(y, \hat{y})$. The percent correctly predicted is given by $\frac{n(y=\hat{y})}{n(S)}$, where $n(y = \hat{y})$ is the number of observations such that $y = \hat{y}$ and $n(S)$ is the total number of available observations in the sample. Sometimes, we need be careful about the percent correctly predicted for some particular case, for example the number of observations with y=1 is much bigger than that with y=0 or vice versa. For this reason, we need report the percent correctly predicted for each of the two groups, event group and nonevent group. More precisely, we can define the four pairs of groups by $(y, \hat{y}) = (1, 1)$, $(y, \hat{y}) = (1, 0)$, $(y, \hat{y}) = (0, 0)$ and $(y, \hat{y}) = (0, 1)$. The $P((y, \hat{y}) = (1, 1))$, that is, the percentage of events correctly predicted is known as the Sensitivity. The $P((y, \hat{y}) = (0, 0))$, that is, the percentage of nonevents correctly predicted, known as Specificity. Focusing on errors in prediction, the False Positive rate is given by the $P((y, \hat{y}) = (0, 1))$, the percentage of predicted events which are incorrect, or the False Negative rate, $P((y, \hat{y}) = (1, 0))$, the percentage of predicted nonevents which are incorrect. Lower λ values will be associated with greater sensitivity and fewer false negatives, but less specificity and more false positives. Higher λ values will be associated with greater specificity and fewer false positives, but lower sensitivity and more false negatives. Some programs just use 0.5 as the λ, but you may pick any value you want.

I have the prediction rate table for five models by SAS programming (See Table 4.4). Five models have the same independent variables with different response variable. Model1-5 are corresponding to response variable, Aures, Extagent, Auprof, Auweb and Agriclin respectively. Column E means "Event", whose response variable is equal to 1 and NE means "Non-Event", whose response variable is equal to 0. ES means the correctly predicted in group Events and EF means incorrectly predicted in group Events. NES means the correctly predicted in group Non-Events and NEF means incorrectly predicted in group Non-Events. ST means the correctly predicted in total. For example, the table shows the number of farmers who used AU Research Stations is 119 and the number of farmers who did not use AU Research Stations is 242. The number that the model correctly predict in group Event is 80 out of 119 and fail 39 out of 119. The number that the model correctly predict in group Non-Event is 24 out of 242 and fail 218 out of 242 Prediction successes is 28.81%([80+24]/361).

The -2 LOG L statistic measures how well the model predicts the probability of the event. The model with the smallest -2 LOG L is considered the best, although the -2 LOG L value itself is not meaningful. For example, we are given this statistic for model that contains only the intercept (-2 LOG L = 457.693, df = 1), Three variables (Grossvalue, Age, Educ), Four variables (Grossvalue, Farming Experience, Age, Educ), Five Variables (Started, Grossvalue, Farming Experience, Age, Educ), and full Model (with intercept and all 7 predictors, -2 LOG L = 393.900, df = 8). The difference between these values of -2 LOG L indicates that Model I with the 7 predictors is the best model among the five options. I have the -2 LOG L statistic table for five models by SAS programming (See Table 4.11). For the other models, I have the similar results, that is, the full models are the best models.

Table 4.6: Odds Ratio Table

	Aures	Extagent	Auprof	Auweb	Agriclin
Started	1.010	1.000	1.051	1.026	1.008
Total Assets					
Less than 5000	1.296	0.477	0.777	0.891	0.663
[5K,10K)	0.805	1.337	0.517	0.566	0.839
[10K,25K)	1.551	1.528	0.621	1.707	1.163
[25K,50K)	0.868	0.732	0.526	0.535	0.299
[50K,100K)	1.572	1.133	1.592	1.737	0.589
[100K,250K)(Reference)	1.000	1.000	1.000	1.000	1.000
[250K,500K)	0.695	0.702	0.841	0.960	0.601
[500K,1M)	0.435	0.569	0.445	0.928	1.012
More than 1M	0.168	0.090	1.293	0.945	<0.001
Total gross value of 2009 sales					
Less than 5K (Reference)	1.000	1.000	1.000	1.000	1.000
[5K,10K)	0.735	1.155	1.076	0.525	0.407
[10K,25K)	2.241	2.113	1.825	0.848	0.782
[25K,50K)	3.056	6.553	3.104	1.184	0.788
[50K,100K)	1.899	3.376	4.329	0.888	0.429
[100K,250K)	7.716	3.264	5.691	2.002	0.664
[250K,500K)	20.484	27.999	1.990	9.355	<0.001
More than 1M	11.529	>1,000	14.212	1.937	<0.001
No record	0.749	0.603	0.799	0.367	0.316
Farming Experience	1.024	1.006	1.004	1.013	1.022
Age	0.991	1.004	0.983	0.971	0.967
Education					
Some high school or less(Reference)	1.000	1.000	1.000	1.000	1.000
Graduated high school	3.321	1.577	1.648	3.505	1.359
Some college or technical school	5.480	1.580	2.237	4.126	1.272
College graduate	5.626	2.097	3.795	5.712	1.305
Some graduate school	18.796	>1,000	4.590	10.306	8.369
Masters degree or higher	11.967	5.147	5.206	10.211	2.337
Return of Investments					
Decreased	0.398	0.396	0.390	0.471	0.743
Unchanged	0.337	0.306	0.188	0.360	0.821
Increased(Reference)	1.000	1.000	1.000	1.000	1.000

Table 4.7: Marginal Effects for AU Research Stations Use

Variable	Coefficient	Standard error	AMEs
Constant	-1.664	1.065	
Started	0.010	0.016	0.002
Total Assets			
[100K,250K)(Reference)			
Less than 5,000	0.260	0.421	0.047
[5K,10K)	-0.216	0.527	-0.040
[10K,25K)	0.439	0.483	0.080
[25K,50K)	-0.142	0.474	-0.026
[50K,100K)	0.452	0.489	0.083
[250K,500K)	-0.364	0.516	-0.067
[500K,1M)	-0.834	0.662	-0.152
more than 1M	-1.784*	0.918	-0.326
Total gross value of 2009 sales			
Less than 5K (Reference)			
[5K,10K)	-0.308	0.416	-0.056
[10K,25K)	0.806*	0.486	0.148
[25K,50K)	1.117*	0.646	0.204
[50K,100K)	0.641	0.689	0.117
[100K,250K)	2.043***	0.612	0.374
[250K,500K)	3.020**	1.234	0.552
More than 1M	2.445*	1.313	0.447
No record	-0.290	0.410	-0.053
Farming Experience	0.024**	0.010	0.004
Age	-0.009	0.011	-0.002
Education			
Some high school or less(Reference)			
Graduated high school	1.200	0.756	0.219
Some college or technical school	1.701**	0.748	0.311
College graduate	1.727**	0.773	0.316
Some graduate school	2.934***	0.983	0.536
Masters degree or higher	2.482***	0.792	0.454
Return of Investments			
Increase in ROI(Reference)			
Decrease in ROI	-0.921*	0.482	-0.168
Unchange in ROI	-1.088**	0.497	-0.199

Table 4.8: Marginal Effects for County or Regional Extension Agents Use

Variable	Coefficient	Standard error	AMEs
Constant	0.248	0.930	
Started	0.000	0.015	4.952E-6
Total Assets			
[100K,250K)(Reference)			
Less than 5,000	-0.740*	0.385	-0.152
[5K,10K)	0.291	0.479	0.060
[10K,25K)	0.424	0.449	0.087
[25K,50K)	-0.313	0.451	-0.064
[50K,100K)	0.125	0.462	0.026
[250K,500K)	-0.354	0.488	-0.073
[500K,1M)	- 0.564	0.606	-0.116
more than 1M	- 2.406**	0.960	-0.494
Total gross value of 2009 sales			
Less than 5K (Reference)			
[5K,10K)	0.144	0.360	0.030
[10K,25K)	0.748	0.511	0.154
[25K,50K)	1.880**	0.844	0.386
[50K,100K)	1.217*	0.664	0.250
[100K,250K)	1.183**	0.589	0.243
[250K,500K)	3.332**	1.475	0.684
More than 1M	15.694	1,144	4.266
No record	- 0.505	0.378	-0.104
Farming Experience	0.006	0.009	0.001
Age	0.004	0.011	0.001
Education			
Some high school or less(Reference)			
Graduated high school	0.455	0.531	0.094
Some college or technical school	0.458	0.517	0.094
College graduate	0.741	0.549	0.152
Some graduate school	15.8208	538.6	4.632
Masters degree or higher	1.638***	0.600	0.337
Return of Investments			
Increase in ROI(Reference)			
Decrease in ROI	-0.926*	0.545	-0.190
Unchange in ROI	-1.185**	0.553	-0.243

Table 4.9: Marginal Effects for AU Professors Use

Constant	-1.026	1.146	
Started	0.050***	0.018	0.006
Total Assets			
[100K,250K)(Reference)			
Less than 5,000	-0.253	0.522	-0.033
[5K,10K)	- 0.660	0.627	-0.086
[10K,25K)	- 0.476	0.653	-0.062
[25K,50K)	- 0.350	0.464	-0.084
[50K,100K)	0.465	0.548	0.061
[250K,500K)	-0.173	0.546	-0.023
[500K,1M)	-0.811	0.705	-0.106
More than 1M	0.257	0.845	0.033
Total gross value of 2009 sales			
Less than 5K (Reference)			
[5K,10K)	0.074	0.489	0.010
[10K,25K)	0.602	0.570	0.078
[25K,50K)	1.133*	0.681	0.148
[50K,100K)	1.465**	0.710	0.191
[100K,250K)	1.739***	0.612	0.227
[250K,500K)	0.688	1.091	0.090
More than 1M	2.654**	1.330	0.346
No record	-0.225	0.518	-0.029
Farming Experience	0.004	0.011	0.001
Age	- 0.018	0.013	-0.002
Education			
Some high school or less(Reference)			
Graduated high school	0.500	0.794	0.065
Some college or technical school	0.805	0.777	0.105
College graduate	1.334*	0.798	0.174
Some graduate school	1.524	1.056	0.199
Masters degree or higher	1.650**	0.818	0.215
Return of Investments			
Increase in ROI(Reference)			
Increase in ROI	-0.941*	0.485	-0.123
Unchange in ROI	-1.672***	0.528	-0.218

Table 4.10: Marginal Effects for AU Website Use

Variable	Coefficient	Standard error	AMEs
Constant	-0.452	1.111	
Started	0.026	0.016	0.004
Total Assets			
[100K,250K)(Reference)			
Less than 5,000	-0.115	0.441	-0.020
[5K,10K)	-0.570	0.554	-0.098
[10K,25K)	0.535	0.483	0.092
[25K,50K)	- 0.625	0.516	-0.107
[50K,100K)	0.552	0.485	0.095
[250K,500K)	-0.041	0.503	-0.007
[500K,1M)	-0.075	0.623	-0.013
more than 1M	- 0.056	0.836	-0.010
Total gross value of 2009 sales			
Less than 5K (Reference)			
[5K,10K)	- 0.645	0.435	-0.111
[10K,25K)	-0.164	0.522	-0.028
[25K,50K)	0.169	0.640	0.029
[50K,100K)	-0.119	0.739	-0.020
[100K,250K)	0.694	0.555	0.119
[250K,500K)	2.236*	1.288	0.385
More than 1M	0.661	1.352	0.114
No record	-1.003**	0.473	-0.172
Farming Experience	0.013	0.010	0.002
Age	-0.029**	0.012	-0.005
Education			
Some high school or less(Reference)			
Graduated high school	1.254	0.822	0.216
Some college or technical school	1.417*	0.814	0.244
College graduate	1.743**	0.831	0.300
Some graduate school	2.333**	1.033	0.401
Masters degree or higher	2.324***	0.854	0.400
Return of Investments			
Increase in ROI(Reference)			
Increase in ROI	-0.753	0.466	-0.130
Unchange in ROI	-1.022**	0.483	-0.176

Table 4.11: Marginal Effects for Climate Forecasts Use

Variable	Coefficient	Standard error	AMEs
Constant	-0.011	1.254	
Started	0.009	0.021	0.001
Total Assets			
[100K,250K)(Reference)			
Less than 5,000	-0.411	0.543	-0.043
[5K,10K)	-0.176	0.620	-0.018
[10K,25K)	0.151	0.573	0.016
[25K,50K)	-1.207*	0.703	-0.125
[50K,100K)	-0.529	0.652	-0.055
[250K,500K)	-0.510	0.680	-0.053
[500K,1M)	0.012	0.804	0.001
more than 1M	-13.231	461.9	-1.432
Total gross value of 2009 sales			
Less than 5K (Reference)			
[5K,10K)	-0.899	0.602	-0.093
[10K,25K)	-0.247	0.657	-0.026
[25K,50K)	-0.238	0.760	-0.025
[50K,100K)	-0.847	1.114	-0.088
[100K,250K)	- 0.409	0.729	-0.042
[250K,500K)	-12.714	631.9	-1.071
More than 1M	-13.450	1093.4	-0.952
No record	-1.151*	0.650	-0.119
Farming Experience	0.021*	0.012	0.002
Age	-0.0341**	0.015	-0.004
Education			
Some high school or less(Reference)			
Graduated high school	0.307	0.847	0.032
Some college or technical school	0.241	0.835	0.025
College graduate	0.266	0.866	0.028
Some graduate school	2.125*	1.089	0.221
Masters degree or higher	0.849	0.897	0.088
Return of Investments			
Increase in ROI(Reference)			
Increase in ROI	-0.297	0.590	-0.031
Unchange in ROI	-0.197	0.602	-0.020

Table 4.12: Predicted Probabilities Table

Model	E	NE	ST(%)	ES	EF	NES	NEF
Aures	119	242	71.29	39	80	218	24
				(32.77)	(67.23)	(90.08)	(9.92)
Extagent	205	156	65.37	153	52	83	73
				(74.63)	(25.37)	(53.21)	(46.79)
Auprof	73	288	77.12	21	52	279	9
				(28.77)	(71.23)	(96.88)	(3.13)
Auweb	103	258	83.10	26	77	242	16
				(25.24)	(74.76)	(93.80)	(6.20)
Agriclin	47	314	87.26	3	44	312	2
				(6.38)	(93.62)	(99.36)	(0.64)

Table 4.13: -2LOG L Statistic Table

	Aures	Extagent	Auprof	Auweb	Agriclin
Intercept only	457.693	493.781	363.498	431.689	279.237
Three variables	415.636	447.225	325.173	391.622	262.511
Four variables	408.075	446.452	324.037	388.623	259.204
Five variables	407.651	446.270	318.254	387.241	259.114
Seven variables	393.900	425.781	300.259	374.787	250.507

Chapter 5

Impact of Auburn Related Services on Return of Investments

In this chapter, I investigate impact of Auburn related services on return of investments. To this end, I perform a sequence of analyses. I employ ordinal logit regression. I compute odd ratios and marginal effects to measure effects of each independent variable and investigate goodness of fit for logistic regression.

I consider an ordinal logit model. The dependent variable Return of Investments Y takes values 1 (decreased), 2 (unchanged) and 3 (increased) and let p1 = P(Y=1) be the probability of decrease in Return of Investments, p2 = P(Y=2) be the probability of unchange in Return of Investments and p3 = P(Y=3) be the probability of increase in Return of Investments. Our logit model is of the following form

$$Log[P1/(1-P1)] = \beta_{01} + \beta X,$$

and

$$Log[(P1+P2)/P3] = \beta_{02} + \beta X,$$

where X=(Started, Farming Experience, Grossvalue, Begass, Age, Educ, Service Variable). Note that the intercepts are different, but the remaining regression parameters are the same for the above equations. In this chapter, I fit five models with service variable by Aures, Extagent, Auprof, Auweb or Agriclin respectively.

5.1 Odds Ratios and Marginal Effects

Under the Proportional Odds Assumption, the odds P1/(1-P1) and (P1+P2)/P3 are proportional, and we have

$$[P1/(1-P1)] = exp(\beta_{01} + \beta X),$$

and

$$[(P1+P2)/P3] = exp(\beta_{02} + \beta X).$$

We can obtain the model parameters by maximum likelihood estimation. We have

$$P1 = exp(\beta_{01} + \beta X)/(1 + exp(\beta_{01} + \beta X)),$$

$$P1 + P2 = exp(\beta_{02} + \beta X)/(1 + exp(\beta_{02} + \beta X)),$$

and

$$P3 = 1 - (P1 + P2).$$

Assume that x_i is changed to $x_i + 1$ and the others hold constant. Then the coefficient β_i for x_i is the difference in the log odds. In other words, for a one-unit increase in x_i, the expected change in odds is $e^{\beta_i} - 1$ times. If parameter β_i is positive, then p1, predicted probability of decrease in return of investments (Y=1), as well as cumulative probability of decrease or unchange in return of investments, p1+p2, are higher for higher values of x_i. If parameter β_i is negative, p1 and p1+p2 are lower for higher values of x_i. For example, for the Farming Experience variable, we have $e^{-0.003} = 0.997$. For a one year increase in experience on farmer's investment, the expected change of odds is 0.7%.

I obtained the information about the marginal effects to our five models in Table 5.2-5.6.

5.2 Impact Factors on Return of Investments

According the parameter estimate, the coefficients of the model are directly related to the probability of the odds of farmers who used the services and those who did not. The coefficients in the regression results represent the marginal effect of a change in factors on the log-odds. The effect on the odds ratio is obtained simply by exponentiating the parameter estimate. On the other hand, odds ratios show the preferences. For example, odds ratio of some graduate school versus some high school or less (Education 5 vs 1) for five models are 0.375, 0.340, 0.336, 0.344 or 0.352. Compared to those farmers with some high school or less, the better educated farmers are less likely to obtain a decreased return of investments and more likely to obtain an increased return of investments.

Intercept 1 and Intercept 2 are the estimated ordinal logits for the adjacent levels of the response variable, decreased return of investments versus unchanged or increased, and decreased or unchanged return of investments versus increased, respectively, when the explanatory variables are evaluated at zero. Standard interpretation of an ordinal logit coefficients is that for a one unit increase in the predictor, the response variable level is expected to change by its respective regression coefficient in the ordinal logit scale while holding the other variables in the model constant. A positive coefficient indicates that the corresponding variable will increase the odds of $[P1/(1-P1)]$ and $[(P1+P2)/P3]$ and then also increase the probability of decrease in return of investments. I consider seven factors in the following.

Choices of farmers on the Auburn Related Services (Service Variable)

The table indicates that the Auburn related services are important factors on return of investments. All coefficients of Service Variable are negative except Auprof, which means that farmers who use these Auburn related services have a lower probability of decrease in return of investments. For example of using or not Auburn research stations, the difference of probability of decrease in return of investments(P1) is -0.033 (AMEs1). Note that P1+P2=1-P3, and AMEs2 is the change of probability of decrease or unchange in return of investments(P1+P2). AMEs2 =- 0.019 means the difference of probability of increase in return of investments(P3) is 0.019. In other words, the farms in User group are less likely(3.271% less in probability) to have decrease in return of investments and more likely (1.914% more in probability) to have increase in return of investments.

Education of farmers(Educ)

The table indicates that the education of farmers is also an important factor for return of investments. All coefficients of Educ are negative and all odds are less than 1 for all five services, which means that it will lead to a lower probability of decrease in return of investments for better educated farmers. We can conclude that better educated farmers are less likely to have decrease in return of investments and more likely to have increase in return of investments. Refer to the numbers of average marginal effects (AMEs1 and AMEs2) for the change of probabilities for decrease in return of investments and increase in return of investments.

Age of farmers(Age)

The positive coefficient of the variable Age and the average marginal effects indicate that older farmers tends to have lower probability of increase in return of investments but the change is very small in magnitude for every year of age increase since its coefficient is near 0 and all average marginal effects AMEs1 and AMEs2 are less than 0.001.

Years of experience of farmers(Farming Experience)

All coefficients of Years of experience of farmers are negative, which means that Years of experience will lead to a higher probability of increase in return of investments. Also the change is very small in magnitude for every year of age increase since its coefficient is near 0 and all average marginal effects AMEs1 and AMEs2 are less than 0.001.

Years of the current farming operation of farmers(Started)

All coefficients of Years of experience in current farming operation are positive, which means that it will lead to a lower probability of increase in return of investments. Also the change is very small in magnitude for every year of age increase since its coefficient is near 0 and all average marginal effects AMEs1 and AMEs2 are less than 0.000.

Total Gross Value of Sales of Farmers(Grossvalue)

All coefficients of total gross value of Sales of the Farmers between \$5,000-\$24,999 (Grossvalue=2 and 3) and their average marginal effects are negative, which means that, compared with the total gross value of farmers less than \$5,000, these farmers with total gross Value of Sales will have a higher probability of increase in return of investment and a lower probability of decrease in return of investment. However, the total gross value of sales of farmers more than \$25,000 (Grossvalue=4-9) will lead to a higher probability of decrease in return of investments and a lower probability of increase in return of investment.

The Value of Total Assets of Farmers(Begass)

The results for Begass can by followed by arguments similar to those for Grossvalue. All coefficients of the value of total assets of the Farmers between \$50,000 and \$99,999 (Begass=5) are negative, which means that, compared with the value of the total assets of farmers between \$100,000 and \$249,999, these values of total assets of Farmers will lead to a higher probability of increase in return of investment. However, the value of total assets of farmers less than \$49,999 or more than \$100,000(Begass=1-4,7-9) will lead to a higher probability of decreased in return of investments and a lower probability of increase in return of investment.

5.3 Goodness of Fit Statistics

By SAS, I have the following -2 LOG L statistic table. We are given this statistic for model that contains only the intercept (-2 LOG L = 649.454, df = 1), Three variables(Farming Experience, educ and Service Variable), Four variables(Farming Experience, Age, educ and Service Variable), Five Variables(Started, Farming Experience, Age, educ and Service Variable), and full Model (with intercept and all 7 predictors, -2 LOG L = 613.085, df = 8). The difference between these values of -2 LOG L indicates that Model I with the seven predictors is the best model among the five options. We have the -2LOG L statistic table for five models by SAS programming(See Table 5.7). For the other five models, I have the similar results, that is, the full models are the best among these models.

Table 5.1: Odds Ratio Table

	Aures	Extagent	Auprof	Auweb	Agriclin
Started	1.014	1.014	1.014	1.014	1.014
Total Assets					
Less than 5000	2.043	2.025	2.028	2.024	2.008
[5K,10K)	1.566	1.586	1.589	1.579	1.580
[10K,25K)	3.609	3.546	3.544	3.552	3.549
[25K,50K)	1.533	1.544	1.548	1.538	1.517
[50K,100K)	0.986	0.966	0.964	0.969	0.960
[100K,250K)(Reference)	1.000	1.000	1.000	1.000	1.000
[250K,500K)	1.713	1.742	1.743	1.740	1.731
[500K,1M)	1.258	1.309	1.315	1.309	1.307
More than 1M	3.386	3.612	3.608	3.619	3.541
Total gross value of 2009 sales					
Less than 5K (Reference)	1.000	1.000	1.000	1.000	1.000
[5K,10K)	0.927	0.939	0.939	0.936	0.927
[10K,25K)	0.522	0.502	0.500	0.502	0.499
[25K,50K)	1.277	1.232	1.221	1.232	1.218
[50K,100K)	1.116	1.082	1.072	1.081	1.066
[100K,250K)	1.928	1.778	1.765	1.783	1.760
[250K,500K)	1.820	1.638	1.633	1.664	1.597
More than 1M	>1,000	>1,000	>1,000	>1,000	>1,000
No record	2.385	2.412	2.417	2.398	2.376
Farming Experience	0.997	0.996	0.996	0.996	0.996
Age	1.003	1.004	1.004	1.004	1.003
Education					
Some high school or less(Reference)	1.000	1.000	1.000	1.000	1.000
Graduated high school	0.424	0.413	0.412	0.414	0.414
Some college or technical school	0.619	0.592	0.591	0.596	0.594
College graduate	0.765	0.730	0.727	0.736	0.731
Some graduate school	0.375	0.340	0.336	0.344	0.352
Masters degree or higher	0.539	0.499	0.495	0.503	0.503
Service variable	0.822	0.991	1.024	0.962	0.873

Table 5.2: Marginal Effects for ROI on AU Research Stations Use

Variable	Coefficient	Standard error	AMEs1	AMEs2
Intercept(1)	-0.132	0.757		
Intercept(2)	2.335***	0.772		
Started	0.014	0.015	0.002	0.001
Total Assets				
[100K,250K)(Reference)				
Less than 5,000	0.714**	0.354	0.096	0.056
[5K,10K)	0.448	0.434	0.055	0.032
[10K,25K)	1.284***	0.437	0.184	0.108
[25K,50K)	0.427	0.393	0.064	0.038
[50K,100K)	-0.014	0.415	-0.008	-0.005
[250K,500K)	0.538	0.422	0.070	0.041
[500K,1M)	0.229	0.530	0.036	0.021
more than 1M	1.220	0.791	0.162	0.095
Total gross value of 2009 sales				
Less than 5K (Reference)				
[5K,10K)	-0.075	0.336	-0.016	-0.009
[10K,25K)	-0.651	0.435	-0.101	-0.059
[25K,50K)	0.245	0.555	0.028	0.016
[50K,100K)	0.110	0.577	0.011	0.006
[100K,250K)	0.657	0.516	0.080	0.047
[250K,500K)	0.599	1.048	0.102	0.060
More than 1M	14.015	651.6	1.063	0.622
No record	0.869**	0.365	0.115	0.067
Farming Experience	-0.003	0.009	-0.001	-0.000
Age	0.003	0.009	0.001	0.000
Education				
Some high school or less(Reference)				
Graduated high school	-0.857	0.527	-0.126	-0.074
Some college or technical school	-0.480	0.520	-0.074	-0.043
College graduate	-0.268	0.548	-0.046	-0.027
Some graduate school	-0.982	0.774	-0.141	-0.082
Masters degree or higher	-0.618	0.571	-0.097	-0.057
Aures	-0.196	0.244	-0.033	-0.019

Table 5.3: Marginal Effects for ROI on County or Regional Extension Agents Use

Variable	Coefficient	Standard error	AMEs1	AMEs2
Intercept(1)	-0.134	0.762		
Intercept(2)	2.328***	0.777		
Started	0.014	0.015	0.002	0.001
Total Assets				
[100K,250K)(Reference)				
Less than 5,000	0.705**	0.355	0.093	0.055
[5K,10K)	0.461	0.434	0.057	0.034
[10K,25K)	1.266***	0.437	0.182	0.107
[25K,50K)	0.434	0.393	0.065	0.038
[50K,100K)	-0.034	0.414	-0.011	-0.006
[250K,500K)	0.555	0.422	0.072	0.042
[500K,1M)	0.270	0.529	0.041	0.024
more than 1M	1.284	0.792	0.171	0.100
Total gross value of 2009 sales				
Less than 5K (Reference)				
[5K,10K)	-0.063	0.335	-0.014	-0.008
[10K,25K)	-0.689	0.434	-0.106	-0.062
[25K,50K)	0.208	0.557	0.024	0.014
[50K,100K)	0.079	0.579	0.009	0.005
[100K,250K)	0.576	0.510	0.070	0.041
[250K,500K)	0.493	1.046	0.088	0.0518
More than 1M	13.887	657.9	1.066	0.626
No record	0.880**	0.365	0.117	0.117
Farming Experience	-0.004	0.009	-0.001	-0.000
Age	0.004	0.010	0.001	0.000
Education				
Some high school or less(Reference)				
Graduated high school	-0.885	0.528	-0.129	-0.076
Some college or technical school	-0.523	0.518	-0.081	-0.048
College graduate	-0.315	0.547	-0.053	-0.031
Some graduate school	-1.080	0.777	-0.154	-0.090
Masters degree or higher	-0.696	0.570	-0.108	-0.063
Extagent	-0.009	0.231	-0.008	-0.005

Table 5.4: Marginal Effects for ROI on AU Professors Use

Variable	Coefficient	Standard error	AMEs1	AMEs2
Intercept(1)	-0.139	0.759		
Intercept(2)	2.322***	0.773		
Started	0.014	0.015	0.002	0.001
Total Assets				
[100K,250K)(Reference)				
Less than 5,000	0.707**	0.353	0.094	0.056
[5K,10K)	0.463	0.435	0.055	0.032
[10K,25K)	1.265***	0.436	0.181	0.106
[25K,50K)	0.437	0.394	0.064	0.037
[50K,100K)	-0.037	0.414	-0.010	-0.006
[250K,500K)	0.556	0.422	0.072	0.042
[500K,1M)	0.274	0.531	0.038	0.023
more than 1M	1.283	0.788	0.174	0.102
Total gross value of 2009 sales				
Less than 5K (Reference)				
[5K,10K)	-0.063	0.335	-0.015	-0.009
[10K,25K)	-0.693	0.434	-0.106	-0.062
[25K,50K)	0.200	0.557	0.024	0.014
[50K,100K)	0.069	0.580	0.011	0.006
[100K,250K)	0.568	0.515	0.073	0.043
[250K,500K)	0.491	1.041	0.085	0.050
More than 1M	13.864	658.3	1.201	0.704
No record	0.8825**	0.3648	0.117	0.069
Farming Experience	-0.004	0.009	-0.001	-0.000
Age	0.004	0.009	0.001	0.000
Education				
Some high school or less(Reference)				
Graduated high school	-0.887*	0.527	-0.130	-0.076
Some college or technical school	-0.526	0.518	-0.081	-0.047
College graduate	-0.319	0.547	-0.052	-0.030
Some graduate school	-1.092	0.767	-0.154	-0.091
Masters degree or higher	-0.704	0.567	-0.107	-0.063
Auprof	0.024	0.286	-0.017	-0.010

Table 5.5: Marginal Effects for ROI on AU Website Use

Variable	Coefficient	Standard error	AMEs1	AMEs2
Intercept(1)	-0.126	0.761		
Intercept(2)	2.336***	0.776		
Started	0.014	0.015	0.002	0.001
Total Assets				
[100K,250K)(Reference)				
Less than 5,000	0.705**	0.353	0.094	0.055
[5K,10K)	0.457	0.435	0.055	0.032
[10K,25K)	1.268***	0.436	0.182	0.107
[25K,50K)	0.430	0.394	0.064	0.037
[50K,100K)	-0.032	0.415	-0.009	-0.005
[250K,500K)	0.554	0.422	0.072	0.042
[500K,1M)	0.269	0.528	0.041	0.024
more than 1M	1.286	0.787	0.173	0.101
Total gross value of 2009 sales				
Less than 5K (Reference)				
[5K,10K)	-0.067	0.336	-0.016	-0.009
[10K,25K)	-0.688	0.432	-0.107	-0.063
[25K,50K)	0.209	0.553	0.021	0.013
[50K,100K)	0.078	0.576	0.007	0.004
[100K,250K)	0.578	0.509	0.070	0.041
[250K,500K)	0.509	1.048	0.093	0.055
More than 1M	13.888	657.1	0.957	0.561
No record	0.875**	0.366	0.116	0.068
Farming Experience	-0.004	0.009	-0.001	-0.000
Age	0.004	0.010	0.001	0.000
Education				
Some high school or less(Reference)				
Graduated high school	-0.882	0.528	-0.128	-0.075
Some college or technical school	-0.518	0.519	-0.080	-0.047
College graduate	-0.307	0.548	-0.051	-0.030
Some graduate school	-1.067	0.770	-0.152	-0.089
Masters degree or higher	-0.687	0.570	-0.105	-0.062
Auweb	-0.039	0.251	-0.015	-0.010

Table 5.6: Marginal Effects for ROI on Climate Forecasts Use

Variable	Coefficient	Standard error	AMEs1	AMEs2
Intercept(1)	-0.092	0.764		
Intercept(2)	2.372***	0.780		
Started	0.014	0.015	0.002	0.001
Total Assets				
[100K,250K)(Reference)				
Less than 5,000	0.697**	0.354	0.093	0.055
[5K,10K)	0.457	0.434	0.056	0.033
[10K,25K)	1.267***	0.437	0.182	0.107
[25K,50K)	0.417	0.395	0.063	0.037
[50K,100K)	-0.041	0.415	-0.012	-0.007
[250K,500K)	0.549	0.423	0.071	0.042
[500K,1M)	0.267	0.528	0.041	0.024
more than 1M	1.264	0.790	0.171	0.100
Total gross value of 2009 sales				
Less than 5K (Reference)				
[5K,10K)	-0.076	0.337	-0.016	-0.010
[10K,25K)	-0.695	0.432	-0.108	-0.063
[25K,50K)	0.197	0.552	0.021	0.012
[50K,100K)	0.064	0.577	0.006	0.004
[100K,250K)	0.565	0.508	0.070	0.039
[250K,500K)	0.468	1.041	0.081	0.048
More than 1M	658.1	0.000	-0.906	0.532
No record	0.866**	0.366	0.116	0.068
Farming Experience	-0.004	0.009	-0.001	-0.000
Age	0.003	0.010	0.001	0.000
Education				
Some high school or less(Reference)				
Graduated high school	-0.881*	0.527	-0.130	-0.076
Some college or technical school	-0.521	0.517	-0.082	-0.048
College graduate	-0.314	0.545	-0.054	-0.032
Some graduate school	-1.045	0.771	-0.153	-0.090
Masters degree or higher	-0.686	0.564	-0.109	-0.064
Agriclin	-0.136	0.322	-0.018	-0.011

Table 5.7: -2 LOG L Table

	Aures	Extagent	Auprof	Auweb	Agriclin
Intercept only	649.454	649.454	649.454	649.454	649.454
Three variables	644.018	644.372	644.318	644.379	643.622
Four variables	643.765	644.060	643.956	644.082	643.390
Five variables	642.786	643.123	643.131	643.146	642.427
Seven variables	613.085	613.703	613.698	613.682	613.527

Chapter 6

Impact Analysis of Auburn Related Services on Performance of Farmers in Alabama

We are interested in the effects of service variables on performance of farmers in Alabama. For example, is the total gross value of sales for those who used certain Auburn Related Services more than that for those Non-users? Another concern is that we want to better predict the performance change of the Alabama farmers those used the services of AU.

The original dataset of Grossvalue is a categorical data. I take values of Grossvalue at the mean and use as a continuous variable (MGrossvalue). MGrossvalue will take values as the following:

2,500 if "1" less than 5,000;

7,500 if "2" 5,000-9,999;

17,500 if "3" 10,000-24,999;

37,500 if "4" 25,000-49,999;

75,000 if "5" 50,000-99,999;

175,000 if "6" 100,000-249,999;

375,000 if "7" 250,000-499,999;

750,000 if "8" 500,000-999,999;

1,000,000 if "9" 1,000,000 or more;

0 if "10" None.

In this chapter, I view MGrossvalue as the response or dependent variable and Started*Farming Experience, Squared Started, Cubic Started, Squared Farming Experience, Cubic Farming Experience, Roa, Begass, Age, Educ, Sex, Race, Marital Status and Service Variable as predictors (See Tables 2.1-2.4 for the descriptions), which are chosen by VIF analysis and Backward selection. I go through a sequence of analyses. Firstly, I use the variance inflation factor (VIF) to test for multicollinearity. Secondly, I use backward selection method to choose variables. Thirdly, I analyze the residual to check the assumptions for the linear regression. Fourthly, I fit the model with the refined data. Finally, I include dummy variables in the model to analyze the impact.

6.1 Use the VIF to Test for Multicollinearity

In the presence of multicollinearity, the estimate of one variable's impact on y tends to be less precise. In some sense, the collinear variables contain the same information about the dependent variable. When severe multicollinearity occurs, the standard errors for the coefficients tend to be very large, and then the estimated regression coefficients can be highly unreliable. To get better regression models, I use the variance inflation factor (VIF) to test for multicollinearity. Refer to [28] for the VIF test. The idea is to calculate the VIF factor for $i-th$ regression variable with the following formula:

$$tolerence = 1 - R_i^2$$

and

$$\text{VIF} = \frac{1}{tolerence}$$

where R_i^2 is the coefficient of determination of the regression equation and to analyze the magnitude of multicollinearity by considering the size of the VIF. A common rule of thumb is that if VIF is larger than 5 then multicollinearity is high. In some literature (See [24]), also VIF of 10(or a tolerance of 0.1 or less) has been proposed as a cut off value.

Notice that the R^2 of Roa is 0.036. Therefore, the tolerance is $1 - 0.036 = 0.964$. The VIF is $1/0.964 = 1.037$. Now we have seen what tolerance and VIF measure and we have been convinced that there is a light collinearity problem with other variable for Roa.

According to Table 6.1, since VIFs for all are close to 1, the multicollinearity problem is not a cause for concern.

Table 6.1: Tolerance and VIF

Variable	DF	R-Square	Tolerance	Variance Inflation
Constant	1			
Started	1	0.065	0.936	1.070
Farming Experience	1	0.046	0.954	1.049
Age	1	0.043	0.957	1.045
Percent changed of ROI	1	0.036	0.964	1.037
Sex	1	0.018	0.982	1.019
Marital Status	1	0.030	0.970	1.031

6.2 Use Model-Selection Methods to Select Variables

In this section, our goal is to relate the characters of farmers to performance via a simple linear regression, but the problem is that the different groups of farmers may cause the relationships to differ. To find a linear model that best predicts the dependent variable from the independent variables, refer to Model-Selection Methods by Hocking (1976) [16] and Judge et al. (1985)[20] for details. Among these methods, we have Full Model Fitted (NONE), Forward Selection (FORWARD), Backward Elimination (BACKWARD), Stepwise (STEPWISE), Maximum R Improvement (MAXR), Minimum R (MINR) Improvement, R Selection (RSQUARE), and so on. I choose the backward selection method to select variables from Started, Roachange, Farming Experience, Begass, Age, Educ, Started*Farming Experience (StEx), Squared Started (SqSt), Squared Farming Experience (SqEx), Cubic Started (CbSt), Cubic Farming Experience (CbEx) and the Service Variable, where Service Variable is dummy variable in Table 2.2 respectively corresponding to five models. I get the following results. For example, for this linear model 1, dummy is corresponding to Aures, and R-Square = 0.285080 which means 28.51% of the variation in MGrossvalue can be explained by this model.

The R-square and adjusted R-square figures indicate that approximately 27-28% of variance in total gross value of sales is explained by age, education, years of experience in farming and use of the services of AU. We are interested in the impact of the use of services of AU. The d values for dummy indicate the direction and number of units of change in the dependent variable due to a one unit change in each dummy variable.

6.3 Estimation of Parameters and Tests of Hypotheses

If the sample can be random assign to two groups, the control and treatment groups, we may apply simple dummy analysis. The dummy variable will be 1 if it belongs to treatment group, otherwise 0. For example, our data can be divide into two groups, User group and Non-user group for certain AU related service (Aures or the other). Our model is the following, $y = \beta_0 + \beta X + dD$, where X=(Begass, Age, Educ, Roa, Race, Sex, Marital, StEx, SqSt, CbSt, Service Variable D). d shows the impact when D=1 where D is a service variable.

Next I would like to test $H_0 : d = 0$, $H_a : d \neq 0$. For example, I test the hypotheses for the case with Aures. The test statistic is $t_0 = 33,375/11,485 = 2.91$. Our obtained, or calculated t value is 2.91. I see that for alpha = 0.05, my calculated value is smaller than the critical value at alpha =0.05, so I reject the null hypothesis and fail to reject the alternative hypothesis, namely, that $d \neq 0$. The results for all five cases are shown in Table 6.5.

6.4 Conclusions

The results show that, a change in Aures from 0 to 1 will result in a change of the average gross value of sales $33,375. This means that the gross value of sales of farmers in User group is $33,375 greater than that in Non-user group. Similarly, we can see the gross value of sales of farmers in User group is greater than that in Non-user group for each of the other three models except for Agriclin.

For the race, the number of observations of group with White or Caucasian (Race =4) is the largest and I choose it as the reference category. Compared with White or Caucasian, American Indian or Alaska Native earn $200,000 and over for all five models. For the other unspecified races, farmers earn $2,871 to $12,872 less than farmers of White or Caucasian for five models. Female will earn $5,935 to $10,093 less than male for five modes. Married will earn $3,230 to $6,151 more than single or divorced or widowed. As percent changed on return of investments (Roa) increases one unit, Mgrossvalue will increase $105 to $248 for five models. Those who have the value of total assets more than $500,000 will have relatively higher Mgrossvalue than those who have the value of total assets between $100,000 and $249,999. Age has a negative effect for the Mgrossvalue. Every year of age increase will result in $931 to $1,138 less.

Table 6.2: Parameter Estimates for Aures, Extagent and Auprof

Variable	AU res.	Extent Agents	AU prof.
Constant	104,831***	102,501***	102,965***
Total Assets			
[100K,250K)(Reference)			
Less than 5,000	-2,306	4,458	1,078
[5K,10K)	2,927	1,109	5,018
[10K,25K)	-14,276	-14,318	-9,883
[25K,50K)	-7,643	-7,122	-4,830
[50K,100K)	-16,733	-15,271	-16,007
[250K,500K)	33,462	33,604	33,300
[500K,1M)	123,386***	123,893***	125,347***
more than 1M	178,335***	184,091***	171,654***
Age	-931.3 **	-1013.0**	-936.3**
Education			
Some high school or less(Reference)			
Graduated high school	-54,253**	-52,192**	-49,042**
Some college or technical school	-64,037***	-58,830**	-57,882**
College graduate	-67,436***	-64,015**	-64,013**
Some graduate school	-62,659*	-61,426	-54,190
Masters degree or higher	-65,168**	-61,048**	-57,489**
Percent changed of ROI	204.1	143.8	104.7
Race			
White or Caucasian(Reference)			
Black or African American	8,268	2,741	7,301
Asian or Pacific Islander	-15,687	11,932	4,211
Other	-4,963	-6,602	-2,871
American Indian or Alaska Native	207,051***	207,115***	210,124***
Spanish,Hispanic or Latino Origin	-4,268	-2,462	11,436
Sex	-5,935	-10,993	-9,101
Marital	6,151	3,230	5,404
Interaction			
Stex	20.7	25.5	29.0
Sqst	89.9	93.2	85.7
Cbst	-1.3	-1.4	-1.4
Service variable D	33,375***	27,186**	35,649***

Table 6.3: Parameter Estimates for AU website and Agriclin

Variable	AU website	Agriclin
Constant	108,063***	121,028***
Total Assets		
[100K,250K)(Reference)		
Less than 5,000	681.9	-559.1
[5K,10K)	4,273	2,345
[10K,25K)	-13,631	-12,281.1
[25K,50K)	-5,855	-9,614
[50K,100K)	-16,079	-14,306
[250K,500K)	33,710	33,097
[500K,1M)	124,975***	126,686,***
more than 1M	175,110***	177,319,***
Age	-960.0**	-1138.0**
Education		
Some high school or less(Reference)		
Graduated high school	-52,772**	-50,210**
Some college or technical school	-60,582**	-57,266**
College graduate	-65,565**	-61,111**
Some graduate school	-53,559	-42,075
Masters degree or higher	-59,062**	-51,497*
Percent changed of ROI	177.1	247.9
Race		
White or Caucasian(Reference)		
Black or African American	4,321	5,564
Asian or Pacific Islander	1,278	-5,653
Other	-12,872.0	-7,524
American Indian or Alaska Native	216,186***	222,938***
Spanish,Hispanic or Latino Origin	3,623	7,051
Sex (1:Female;0:Male)	-9,313	-10,033
Marital	5,167	5,278
Interaction		
Stex	30.4	30.4
Sqst	92.7	93.1
Cbst	-1.4	-1.4
Service variable D	20,473*	-15,294

Table 6.4: Model Diagnostics

Variable	Aures	Extagent	Auprof	Auweb	Agriclin
R-square	0.285	0.280	0.282	0.273	0.269
F	5.12	4.99	5.04	4.83	4.73
Pr>F	<.0001	<.0001	<.0001	<.0001	<.0001

Table 6.5: T-Test for D

	Aures	Extagent	Auprof	Auweb	Agriclin
T0	2.91	2.44	2.61	1.68	-0.96
P-value	0.004	0.015	0.010	0.094	0.337
Test results	reject	reject	reject	fail to reject	fail to reject

Chapter 7

Summary, Conclusions and Discussions

Agricultural extension services, provided by universities, play critical roles in assisting farmers. This thesis studies the impact of Agricultural extension services provided by Auburn University on Alabama farmers. However, the data showed only small percent (around one third) of farmers who responded to the survey use services offered by Auburn University. The number of the sample is 389 out of a population of 1,750 of farmers who started any part of their operation since 2005. According to United States Department of Agriculture (USDA) Hired Farm Labor - Workers and Payroll: 2007, there are 9,541 farms with 30,932 workers in Alabama.

Even though the number of sample points in our data is only 389, it is sufficient to employ standard statistical techniques. Important conclusions can be drawn from the results. For instance, it has been apparent that as the average of age of farmers increases and younger farmers would be more likely to seek help from the Auburn Research Extensions and Auburn Professors, visit Auburn websites and agroclimate.org. The statistical data indicates that the percent of farmers who used services from Auburn University with age 50 or less is higher than those the age 51 or more.

Education plays an important role in extension efforts, such as seeking help from Auburn professors. It was found that the farmers with some graduate experience (master or above degree) are more likely turn to Auburn University professors for help. As they acquire more education, they would have greater understanding of new technologies. These younger better educated farmers would also rely more on the information provided by the auburn websites or agroclimate.org.

I applied the two sample Hoteling T-squared method to test whether there is some difference between two means of the groups of User and Non-user of AU services. The results confirmed that there did exist a difference. To investigate the impact factors on the choices of Auburn related services, I employed Binary Logit Model. It was found that years since the year current farming operation was established, the value of total assets, the total gross value of 2009 sales, years of experience in farming, age, education and return of investments are important factors affecting the choices to use or not Auburn related services of farmers in Alabama. The year current farming operation was established, the value of the total assets, the total gross value of 2009 sales, years of experience in farming, age, education and AU related service are also important factors affecting the return of investment. As expected, the use of the AU related services for Aures, Extagent, Auweb and agriclin are associated with higher probability of obtaining an increase in return of investments.

I apply statistical regression involving dummy variables to study the relation between the Auburn University related services and the performance of farmers in Alabama. The results indicate that farmers who used Auburn University services are more likely to perform better than those who did not. The results in Chapter 6 by statistical regression also confirm that the age of the farm, the value of total assets, the total gross value of 2009 sales, years of experience in farming, the age of the farmer and their education, Percent changed in return of investments, race, sex, marital status and AU related service are important factors affecting the performance of farmers in Alabama.

However, the limitations of a cross-sectional study using the current data are apparent. For the impact of use the Auburn related services, if we have a panel data, for example, during several years from time before the use to after 5 years, then we can investigate the difference of performance before and after the use. Another reason to use panel data is that the technology use needs some time to affect the performance of farmers. Due to the cost of survey, unfortunately I did not have such data. Another drawback of this model is the cost of Auburn related services was not used.

Bibliography

[1] Aldrich, J. H. and Nelson,F. D. (1984), Linear Probability, Logit, and Probit Models, Beverly Hills, CA:Sage.

[2] Araji, A.A., R.J. Sim, and R.I. Gardner. (1978). Returns to agricultural research and extension: An ex ante approach. American Journal of Agricultural Economics 60:164-168.

[3] Anderson, J.R. (2007), Agricultural Advisory Services, Background Paper for the World Development Report 2008.

[4] Anderson, J.R. and Feder, G. (2003), Rural Extension Services, World Bank Policy Research Working Paper 2976, World Bank, Washington D.C.

[5] Anderson, J. R., Feder, G. and Ganguly, S. (2006), The Rise and Fall of Training and Visit Extension: An Asian Mini-drama with and African Epilogue, World Bank Policy Research Working Paper 3928, World Bank, Washington D.C.

[6] Birkhaeuser, D., Evenson, R. E., Feder, G., (1991). The economic impact of agricultural extension: a review. Econ. Dev. Cult. Change 39, 610-650.

[7] Bohner, S.A. and R.S. Arnold, Eds. (1996). Software Change Impact Analysis. Los Alamitos, California, USA, IEEE Computer Society Press.

[8] Chavas J, Aliber M (1993) An analysis of economic efficiency in agriculture: a nonparametric approach. J Agric Resour Econ 18:1-16

[9] Dinar, A., Karagiannis, G. and Tzouvelekas, V. (2007) Evaluating the impact of agricultural extension on farms? performance in Crete: a nonneutral stochastic frontier approach. Agricultural Economics, 36 (2):135-146.

[10] Ekboir, J., (2003).Why impact analysis should not be used for research evaluation and what the alternatives are. Agricultural Systems 78 (2), 166-184.

[11] Evenson, Robert, and Mwabu Germano, (1998), The Effects of Agricultural Extension on Farm Yields in Kenya. Economic Growth Center Discussion Paper No. 798.

[12] Feder, G., R. Just, et al. (1985). Adoption of Agricultural Innovations in Developing Countries: A Survey.Economic Development and Cultural Change 33(2): 255-298.

[13] Feder, Gershon, and Roger Slade (1986), The Impact of Agricultural Extension: The Training and Visit System in India, The World Bank Research Observer, 1(2): 139-161.

[14] Godtland E., Sadoulet, E., De Janvry, A., Murgai, R. and Ortiz, O.(2004), The Impact of Farmer Field Schools on Knowledge and Productivity: A Study of Potato Farmers in the Peruvian Andes, Economic Development and Cultural Change, 53(1): 63-92.

[15] Greene, W. H. (1997), Econometric Analysis, Third edition, Prentice Hall, 339?C350.

[16] Hocking, R. R. (1976). The analysis and selection of variables in linear regression. Biometrics, 32, 1-49.

[17] H.James, L. Kenneth and E. Timothy, Evaluating the benefit of NGO Agricultural Assistance: the case of season extenders in Bosnia-Herzegovina, preprint.

[18] Jamison, D. and Moock, P. (1984). Farmer education and farm efficiency in Nepal: The role of schooling, extension services and cognitive skills. World Development, 12.

[19] Johnson, N. L., N. Lilja, and J. A. Ashby. 2003. Measuring the impact of user participation in agricultural and natural resource management research. Agricultural Systems 78:287-306.

[20] Judge, G. G., Griffiths, W. E., Hill, R. C., L*ü*tkepohl, H., Lee, T. C., (1985) The Theory and Practice of Econometrics, Second Edition, New York: Wiley & Sons.

[21] Lockheed, M. E., D. T. Jamison, and L. J. Lau. 1980. Farmer education and farm efficiency: A survey. Economic Development and Cultural Change 29 (1): 37-76.

[22] Lohr L., Park T.A., (2003). Improving extension effectiveness for organic clients: Current status and future directions. Journal of Agricultural and Resource Economics, 28 (3): 634- 650.

[23] Long, S. J. (1997) Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: SAGE Publications, Inc.

[24] Kutner, Nachtsheim, Neter, Applied Linear Regression Models, 4th edition, McGraw-Hill Irwin, 2004.

[25] Owens, T., Hoddinott, J. and Kinsey, B. (2003) The impact of agricultural extension on farm production in resettlement areas of Zimbabwe. Economic Development and Cultural Change, 51, pp. 337-58.

[26] Pedro, C., Maffioli, A. and Ubfal, D. (2008). The Impact of Agricultural Extension Services: The Case of Grape Production in Argentina.Working Paper prepared by the Inter-American Development Bank.

[27] Ransom, J.K., K. Paudyal, and K. Adhikari. (2003). Adoption of Improved Maize Varieties in the Hills of Nepal.Agricultural Economics 29: 299-305. Reilly, J., Schimmelpfennig, D., (1999). Agricultural impact assessment, vulnerability, and the scope for adaptation.Clim. Change 43, 745-788.

[28] O'Brien, Robert M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors," Quality and Quantity 41(5)673-690.

[29] SAS Institute Inc., SAS/STAT User's Guide, Version 9.1. Cary, NC, SAS Institute Inc, 2002-2003

[30] Thirtle, C., L. Lin, and J. Piesse. (2003). The Impact of Research Led Agriculture Productivity Growth on Poverty Reduction in Africa, Asia and Latin America.World Development 31 (12): 1959-75.

[31] Valentin, L., Bernardo,D. J. and Kastens, T. L..(2004).Testing the Empirical Relationship between Best Management Practice Adoption and Farm Profitability.Review of Agricultural Economics,Vol 26, No 4, pp 489-504.

[32] Wu, L., Praneetvatakul, S., Waibel, H. and Wang, L. (2005), The Impact of FFS on Yield, Pesticide Cost and Gross Margin in Shandong Province, P.R. China: an Econometric Approach, in Ooi, P.A.C., Praneetvatakul, S., Waibel, H. and Walter-Echols, G. (eds), The Impact of the FAO-EU IPM Programme for Cotton in Asia, Pesticide Policy Project Publication Series, Special Issue No. 9, University of Hannover, Germany.