2010

Off-farm labor supply by farm operators and spouses: a comparison of estimation methods

Mahesh Pandit
Louisiana State University and Agricultural and Mechanical College

OFF-FARM LABOR SUPPLY BY FARM OPERATORS AND SPOUSES: A COMPARISION OF ESTIMATION METHODS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Agricultural Economics and Agribusiness

By
Mahesh Pandit
B.S. Tribhuvan University, 2004
M.S., Tribhuvan University, 2006
August 2010

To my parents

&

Wife

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my major professor Dr. Ashok K. Mishra for his invaluable guidance to complete this thesis. He always inspired and challenged me to write a thesis that contributes to and furthers the knowledge in this field. My gratitude also extended to the committee members Drs. Krishna P. Paudel, Luis A. Escobar and Joshua D. Detre for their encouragement and kindly help.

My deepest gratitude goes to my parents (Tek Nidihi Pandit and Ambika Devi Pandit), and my wife Saraswati Pandit Paudel. My wife has shown great strength and support for this achievement. Without her love and devotion, my study could not have been accomplished. Back in Nepal, my appreciation goes to all my brothers, sisters, sisters in laws, nephews and nieces.

I especially appreciate the contributions of Drs. Murali Adhikari and Laxmi Paudel who helped and encouraged me throughout and prior to my studies at Louisiana State University. My appreciation is extended to the families of Sheshkanta Buba, Mohan Dai, Pradip Dai, Dipak Dai, and to Suniti Bhauju.

I would also like to thank Brian M. Hilbun, a colleague, and classmate Patrick Hatzenbuehler. In addition, a special thanks goes to the Nepalese community at LSU for their help and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis studies the off-farm labor supply decision of farm operators and their spouses in the United States. The data used in this study is from the Agricultural Resource Management survey, 2006. The objective of this study is twofold. First, to identify those factors that affect off-farm labor supply of farm operators and their spouses. In particular, this study investigates the impact of human capital of farm operators and spouses, personal, family, farm and location characteristics on labor allocation for on- and off-farm work. Empirical results indicate that farm operators' and their spouses' human capital are positively correlated with off-farm labor supply. In addition, the number of children in a household is inversely related to a spouse's off-farm employment. Similarly, a household's net worth and farm size have a negative impact on off-farm labor allocation decisions by both farm operators and their spouses. Payments from government programs have a negative effect on labor allocation for non-farm work. The availability of health insurance to farm operators and their spouses from off-farm employment has a positive effect on labor supply for off-farm work.

The second objective of this study is to compare results obtained from a parametric probit model and a semiparametric additive probit model of off-farm labor supply by farm operators and spouses. One of the most important aspects of semiparametric analysis is to identify smoothing or nonparametric variables in a regression model. The Blundell and Duncan (1998) approach shows that farm size is such a smoothing variable in the off-farm labor supply model. A semiparametric additive regression model identifies a few significant covariates as compared to a parametric probit model; however, the Hong and White (1995) specification test and likelihood ratio test favor a semiparametric model in this study. In particular, the graphical plots of fitted values from parametric and semiparametric models also show that a semiparametric

model is preferred. The semiparametric model helps to formulate appropriate functional form of off-farm labor supply in the United States, which might be the subject of further study of this research.

CHAPTER 1

INTRODUCTION

According to United States Department of Agriculture (USDA, 2005), agricultural production commands an overall smaller share in both the national and rural economies of the United States. Agricultural industries contribute 2.5 percent of the U.S.'s gross domestic product (GDP). A report written by Dimitri, Effland and Conklin (2005) shows that American agriculture has undergone a tremendous transformation in the 20th century in terms of the structure of farms and farm households including off-farm work. Decision makers, mainly farm operators and their spouses, are providers of agricultural labor. Table 1 shows the number of farms, total farm acreage, gross farm income and net farm income in the United States from 1950 to 2007. Over that time, the number of farms has decreased by almost 80 percent over the past six decades. Similarly, the data shows a decrease in the overall amount of acreage in farms. For example, the total acreage of land devoted to farming in 2007 was 930 million acres, a nearly 23 percent decrease in acreage since 1950. The number of farm operators decreased by 68 percent over the same period.

Some of the main factors that have caused these transformations are low price and income elasticities for agricultural products and technological advancement, which, suggests that the supply of agricultural products has outpaced the growth in demand for these products. The increase in the relative wage rate in non-farm sectors is also responsible for the reallocation of farm labor in non-farm sectors. As a result, labor demand in the farm sector was declined. These changes resulted in a decrease in the number of farm operators and an increase in the total number of hours worked off-farm by farm operators. According to the 2007 Census of Agriculture, more than 65 percent of farm operators worked off the farm.

Table 1. Selected Statistics of Farms in the United States, 1950-2008

Year	Farms	Land In Farms (000)	Net Farm Income ($000)	Gross Farm Income ($000)
1950	5,647,800	1,202,019	13,634	33,089
1951	5,427,600	1,203,500	15,909	38,257
1952	5,197,500	1,204,930	14,945	37,735
1953	4,983,600	1,205,740	12,961	34,429
1954	4,798,200	1,206,355	12,365	34,174
1955	4,653,800	1,201,900	11,277	33,448
1956	4,514,100	1,197,070	11,242	33,948
1957	4,371,700	1,191,340	11,063	34,765
1958	4,232,900	1,184,944	13,179	38,969
1959	4,104,520	1,182,563	10,707	37,885
1960	3,962,520	1,175,646	11,212	38,588
1961	3,825,410	1,167,699	11,957	40,547
1962	3,692,410	1,159,383	12,064	42,343
1963	3,572,200	1,151,572	11,770	43,369
1964	3,456,690	1,146,106	10,492	42,304
1965	3,356,170	1,139,597	12,899	46,549
1966	3,257,040	1,131,844	13,960	50,468
1967	3,161,730	1,123,456	12,339	50,520
1968	3,070,860	1,115,231	12,322	51,847
1969	3,000,180	1,107,811	14,293	56,408
1970	2,949,140	1,102,371	14,366	58,818
1971	2,902,310	1,096,863	15,012	62,118
1972	2,859,880	1,092,065	19,455	71,143
1973	2,823,260	1,087,923	34,356	98,910
1974	2,795,460	1,084,433	27,267	98,247
1975	2,521,420	1,059,420	25,510	100,554
1976	2,497,270	1,054,075	20,175	102,917
1977	2,455,830	1,047,785	19,881	108,766

Table 1. Contd.

Year	Farms	Land In Farms (000)	Net Farm Income ($000)	Gross Farm Income ($000)
1978	2,436,250	1,044,790	25,198	128,447
1979	2,437,300	1,042,015	27,415	150,720
1980	2,439,510	1,038,885	16,141	149,280
1981	2,439,920	1,034,190	26,879	166,323
1982	2,406,550	1,027,795	23,841	164,146
1983	2,378,620	1,023,425	14,262	153,855
1984	2,333,810	1,017,803	25,975	167,981
1985	2,292,530	1,012,073	28,509	161,069
1986	2,249,820	1,005,333	31,117	156,125
1987	2,212,960	998,923	38,005	168,408
1988	2,200,940	994,423	39,633	177,916
1989	2,174,520	990,723	46,487	191,603
1990	2,145,820	986,850	46,261	197,795
1991	2,116,760	981,736	40,234	192,049
1992	2,107,840	978,503	50,171	200,555
1993	2,201,590	968,845	46,725	205,022
1994	2,197,690	965,935	52,567	216,107
1995	2,196,400	962,515	39,771	210,833
1996	2,190,500	958,675	58,936	235,807
1997	2,190,510	956,010	51,306	238,029
1998	2,192,330	952,080	47,119	232,592
1999	2,187,280	948,460	47,690	234,934
2000	2,166,780	945,080	50,569	243,622
2001	2,148,630	942,070	54,874	251,779
2002	2,135,360	940,300	39,594	232,649
2003	2,126,860	936,750	60,462	260,047
2004	2,112,970	932,260	85,801	295,638
2005	2,098,690	927,940	79,335	301,105

Table 1. Contd.

Year	Farms	Land In Farms (000)	Net Farm Income ($000)	Gross Farm Income ($000)
2006	2,088,790	925,790	58,509	292,380
2007	2,204,950	921,460	86,778	341,148
2008	2,200,000	919,900	89,320	379,905

Source: Economic Research Services, U.S. Department of Agriculture (1950-2008) and Mishra (1996)

Furthermore, almost forty percent of farm operators worked off-farm 200 days or more in 2007. Hence, off-farm employment is an integral part of the agricultural transformation, and is one of the most remarkable changes in agricultural production. The income earned from off-farm wages and salaries accounted for nearly 80% of total farm household income. (Mishra et al., 2002). This income is important for the economic well-being of these households. Off-farm income appears to smooth household income (Mishra and Goodwin, 1997; Mishra and Sandretto, 2002) and most farmers view off-farm work as permanent rather than just a temporary or transitional pursuit (Ahearn, Perry, El-Osta, 1993). In addition, growth of off-farm income over the past four decades has significantly reduced income inequality between both farm and non-farm households (Gardner, 2002).

The above discussion shows that, like non-farm industries, many farm households (operator and spouse) are dual career households with respect to income sources (farm and off-farm). The alternative source of off-farm income (or off-farm hours) varies by farm size, region, farm characteristics, and the human capital of operators and spouses. According to Fernandez-Cornejo (2007), off-farm income varies inversely with farm size; i.e. operators of smaller farms have higher off-farm income than large farms. This is due to greater off-farm employment for operators of small farms (Mishra and Goodwin, 1997).

Government farm programs affect the labor allocation decision by farm operators and their spouses. Although several farm policies have been introduced since 1933, the Federal Agriculture Improvement and Reform Act of 1996 replaced the traditional price support program with a system of direct payments based on historical production with one of complete planting flexibility. The Farm Security and Rural Investment Act of 2002 introduced a counter-cyclical payments program that paid farmers on historical production for times when market prices fall below a pre-specified target level. For example, peanut producers are qualified to receive counter-cyclical payments when market price falls below the target price of $495 per ton during 1998-2001. Whether government policies influence the time allocation decisions of farm operators is an important issue in the study of farm household behavior.

From an econometric estimation approach, most of the research in off-farm labor supply functions has focused primarily on parametric functional forms. An alternative approach is semiparametric functional form. A semiparametric model is more flexible in estimation than the nonparametric model, because the sample size required to obtain reliable estimates is not as large the sample size required for a parametric model and it captures the non-linearity of covariates. Goodwin and Holt (2002) use semiparametric model to study off-farm labor supply of agrarian households in transition in Bulgaria.

1.1. Objectives

The main objective of this study is to estimate the off-farm labor supply model of farm operators and their spouses using a semiparametric methodology. The study presented here also discusses the advantages and disadvantages of both traditional and semiparametric estimation procedures. This study find parameter estimates using probit estimation and generalized additive semiparametric model. Finally, a Hong and White test is used to compare results of the parametric and the semiparametric models. Thus, the researchers, who are interested in the study of functional form of off-farm labor supply, can be benefited to apply appropriate functional form.

CHAPTER 2

LITERATURE REVIEW

Numerous studies of agriculture and farm households have shown that changes have occurred in the structure of U.S. agriculture. For example, Gebremdhin and Christy (1996) illustrated the transformation of the structure of production agriculture using a descriptive analysis. They focused on the implications of structural changes on resource usage, population distribution in rural communities, and the survival of small farms. Gebremdhin and Christy (1996) argue that this structural change in agriculture is not a temporary phenomenon. One of the structural changes is that farm operators report more off-farm work. The term 'off-farm work' is described in many different ways by different economists, for example Fuguitt (1959) described off-farm work as a "push-pull" hypothesis and concluded that off-farm employment is directly related to off-farm opportunities and inversely related to opportunities in farming.

Previous studies were done using utility maximization subject to budget constraints (Bollman, 1979). Huffman (1980) studied off-farm labor supply in Iowa, North Carolina, and Oklahoma using household utility maximization as its objective, subject to constraints on time, income, and farm production. Using these three factors (a vector of time endowments of family members, allocated between farm work, off-farm work, and leisure; household income received by members through off-farm work, and a vector of farm production), Huffman (1980) concluded that off-farm labor supply is directly related to operators' education and extension inputs. His result also implied that farmers who attain more education tend to reallocate their labor services from self-employed farm work to off-farm work faster than less educated farmers.

A study of farm operators in Dodge Country, Georgia (Barlett, 1986), claimed that the primary reason farmers choose to work off-farm is due to the variability associated with farm income. However, Barlett (1986) did not use any empirical methodology to reach this conclusion. Supporting this view, Mishra and Goodwin (1997) found that the off-farm labor supply of farmers was positively correlated with the variability in their farm income. Mishra and Sandretto (2002) found that even though real net farm income has been declining over the past fifty years, income variability has declined over that same time period on either the aggregate or per-farm basis. This evidence suggests that the declining mean income and its high variability are a major shift in labor decisions of farm households.

Perhaps off-farm income reduces the impacts of these problems. For example, Mishra and Sandretto (2002) show that off-farm income has contributed significantly to raising the level of and reducing the variability in farm household incomes. Furthermore, Mishra and El-Osta (2001) agreed with previous studies in concluding that the farm income component is the primary source of variability in total farm household income for those farms participating in federal commodity programs, while the major source of income variability for nonparticipating households is income from off-farm sources. Empirical results obtained by Mishra and Holthausen (2002) confirms that, as expected, variability in farm income and off-farm incomes have a positive and negative effect, respectively, on off-farm hours worked by farm operators. Kwon, Orazem, and Otto, (2006) found that Iowa farm couples adjust their off-farm labor supply in response to both permanent and transitory farm income shocks.

Human capital plays an important role in off-farm labor supply (Bollman 1979). Investment in education increases the human capital of individuals. Korb (1999) indicated that younger, better-educated farmers and their spouses are most likely to work off-farm. Although, Furtan, Van Kooten and Thompson (1985), and Huffman and Lange (1989) found education elasticities of wages for both men and women to be inelastic. Huffman (1980) found a positive elasticity and a significant direct effect of education on the number of days worked off-farm. In addition, education influence off-farm labor supply (e.g. Gould and Saupe, 1989; Huffman and El-Osta, 1997; Thompson, 1985; Goodwin and Holt, 2002). Gould and Saupe (1989) concluded that vocational training increased the probability of off-farm work. Yang (1997) states that knowledge sharing within a household is a cause of educational selectivity of off-farm family labor supply. El-Osta, Mishra and Morehart (2008) found that educational attainment of the spouse to be positively related with the decision of the wife to work off-farm.

Experience (both on and off-farm) is another factor affecting labor supply. Past research shows that experience working off-farm has a positive effect on off-farm labor supply. For example, Sumner (1982) found that the off-farm work experience is positive and inelastic. Lass, Findeis and Hellberg (1991) find similar result. Alternatively, some economists find that farming experience affects off-farm labor supply directly through the labor supply function and indirectly through farm production. Furtan, Van Kooten and Thompson (1985) estimated the direct impact of farming experience on the off-farm labor supply and found it to be negative and statistically significant. Mishra and Goodwin (1997) concluded that farmers with more farming experience were less likely to work off the farm. They also found that the off-farm labor supply of farmers and their spouses was positively correlated with off-farm experience. Goodwin and Holt (2002) also support previous findings that previous off-farm experience farm operators were more likely to work off the farm.

Family and location characteristics such as the number of children in the household and household net worth play a vital role in affecting off-farm labor supply. Since most farm operator's spouses are female, who were traditionally responsible for taking care of the household's children. It is likely that the number of children in a household determines whether an operator's wife works on- or off-farm. Furtan, Van Kooten and Thompson (1985) concluded that the total number of children has a negative effect on the number of off-farm work hours supplied by farm wives. Thompson (1985) separated children into two age groups; 6-12 years and 12 years or more. She found that the first group reduces the number of hours worked off the farm by spouses. Gould and Saupe (1989) found that number of children under six years old had a significant negative effect on the probability of the spouse's off-farm work. Further, they concluded that marginal value of the spouse's home time due to a child's birth increased households work decreasing off-farm work. Mishra and Goodwin (1997), El-Osta, Mishra, and Morehart, (2008), and Goodwin and Holt (2002) draw similar conclusions from their research on off farm work by the female spouse of the farm operator.

The effect that the number of children have on the off labor supply of male farmers is not conclusive. Huffman (1980) found that the number of children who were less than five years of age significantly increased the off-farm labor supply of farm operators. Furtan, Van Kooten, and Thompson (1985) concluded that for each additional child, the hours worked off-farm by a farmer increased by 11%. From these findings, one can conclude that more children perhaps require the farmer to seek off-farm income, so farm operators are more likely to work off-farm. Furthermore, using a family farm survey conducted in Israel, Kimihi (2004) concluded that both fathers and mothers tended to reduce their participation in off-farm work as the number of children rose. However, Kimihi (1994) also found that farm couples are more likely to work off the farm when the number of other adults in the household increased, because adults might be able to help some farm and household work.

Government farm programs started in 1933 as part of the Agricultural Adjustment Act (AAA). The first farm bill proposed to establish a 'New Deal' mix of commodity-specific price and income support programs. In 1985, the U.S. government enacted the Food Security Act (1985) that introduced marketing load provisions to commodity loan programs by allowing the repayment of loans at lower rates when market prices fell. This was done with the intention of both aiding U.S. farmers and in reducing government-held surplus.

Mishra and Goodwin (1997), El-Osta and Ahearn (1996) found that off-farm work by farm operators and their spouses were significantly correlated with receipts of government payments. The U.S. government introduced the Federal Agricultural Improvement and Reform Act (FAIR) in 1996 known as 'freedom to farm', will stipulated that government payments are not tied to production. The FAIR Act pertains to use fixed "Production Flexibility Contract" (PFC) or "Agricultural Market Transition Act (AMTA) payments and described as decoupled payments (El-Osta, Mishra and Ahearn, 2004; Ahearn, El-Osta and Dewbre, 2006). Decoupled payments are not related to a farmer's current production, output levels or market conditions (e.g., price), so these payments do not have a direct effect on production decisions for specific commodities, i.e. the payments would be expected to only have a wealth effect.

The motivation for introducing AMTA/PFC payments is to minimize the trade-distorting impact of traditional coupled farm commodity payments. These payments play an important role in off-farm labor supply (Ahearn, El-Osta and Dewbre, 2006). Although Ahearn, El-Osta and Dewbre (2002) did not find any difference in the impact of coupled payments in the decision to participate in working off the farm, Dewbre and Mishra (2007) concluded that transition payments increased leisure hours for both farm operators and their spouses. El-Osta, Mishra, and Ahearn (2004) analyzed separately the effect of coupled and decoupled payments on off-farm labor supply and found that both types of government payments tended to increase the number of hours operators worked on farm and decreased the hours they work off the farm. Ahearn El-Osta and Dewbre (2006) concluded that the increase in off-farm participation of farm operators who received payments was not the result of the 1996 policy change, instead it was due to the continuation of the long-term trend of a greater reliance on off-farm work by farm households. Other payments (whether coupled or decoupled) have a negative effect on off-farm labor participation. Dewbre and Mishra (2007) also found that AMTA payments increased leisure hours for both farm operators and their spouses.

In 2008, El-Osta, Mishra, and Morehart (2008) found that total government payments are an important factor in decreasing the likelihood of off-farm work strategies involving work by a husband or a wife or both (husband and wife). In addition, they found that the marginal impact of this payment on the probability of the wife working off the farm alone is found to be positive. Hennessy and Rehman (2008), using data from Irish farmers, found that the decoupling of direct payments increases the probability of a farmer obtain the U.S. off-farm employment as well as the amount of time allocated to working off-farm. A recent study of the U.S. farmer by Key and Roberts (2009) found that farmers with decreasing marginal utility of income responded to higher decoupled payments by decreasing off-farm labor and increasing farm labor, which subsequently resulted in greater agricultural output.

The AMTA payment under the FAIR act (1996) was supposed to last for seven years, however, periods of low prices and localized yield short falls during the late 1990's led Congress to pass a supplemental payments programs for farmers in 2002. The payment, also known as "Market Loss Assistance" (MLA) payments, was based on historical base acreage; and consequently this payment classified as a 'decoupled' payment. Ad hoc market loss assistance payments started into farm legislation in the form of "Counter-Cyclical Payments" (CCP) to institutionalize Market Loss Assistance payments. Using a farm household resource allocation model and Agricultural Resource Management Survey (ARMS), Dewbre and Mishra (2007) found that loan deficiency payments (LDP) and MLA payments reduced leisure. Dewbre and Mishra (2007) also found that AMTA payments exhibited a much higher degree of income transfer efficiency than did either LDP or MLA payments. Thus, different governmental farm payments have differing effects on the allocation of labor by operators and spouses between farm and non-farm sectors.

Another factor that influences the allocation of labor supply between on and off-farm activities is farm type. Off-farm work is less likely for those farms, which are relatively labor-intensive enterprises like dairy farms (Leistritz *et al*., 1985) and other types of operations involving livestock (Lass, Findeis and Hallberg, 1991). In contrast, Futan, Van Kooten and Thompson (1985) found that the presence of livestock enterprises reduced off-farm labor supply by 41% for the United States. Their model, however, did not account for differences between dairy and beef cattle operations. In contrast, Lass and Gempeaswa (1992) concluded that farm type had little impact on off-farm labor supply when operators' and their spouses' off-farm labor supply function are jointly estimated. Off-farm hours worked, however, are significantly lower for dairy farmers when only the operator works off-farm. Kilkenny (1993) and Kimhi (1994) found that participation in off-farm labor markets differs across farm type and family structure, which support the previous results.

Farm size also has an impact on labor allocation decisions. Operators of small farms typically participate more in off-farm employment activities, work more hours off the farm, and have higher off-farm income than do operators of larger farms (Fernandez-Cornejo, Hendricks and Mishra, 2005). Mishra and Goodwin (1997) indicate that farm size affects labor allocation decisions. They found that an inverse relationship exists between farm size and off-farm work.

The availability of insurance for farm products is also another important determining factor that influences the off-farm labor market. Key, Roberts and O'Donoghoue (2006) found that greater crop insurance coverage reduces the off-farm labor supply of operators who produced at least $1,000,000 of output, and increased the labor supply of small-farm operators who produced less than $25,000 of output.

2.1. Functional Forms of Off-farm Labor Supply

With regard to functional forms of off-farm labor supply, most studies have used parametric models. For example, during the 1970s, Theil (1971), Larson and Yu Hu (1977), and Sexton (1975) used Ordinary Least Squares (OLS) to estimate the off-farm labor supply function. Sumner (1982) applied simple regression to study off-farm labor supply by farm operators. Gould and Saupe (1989) used the same procedure to study hours of off-farm work. In addition, they used a maximum likelihood probit model to study the benefit associated with off-farm labor supply. Gould and Saupe (1989) used a bivariate probit model to study entry and exit off farm labor supply. Huffman (1980) used a logit estimation procedure to find the relationship between off-farm work participation and explanatory variables. During the 1990's, economists Lass and Gempesaw (1992) used a univariate probit model with a random coefficient approach. Kimhi and Lee (1996) used simultaneous equations, with ordered categorical dependent variables, to study off-farm work decisions by farm couples. Mishra and Goodwin (1997) used a Tobit estimation procedure to find the effect of farm income variability on the supply of off-farm labor. These types of models continued to be used by researchers in following decades (Phimister and Roberts, 2006; El-Osta, Mishra and Ahearn, 2004). Another methodology used in decades that are more recent involve the use of bivariate probit models (Ahearn, El-Osta, Dewbre, 2006), and multinomial logit estimation (El-Osta, Mishra and Morehart, 2008).

Although, parametric models have been widely used in economic research, these models have some structural weakness. One weakness of parametric models is that they require strong assumptions regarding functional forms. The possibility of having nonlinear effects of covariates and this assumption does not hold true in many cases. Second, sample sizes required to estimate the coefficients accurately are large and available data are often not of sufficient size to meet this requirement. Third, parametric models are balanced by misspecification problems and a poorly specified model will leads to poor results. If semiparametric fit is more powerful than the a linear of quadratic fit, it should be used. However, if there is little difference between parametric and semiparametric fit, one can proceed with parametric fit (Keele 2008). A semiparametric estimation procedure; however would be used to correct for these weaknesses. Nonparametric components in a semiparametric distribution are distribution free, so a strong assumption such as function form is not required. Second, it can be find better estimates while using smaller sample sizes. Finally, the problem of misspecification is not commonly found when a semiparametric methodology is employed. Overall, semiparametric functional form is better compared to parametric and nonparametric model.

Semiparametric modeling was introduced following the work of Manski (1975). The partial linear model, as proposed by Robinson (1988), is the simplest semiparametric model. Although this type of model is used in many studies, there is limited use for it in labor supply related models. Goodwin and Holt (2002) used a semiparametric model to study the farm labor supply of a sample of agrarian households in Bulgaria, and found that it is consistent under various distributional assumptions. They show that labor supply is positively affected by education and work experience, which are hypothesized to increase off-farm wages. They compare parametric and semiparametric estimates using Housman test and found that semiparametric modeling procedure better than estimates from simple parametric model approach.

CHAPTER 3

THEORETICAL MODEL

3.1. Household Model

The agricultural household model provides a unifying microeconomic framework for understanding agricultural households' decisions on consumption, production, and time allocation. This chapter presents the basic structure of the model, the theoretical underpinnings and the empirical formulation of the model. The model of off-farm labor supply is based upon the model as proposed by Mishra and Goodwin (1997), and Goodwin and Holt (2002).

Consider a household that consists of two members – a head of household and that head's spouse. Assume that each household member has an option for income generation, either in agriculture (supplying F hours of labor to farming) and/or through off-farm employment (supplying M hours of labor to the off-farm market). A household member allocates their time to farm, off-farm, and leisure activities. It is assumed that perfect competition exists in the labor market and, therefore, the farm operator's labor allocation decision has no effect on aggregate demand, supply, and price of labor. Under this condition, a farm operator obtains utility from the consumption of goods, off-farm work, and on-farm work subject to budget and time constraints. If a farm operator and spouse have multiple job choices, they will compare available options and allocate their labor as to maximize total utility that implies equalizing marginal returns to labor in an alternative job and in the consumption of leisure. Let H denote household Head and S represents Spouse. Under the assumption of perfect information, I assume that the farm operator maximizes utility having leisure (L) and the consumption of good q and express it as follows:

$$U = U(q, L_H, L_S, k) \tag{3.1}$$

Subject to:

$$pq + r'X = w_H M_H + w_S M_S + \ P_f Q(X, K_H, K_S, F_H, F_S) + A \tag{3.2}$$

$$T_i = F_i + L_i + M_i, i = H, S \tag{3.3}$$

$$F_i \ \geq 0, L_i \ \geq 0, M_i \geq 0$$

where, q represents consumed good; L_i refers to leisure time $(i = H, S)$; w_i refers to the off-farm wage rate$(i = H, S)$; F_i refers to farm hours supplied; $(i = H, S)$; M_i refers to off-farm work hours; $(i = H, S)$; T_i represents total endowment of time; $(i = H, S)$; p represents the price of consumption goods; P_f represents the price of farm output; *Q(.)* is the farm output; r is a column vector of the price of other farm inputs; X represents a column vector of other input quantities; A represents non labor income; k represents household characteristics; K_i represents human capital $(i = H, S)$.

The utility function is assumed to be concave and twice continuously differentiable. The utility function varies according to operator and spouse characteristics, which is represented by k in the above equation. The utility function has two restrictions. The level of consumption is the sum of farm income, off-farm income, and exogenous non-labor income (budget constraint equation, 3.2). Using farm labor (F_i), human capital (K_i), and other farm inputs (X) as inputs for the production function *Q(.)* the farm produces output. The model assumes that the marginal utilities of leisure and consumption of goods approach infinity as consumption approaches zero, thus ensuring that positive levels of leisure and goods are always consumed (Goodwin and Holt, 2002). Equation (3.3) is the time constraint. Farm operators and spouses have fixed amount of time to allocate among leisure time, farm work, and off-farm work. In general, time spent on farm is always greater than zero $(F_i > 0)$, however, the optimal hours of off-farm work may be zero $(M_i \geq 0)$ since some farmers do not work off-farm. Under the assumption for the differentiable utility function, the optimality condition can be stated with the following Langrangian function.

$$L = U(q, L_H, L_S, k) + \lambda\big(w_H M_H + w_S M_S + P_f Q(X, K_H, K_S, F_H, F_S) + A - pq - r'X\big)$$
$$+\gamma_H(T_H - F_H - L_H - M_H) + \gamma_S(T_S - F_S - L_S - M_S) \qquad (3.4)$$

The Kuhn-Tucker first-order conditions for maximizing the utility function subject to income, time and nonnegativity constraints are:

$$\frac{\partial L}{\partial X} = p_f Q_{Xn} - r_n = 0, \quad \text{for farm inputs } n = 1, \dots N \qquad (3.5)$$

$$\frac{\partial L}{\partial M_H} = \lambda w_H - \gamma_H \leq 0,\ M_H(\lambda w_H - \gamma_H) = 0, \qquad (3.\ 6)$$

$$\frac{\partial L}{\partial M_S} = \lambda w_S - \gamma_S \leq 0,\ M_S(\lambda w_S - \gamma_S) = 0, \qquad (3.7)$$

$$\frac{\partial L}{\partial F_H} = \lambda p_f Q'_{F_H} - \gamma \leq 0, F_H\left(\lambda p_f Q'_{F_H} - \gamma_H\right) = 0, \qquad (3.8)$$

$$\frac{\partial L}{\partial F_S} = \lambda p_f Q'_{F_S} - \gamma \leq 0, F_S\left(\lambda p_f Q'_{F_S} - \gamma_S\right) = 0, \qquad (3.9)$$

$$\frac{\partial L}{\partial q} = U'_q - \lambda p = 0, \qquad (3.10)$$

$$\frac{\partial L}{\partial L_H} = U'_{L_H} - \gamma_H = 0, \qquad (3.11)$$

$$\frac{\partial L}{\partial L_S} = U'_{L_S} - \gamma_S = 0, \qquad (3.12)$$

Where λ and γ represents Langrange multipliers for the household's income and individual's time allocation. If an interior solution exists (implying that both the head and spouse work on and off-farm), equations (3.6) - (3.9) hold as equalities and imply the following known condition:

$$\frac{U'_q}{U'_{L_i}} = \frac{p}{w_i} \qquad (3.13)$$

Goodwin and Holt (2002) show that the marginal rate of substitution between leisure and consumption of good is equal to the ratio of the price of the consumption good to the wage rate.

$$P_f Q_{F_i} = w_i \tag{3.14}$$

Equation (3.14) implies that the value of the marginal product of farm labor is equal to the off-farm wage rate. According to Goodwin and Holt (2002), when an individual works on the farm but not off-farm, a corner solution with $w_i < P_f Q_{F_i}$ is implied. On the other hand, if an individual works off the farm, then a corner solution with $w_i > P_f Q_{F_i}$ is obtained. Assuming that the utility function is additively separable, marginal utilities of operators and spouses are independent. Considering the independence of the levels of productivity for both operators and their spouses, the labor supply decisions of farm operators and their spouses are separately determined.

Solving for first order conditions yields the reduced form supply equations for both farm operators and their spouses. The reduced form of the model consists of six structural endogenous variables $(M_H, M_S, F_H, F_S, q, X)$ and exogenous factors that include wages, prices, and characteristics of the production and utility function (Goodwin and Holt, 2002). The goal of this analysis is to provide estimates of off-farm labor supply decisions. A simple reduced form modeling approach that relate off-farm labor supply decisions influence by operator's and spouse's characteristics is used. Hence, an empirical off-farm labor supply model would take the form:

$$M_i = f_i(k, K_H, K_S, P_F, A) \tag{3.15}$$

Where the variables maintain their previous specification.

3.2. Parametric Model (Probit)

A farm operator's decision to work off-farm can be expressed in the framework of a discrete choice model. The response variable in this analysis is binary, indicating whether the individual (1) decides to work off-farm or 2) not to work off-farm. Let Y denote the decision of the farmers to work off-farm then the response. Y takes the value 1 if the farm operator or spouse decides to work off-farm, 0 otherwise. Generally, the decision to participate in off-farm work depends on the farm operator's level of human capital, family characteristics, farm characteristics, types of government payments received by the farm and other characteristics that are relevant to off-farm work opportunities and the cost associated with working off-farm. Let the vector $\boldsymbol{X}$ represent the information on all of these characteristics and explain off-farm labor supply. In this case, the predicted values from a regression analysis beyond the limits of 0 and 1 are meaningless. Therefore, the appropriate model for this type of analysis is a probit model and the specification is as follows

$$y = X'\beta + \varepsilon, y = 1 \text{ if } y^* > 0, \ \ 0 \text{ otherwise}, \tag{3.16}$$

$$E[\varepsilon|X] = 0 \tag{3.17}$$

$$Var[\varepsilon|X] = 1 \tag{3.18}$$

where, β represents the coefficients of covariates, $\phi(.)$ denotes the probability density function of the normal distribution, and $\Phi(.)$ represents the cumulative density function of the normal distribution. A typical probit model can be written as equation (3.19) and the parameters are estimated by the maximum likelihood estimation procedure. According to Wooldridge (2002), the probit model follows a normal distribution and takes the following forms:

$$Prob(y = 1) = \int_{-\infty}^{X'\beta} \emptyset(t)dt = \Phi(X'\beta) \tag{3.19}$$

$$Prob(y = 0) = 1 - \Phi(X'\beta) \tag{3.20}$$

Combining equations (3.19) and (3.20) to obtain the density of y_i , given X_i , is written as:

$$Prob(y|X) = [\Phi(X'\beta)]^y[1 - \Phi(X'\beta)]^{1-y} \tag{3.21}$$

Let $\{y_i, X_i\}_{i=1}^{n}$ are n independent, identically independently normally distributed observations in model (3.16). Therefore the log likelihood function of this model is:

$$\ln(l_\beta) = \sum_{i=1}^{n} y_i \ln\ \Phi(X'\beta) + (1 - y_i) \ln(1 - \Phi(X'\beta))\ . \qquad (3.22)$$

Hence, the estimation of parameter β is equivalent to maximization of the log likelihood function with respect to the β parameter.

Since this research is interested in assessing the influence of each of the independent variables on the decision of the farmer and /or spouse to work off-farm. The marginal effect of a continuous variable in a probit model is given by:

$$\frac{\partial E[y|X\}}{\partial X} = \phi(X'\beta)\beta \qquad (3.23)$$

(Green, 2008)

In addition, the marginal effect for a binary independent (dummy) variable is

$$ME = Prob[y = 1|d = 1] - Prob[y = 1|d = 0] \qquad (3.24)$$

Marginal effects can be calculated in two ways. The first method involves computing the mean of the data and using expression 3.23 in order to determine marginal effects. The second method is to find the marginal effect at each observation value and average these obtained values to get the scalar value of the marginal effect. The second procedure is more reliable than the first one, so this second procedure is used to compute the marginal effect of the independent variables.

3. 3. Semiparametric Model

One of the major concerns in statistics and economics has to do with finding the appropriate functional form. Generally, functional forms are explained by the relationship between dependent and independent variables via their probability density functions. Raw data does not contain significant information itself; therefore, data needs to be examined from different angles. The first step in analyzing data is to draw diagrams such as bar diagrams, pie charts, and scatter plots. Statistical analysis is not limited to such diagrammatic processes; different distributional, functional forms have been introduced in the history of statistical analysis; however, researchers commonly use three types (parametric, nonparametric and semiparametric) of regression procedures today.

The parametric estimation procedure involves the estimation of regression parameters, and interpretation of the estimated parameter values. The parameters are generally estimated assuming a linear relationship between the dependent and independent variables. Non-linear relationships between variables can be transformed into linear form using log function, which makes parameter computation easier. OLS and Maximum Likelihood Estimation (MLE) are two common methodologies in parametric data analysis. Sometimes, the functional forms for parameters are too rigid causing the parametric form to be inappropriate (Lee, 2001). Some of the assumptions in OLS, such as normality, autocorrelation, and multicollinearity may not hold so the estimators would no longer be considered as the 'best' linear unbiased estimator (BLUE). Consequently, OLS might sometimes be an inappropriate functional form. In addition, if the chosen density is incorrect, the MLE will not be the best (Lee, 2001).

The drawbacks of parametric regression can be overcome by removing the restriction on functional form. This approach leads to the distribution free functional form known as Nonparametric Regression. A straightforward procedure uses scatter plots to figure out the simple functional form without imposing the restrictions required in a parametric model. According to Lee (2001), this estimation procedure is "letting the data speak for themselves." Many procedures for estimating nonparametric regression are available today; however, the procedures are not free of drawbacks. One of the disadvantages of this procedure is the curse of dimensionality. Large samples need accurate measurement. In addition, if the sample size required for analysis rises with an increase in the number of explanatory variables, then the nonparametric estimation procedure can become somewhat tedious, and to address some of these problems, a new technique was introduced which combines elements of both parametric and nonparametric models into a single model. This type of estimation is known as semiparametric model. The purpose of the semiparametric model is to balance the pros and cons of the parametric and nonparametric procedures as described above. The main advantage of semiparametric estimation is the flexibility of functional form. In one sense, this methodology is a hybrid form of the parametric and nonparametric methods (Lee, 2001). Furthermore, this procedure does not restrict the parameter, so it is able to reduce the "curse of dimensionality." Moreover, the nonparametric component in the semiparametric procedure captures the nonlinearity in data. The characteristic of the included variables determines the role that either the parametric or the nonparametric component would play within a model. Although, we can select nonparametric variables using established theory, many literature have proposed a new methodology in order to test the nature of independent variables.

3.3.1. Development of Semiparametric Model

The nonparametric statistical method has been used in economic research since the 1960's, but researchers have widely used since the 1990s. Fan and Gijbels (1992) propose a nonparametric method for estimating the mean regression function, which combines the ideas of a local linear smoother and variable bandwidth. They approach a new minimization of the mean integrated square error for variable bandwidth selection. Generally, there are two types of variables: categorical and continuous. Earlier analyses are based upon just a single variable type, but in this research there are both categorical and continuous variable types i.e. rank, binary and continuous data, which have been simultaneously censored. Through implementation of local linear kernel estimation, Lewbell and Linton (2002) constructed a nonparametric censored and truncated regression model as a latent regression function, which is consistent and asymptotically and normally distributed.

Economics is a social science; so many regressors like family size, gender, etc., often come in the form of categorical variables. It is easy to perform statistical analysis if all variables are continuous; however, mixed data that consists of both categorical and continuous data are tedious to manipulate in semiparametric regression compared to parametric regression analysis. Many authors have proposed new methodologies to account for this phenomenon. For example, Racine and Li (2004) proposed a new methodology of nonparametric regression estimation for both categorical and continuous data. Using kernels along with the cross validation method as a choosing procedure for smoothing parameters, they show that their new estimator performs much better than conventional nonparametric estimators in the presence of mixed data. Blundell and Powell (2004) developed and implemented a single-index binary response model for estimating binary response models with continuous endogenous regressors. They show that semiparametric estimators work well in contrasting probit and linear probability models, and in detecting for significant attenuation bias. Furthermore, multivariate-based distributions used in economic research, is another difficulty in the semiparametric estimation procedure. To account for this phenomenon, Chen and Fan (2006) suggest a Copula-based semiparametric stationary Markov model characterized by a parametric copula and a nonparametric marginal distribution. A copula serves a heuristic in constructing a multivariate regression and represents general types of dependence. Bickel, Ritov and Stoker (2006) constructed a score test, a new framework for general semiparametric hypotheses that has a nontrivial power on the $n^{\frac{1}{2}}$ scale in every direction.

Likewise, the Simar and Wilson (2007) approach is a coherent data-generating process which is utilized by implementing a single and double bootstrap procedure.

Recently, semiparametric functional estimation techniques such as the kernel method have been used to construct consistent model specification tests. Statisticians Robinson (1988), Hong and White (1995) generated the idea of testing parametric and semiparametric regression models. Similarly, Hardle and Mammen (1993) suggest the use of the wild bootstrap procedure. Blundell, Duncan, Pendakur (1998) introduced a specification testing procedure for determining the endogeneity of variables by implementing semiparametric methods in an Engel curve relationship using British family expenditure survey data. They also discussed a useful method for pooling nonparametric Engel curves across households with different demographic compositions. Recently, Hsiao, Li and Racine (2007) proposed using a nonparametric kernel-based model specification test for mixed data (discrete and continuous) by means of using the cross-validation method. Using simulation results, they found that the proposed test has a significant advantage over other conventional frequency based kernel tests.

3.3.2. Semiparametric Smoother

In the semiparametric regression procedure, the regression consists of two forms: a parametric and a nonparametric form. The parametric procedure is simple and straightforward if the relevant variables are available and the appropriate functional form is known, but I have to consider different factors in the nonparametric procedure. The main aim of nonparametric regression is in the smoothing response variables that stem from one or more of the dependent variable. In practice, continuous variables are used to smooth dependent variables with few assumptions about functional forms of nonlinearity (Keele, 2008). The semiparametric regression procedure is often called additive or Generalized Additive Model (GAM). A smoothing spline and/or a

Kernel smoother are two common tools used for smoothing in both non- and semiparametric models.

3.3.2.1. Kernel Smoother

A kernel smoother is a weighting function used in a kernel regression to estimate the conditional expectation of random variables. The function is denoted by *K(u)* and satisfies the following properties

(i) $\int_{-\infty}^{\infty} \mathrm{K(u)du} = 1$ (3.25)

(ii) $\mathrm{K(u)} = \mathrm{K(-u)}$ (3.26)

(iii) $\int \mathrm{u^2K(u)du} = \mathrm{k_2} > 0$ (3.27)

The kernel function is a consistent estimator of nonparametric component $f(x)$ and the symmetry condition (ii) implies that

$\int \mathrm{u\,K(u)du} = 0$ (3.28)

According to Silverman (1986), the kernel estimator with kernel *K* is defined as

$\hat{\mathrm{f}}(\mathrm{x}) = 1/\mathrm{nh} \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K}\left(\frac{\mathrm{x-x_i}}{\mathrm{h}}\right)$ (3.29)

where h denotes bandwidth, which is generally described as a smoothing parameter (Silverman, 1986). The kernel function determines the shape of the bumps while the bandwidth determines their width. In general, there are six types of kernel functions (Silverman, 1986), and they are the Uniform, Epanechnikov, Biweight, Triangular, Gaussian and Rectangular. Each density function and its associated efficiency are shown in Table 3.1.

Another fundamental part of nonparametric statistics is the selection of an appropriate smoothing parameter. If the smoothing parameter (bandwidth *h* for kernel smoothing) is small, the resulting estimator will have a small bias but a large variance. On the other hand, if *h* is large, the resulting estimator will have a large bias but small variance. Therefore, in empirical work, deciding upon the most appropriate smoothing value involves a trade-off between variance and bias. Optimal bandwidth is obtained by the minimizing the Integrated Mean Square Error (IMSE). The detail of this IMSE procedure is available in Li and Racine (2007). There are mainly two principal methods for bandwidth selection: (1) Rule of Thumb or Plug-In Method, and (2) Cross-Validation Methods. A detailed procedure for computing optimal bandwidth is available in Li and Racine (2007),

Table 3.1. Different Types of Kernel Smoothers

Kernel	**K(t)**	**Efficiency**
Epanechnikov	$\frac{\frac{3}{4}\left(1-\frac{1}{5}t^2\right)}{\sqrt{5}}$ for $\|t\| < \sqrt{5}.$ 0 Otherwise	1
Biweight	$\frac{15}{16}(1-t^2)^2$ for $\|t\| < 1$ 0 Otherwise	0.9939
Triangular	$1-\|t\|\ for\ \|t\| < 1,$ 0 Otherwise	0.9859
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$	0.9512
Rectangular	$\frac{1}{2}\ for\ \|t\| < 1, 0\ otherwise$	0.9295

Source: Silverman (1986)

3.3.2.2 Smoothing Spline

The smoothing spline method depends upon minimizing the residual sum of squares between a response variable y and the nonparametric estimate, $f(x_i)$. The Residual Sum of Squares (RSS) for one variable is given by

$$RSS = \sum_{i=1}^{n}[y - f(x_i)]^2 \ . \tag{3.30}$$

When there is more than one independent variable, the spline smoothing needs to be penalized by a factor. Therefore, the minimization of RSS is subject to the penalty for the number of local parameters used for spline smoothing (Keele 2008). The penalty for spline models is $\lambda\int_{x_1}^{x_n}[f''(x)]^2dx$ (Wood, 2006). This term is also known as a roughness penalty constraint. The first part, λ, is the smoothing parameter, and the second term, which consists of the second derivative of $f(x)$, measures the rate of change of function, so the high value of second derivative imply the high curvature and vice versa. In another words, the Hessian matrix measures the amount of curvature around the likelihood maximum (Keele, 2008). Thus, a spline estimate is given by the minimization of the following expression.

$$SSR(f, \lambda) = \sum_{i=1}^{n}[y - f(x_i)]^2 + \lambda \int_{x_1}^{x_n}[f''(x)]^2 dx \ . \qquad (3.31)$$

where, spline smoothing is to minimize the sum of squares between y and the nonparametric estimate. The term $f(x)$ is subject to a penalty factor λ, which represents the smoothing parameter. Similar to kernel estimators, a very small value for λ gives over fitting close to data and a large λ value produce a least square fit. Therefore, I need to find an appropriate smoothing value that fits the semiparametric regression model well. Selection criteria of the smoother in semiparametric model are given in section 3.3.5.

3.3.3. Semiparametric (Generalized Additive Regression) Model

Statisticians and economists propose different types of semiparametric models includes the (1) Single Index Model, (2) Multiple Index Model, (3) General Additive Model, (4) Partial Linear Model and (5) Smooth Coefficient Model. Literately few kernel based semiparametric models are used in agricultural economics; however, the spline based semiparametric model is very rare in this field. To illustrate, Robert and Key (2008) used generalized additive model in the study of agricultural payments and land concentration in the United States, and find that semiparametric models account strong association between government payments and concentration of cropland and farmland. This research applies a semiparametric regression called a Generalized Additive Model (Hastie and Tibshirani, 1990). Likewise, this research is interested to utilize a General Additive Model to account nonlinearity in off-farm labor supply model. The semiparametric additive model with nonparametric and parametric terms and takes the following form:

$$g(Y_i) = \alpha + \beta X_i + f\left(X_{ji}\right) + \varepsilon_i \tag{3.32}$$

In this model, the covariate X is assumed to have a linear effect. The covariates X_{ji} are non-linear covariates and are fitted by a nonparametric estimation procedure. The parametric part of the model allows for the existence of discrete independent variables, such as dummy variables. The nonparametric terms; however, contain only continuous covariates. This model can be solved by using the penalized likelihood maximization method. Iterative algorithmic procedures are required for parameter estimation of this model. A description of this estimation procedure is available in Wood (2006); Following Wood (2006), the semiparametric generalized additive model can be written as

$$g(\mu_i) = X_i\theta + f_1(x_{1i}) + f_1(x_{2i}) + \cdots \tag{3.33}$$

where $\mu_i \equiv E(y_i)$ and $y_i \sim$ is 'an exponential family distribution'; g is a known, monotonic and twice differentiable link function (a probit). X_i is the i^{th} row of parametric model matrix, with parameter vectors β, and f_i serving as smoothing functions for the nonparametric covariates x_j.

To estimate the model specified in equation (3.33), I specify the basis, b_{ji}, for each smooth function, so that the smooth function can be represented as

$$f_j\left(x_j\right) = \sum_{i=1}^{q_j} m_{ji} b_{ji}(x_j), \text{ and implies } \int_{x_1}^{x_n} [f''(x)]^2 dx = \beta^T D^T B^{-1} D\beta. \tag{3.34}$$

Suppose $S \equiv D^T B^{-1} D$ (3.35)

, then S is called the penalty matrix for the basis. where $\boldsymbol{B}$ and $\boldsymbol{D}$ are the basis function and matrices used to define a regression spline as shown in Table 4.2 of Wood (2006); x_j are vectors of nonparametric components, $m_{ji}$are coefficients of the smooth function and, of which, are required to be estimated in this model. Using a basis, I can create the model matrix

$\tilde{X}_j = b_{jk}(x_{ji})$

And

$\tilde{\beta}_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jq_j}]^T$, that implies

$$f_j = \tilde{X}_j \tilde{\beta}_j \tag{3.36}$$

Wood (2006) has shown that equation (3.36) can be reparameterized subject to the constraint$1^T \tilde{X}_j \tilde{\beta}_j = 0$, such that for any $q_j - 1$, the orthogonal column matrix Z satisfies $1^T \tilde{X}_j Z = 0$. After the reparameterization, the new parameter β_j and new model matrix X_j satisfy the following two conditions:

$$\tilde{\beta}_j = Z\beta_j \text{ and } X_j = \tilde{X}_j Z \text{ ;Then} \tag{3.37}$$

$$g(\mu_i) = X_i \beta \tag{3.38}$$

Where, $\beta^T = [\theta^T, \beta^T, \beta^T, \dots]$. Hence, equation (3.38) is similar to a parametric generalized linear model and estimated by a penalized likelihood maximization. Wood (2006) suggests that convenient penalties be applied in this model would be quadratic form in type. Then, the penalized likelihood is expressed as

$$l_p(\beta) = l(\beta) - \frac{1}{2}\sum_j \lambda_j \beta^T S_j \beta = l(\beta) - \frac{1}{2}\beta^T S \beta \tag{3.39}$$

where $= \sum_j \lambda_j S_j$; λ_j is a smoothing parameter that manipulates the tradeoff between goodness of fit of the model and smoothness, and S_j is a matrix of known coefficients. Using penalized the iteratively re-weighted least square maximization process and properties of exponential family distribution. I can write

$$\frac{\partial l_p}{\partial \beta_j} = \frac{\partial l}{\partial \beta_j} - [S\beta] = \frac{1}{\emptyset}\sum_{i=1}^{n} \frac{y_i - \mu_i}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} - [S\beta] = 0 \tag{3.40}$$

The solution of this equation is

$$S_p = \sum_{i=1}^{n} \frac{y_i - \mu_i}{Var(Y_i)} + \beta^T S \beta \tag{3.41}$$

and Wood (2006) has shown that

$$S_p \cong \left\|\sqrt{W^{[k]}}\left(z^{[k]} - X\beta\right)\right\|^2 + \beta^T S\beta \quad (3.45)$$

where, $W^{[k]}$ is diagonal matrix of weight and $z^{[k]}$ is pseudo data. Wood (2006) then outlines the procedure in two steps

1. Given the current $\mu^{[k]}$, calculate the pseudo data $z^{[k]}$ and weights $w_i^{[k]}$.
2. Minimize equation (3.45) with respect to β in order to find $\hat{\beta}^{[k+1]}$.Evaluate the linear predict$or, \eta^{[k+1]} = X\hat{\beta}^{[k+1]}$, and fitted values $\mu^{[k+1]} = g^{-1}({\eta_i}^{[k+1]})$, incrementally.

In this penalized least square estimation problem, the influence matrix A is specified as

$$A = X(X^T WX + S)^{-1} X^T W\mathrm{I} \quad (3.46)$$

and for the un-weighted additive model

$$A = X(X^T X + S)^{-1} X^T \quad (3.47)$$

Wood (2006) shows that the parameter estimates after maximizing the penalized likelihood function is:

$$\hat{\beta} = (X^T X + S)^{-1} X^T y \quad (3.48)$$

3.3.4. Smoothing Parameter Selection Criteria

The criterion for selecting the smoothing parameter is to minimize the mean square error. When the scale parameter of the distribution is known, the minimization of expected mean square error is equivalent to Mallows’ Cp/UBRE (Un-Biased Risk Estimator; Craven and Wahba, 1979). However, when the scale parameter is not known, the cross validation method is useful.

1. For a known scale parameter: UBRE: The mathematical form of this criteria is

$$\upsilon_u(\lambda) = \frac{E\left(\|y - Ay\|^2\right)}{n} - \sigma^2 + \frac{2tr(A)\sigma^2}{n} \quad (3.49)$$

2. For an unknown scale parameter: When the scale parameter is unknown, I am unable to use equation 3.48 for smoothing, as it requires value of σ^2. Instead, prediction error is used as a base smoothing parameter. This method involves omitting a datum, y_i from the model at first and estimating a scale parameter so that the ordinal cross validation score can be computed, and is expressed as

$$\upsilon_0 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\mu}_i^{[i]}\hat{\mu}^{[-i]}{}_i\right)^2 \tag{3.50}$$

where $\hat{\mu}_i^{[i]}$ is predicted value of $E(y_i)$ obtained from the model fitted to all remaining data except y_i itself. Wood (2006) then shows that the ordinary cross validation score is

$$\upsilon_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i^{[i]})^2}{(1-A_{ii})}\,. \tag{3.51}$$

and, the general form of υ_0 is

$$\upsilon_u(\lambda) = \frac{E(\|y-Ay\|^2)}{[n-tr(A)]} \text{ (Hastie and Tibshirani; 1990).} \tag{3.52}$$

This method is called the Generalized Cross Validation Score (GCVS).

3.4. Variable Selection Procedure

The importance of variable selection in a semiparametric model process was highlighted at the beginning of this current chapter; however, in real data analysis, it is not an obvious rule to place covariates in parametric and nonparametric components. For example, in off-farm labor supply, it might be logical to assume that household net worth is a nonparametric variable and that need to be smoothing variable because a financially well-established farm households are less likely to work off-farm. However, I need to confirm it before using it as a nonparametric variable. This assumption needs to be validated. A statistical procedure developed by Blundell and Duncan (1998) is used to select which variables are parametric and which are nonparametric.

According to this methodology, if covariates are endogenous in their nonparametric component, they are nonparametric variables else they are parametric. Let x be endogenous in the model

$$y = g(x) + \epsilon \tag{3.53}$$

In the sense that

$E(\varepsilon|x) \neq 0 \; or \; E(y|x) \neq g(x)$ in which case $\hat{g}_h(x) \nrightarrow g(x)$ so the nonparametric estimator will not be consistent. Instead, consider a different variable, z, which holds the following relation

$$x = \pi z + \upsilon \text{ with } E\left(\frac{\upsilon}{z}\right) = 0 \text{ and} \tag{3.54}$$

$$y = g(x) + \upsilon\rho + \epsilon \;\text{ with }\; E(\epsilon/x\,) = 0 \tag{3.55}$$

Then, let us assume that there is endogeneity with null hypothesis

$H_0: \rho = 0$ Nonparametric variable

$H_1: \rho \neq 0$ Parametric variable (3.56)

This test statistic determines the endogeneity of variables in the semiparametric model.

3.5. Specification Test

Comparison of the models on the results produced by parametric and semiparametric model is one of the important aspects of developing a semiparametric model. Many recent journal articles have examined the significance of the two models. For example, Goodwin and Holt (2002) compare the Single Index model with a Tobit model using Housman test and found that single index model is specified. In similar manner, this research compares a semiparametric model with the parametric model of labor supply. Statisticians and economists like Hong and White (1995), Zheng (1996), and Li and Wang (1998) proposed model specification test. The Hong and White (1995) test which is based on residual of models is used in this study and subsequently discussed in the next section.

3.5.1. Hong and White Test

Hong and White (1995) introduced a consistent test of functional form via nonparametric techniques. They set up a null hypothesis that the parametric model is correct against the nonparametric as an alternative model. The test can is:

H_0: Parametric Model (3.57)

H_1: Semiparametric Model (3.58)

The test statistics $\widetilde{T}_n$ is given by

$$\widetilde{T}_n = (n\widetilde{m}_n/\hat{\sigma}_n^2 - P_n)/(2P_n)^{1/2} \tag{3.59}$$

Where, $\widetilde{m}_n = n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{nt}^2 - n^{-1}\sum_{t=1}^{n}\hat{\eta}_{nt}^2$ (3.60)

where $\hat{\eta}_{nt}$ is the residual from nonparametric estimation, $\hat{\sigma}_n^2$ is a consistent estimator for the variance of the error term under H_0, P_n is dimension of parameter for parametric covariates, $\hat{\varepsilon}_{nt}^2$ regression error from parametric estimation procedure. Hong and White proved that as $n \to \infty$, $T_n \xrightarrow{d} N(0,1)$ under H_0. The hypothesis H_0 is rejected for large values of T_n.

CHAPTER 4

DATA AND EMPIRICAL MODEL

Data collected under the USDA's National Agriculture Statistical Service's Agricultural Resource Management Survey (ARMS) conducted in 2006 is used in this study. The 2006 ARMS data set is a large data set providing a host of information regarding the U.S. agricultural sector including the survey of household activities such as labor allocation for both on- and off-farm labor. Farm operators' and their spouses' characteristics like operator's age, years of formal education and their health insurance status are also available in the data set. The data also provide family characteristics such as number of household members along with their ages. In order to account for non-linearity in both operators' and spouses' ages, an *age squared* variable is included in the model. The ARMS database contains different farm program payments, including income, expenses and type of farms. The descriptive statistics of important factors used in this research in labor allocation to nonfarm employment sectors are presented in Table 4.1.

ARMS collects data on hours of worked on farm and off the farm for both farm operators and their spouses. The average annual off-farm work hours are forty-three hours for operators and sixty six hours for spouses. This figure indicates that the annual off-farm work by spouses is compared to farm operators is greater. Although, annual hours of off-farm work is accessible, I create new dummy variables for both operators and spouses based on whether they supply labor to off-farm work. In my analysis, a value of *1* is assigned if the individual (operator or spouse) works off-farm and *0* otherwise.

Literatures in off-farm labor supply (Huffman, 1980; Goodwin and Mishra, 1997) argue that off-farm work experience is one of the most important factors affecting off-farm labor

Table 4.1. Definition and Summary Statistics of Variables Used in the Analysis

Variable Code	Variables Definition	Mean	Standard Deviation	Minimum	Maximum
Dependent Variables					
of_op	=1 if worked off the farm, 0 otherwise (Operator)	0.30	0.46	0.00	1.00
of_sp	=1 if worked off the farm, 0 otherwise (Spouse)	0.46	0.49	0.00	1.00
Characteristics					
op.age	Age of operator (years)	55.39	12.03	19.00	92.00
sp.age	Age of spouse (years)	52.80	11.87	17.00	92.00
op.educ	Years of formal education, operator	13.46	1.91	10.00	16.00
sp.educ	years of formal education, spouse	13.58	2.21	10.00	16.00
ophthins	=1 if the farm operator received health insurance through off-farm work, 0 otherwise	0.19	0.39	0.00	1.00
sphthins	=1 if the farm spouse received health insurance through off-farm work, 0 otherwise	0.23	0.42	0.00	1.00
Family Characteristics					
hhsize06	Number of household member under age 6	0.15	0.50	0.00	6.00
hhsize13	Number of household member between 13 and 17	0.54	1.00	0.00	7.00
hhnw1	Household net worth ($1000)	2177.45	7224.52	0.00	380292.25
Farm Characteristics					
direct	Direct farm program Payments	8610.71	24943.20	0.00	666349.00
indirect	Indirect farm program payments	8900.35	27490.58	0.00	655200.00
fowner	= 1 if the farm is full owned , 0 otherwise	0.40	0.49	0.00	1.00
powner	= 1 if the farm is partially owned, 0 otherwise	0.49	0.50	0.00	1.00
crppayment	Conservation reserve payments	795.57	7239.23	0.00	295923.00
vprod1	Farm size, value of agricultural output sold ($1000)	750.82	2016.56	0.00	46000.00
insur	= 1 if the farm has crop insurance, 0 otherwise	0.32	0.47	0.00	1.00
entropy	Entropy measure of farm diversification	0.14	0.14	0.00	0.58
Local Economic Condition					
metro1	= 1 if the farm is located in a metro county, 0 otherwise	0.34	0.47	0.00	1.00
Sample size			5144		

allocation. Unfortunately, the 2006 ARMS data used in this study does not contain any information about off-farm work experience. The levels of education for their family members are available and are disaggregated in this paper into four levels of formal education. The four categories are under high school, high school, undergraduate and graduate. The status of operator's and spouse's health insurance received from off-farm work is used in our research as an indicator variable. Both the number of household members and their age is useful in determining the impact children have on their parents' labor allocation for both farm and non-farm activities. In this study, the number of children in a household is categorized into two groups: under the age 6, and between the age of 13 to 17.

Literature also shows that household net worth is an important factors in the off-farm labor allocation. Household net worth is a measure of the financial wealth of a household. Further, different farm characteristics such as ownership and type of subsidies farms received from different governmental subsidy programs are used in this research. To illustrate, information regarding different government farm program payments such as direct, indirect, and conservation reserve payments is available in the 2006 ARMS data. This research is primarily interested in studying the effect of direct, indirect, and conservation reserve payments on the off-farm labor allocation decision. Here, direct payments are decoupled farm program payments and indirect farm program payments are coupled farm program payments. On average, farms received \$8,145 in direct, \$8,337 in indirect farm program payments, and \$780, on average, in conservation reserve payments. This study uses the value of agriculture production by the farm as a proxy for farm size in this study. The status of farm crop insurance available in 2006 ARMS data obtained by farms is also of interest in our area of research. The type of county, metro or non-metro, were included in the data as to assess the influence of farm location on off-farm labor force participation. Finally, after removing some missing observations, 5,144 observations were used in this analysis.

4.1. Empirical Model

The theoretical model discussed in chapter three suggests that the number of hours supplied in off-farm work by farm operators and their spouses depends upon their age, education level, health insurance obtained from non-farm employment, number of children present in the family, and farm characteristics like ownership, farm size, farm diversification (entropy), crop insurance. Different farm program payments they received and location of farm are also a part of off-farm labor allocation decisions. In this research, I am interested in whether the operator and spouses choose to work on or off the farm. Therefore, our dependent variable is a binary variable with a value of *1* if they work in the non-farm sector, *0* otherwise. Our empirical estimation procedure consists of three parts. First the study estimates both a parametric probit model, and semiparametric model of off-farm labor supply. To this end, I test the specification of model using graphical plots of the model, the likelihood ratio test, and a Hong and White (1995) test. The off-farm function of farm operators and their spouses in parametric framework can be written as in equation (4.1).

$$Ofw_i = G(\beta_0 + \ \beta_1\,(age)_i + \beta_1\,(age)_i^{\,2} + \beta_2\,(op.educ)_i + \beta_3\,(phthins)_i + \beta_4\,(hhsize06) + \beta_5\,(hhsize13) + \beta_6\,(hhnw1) + \beta_7\,(direct) + \beta_8\,(indirect) + \beta_9\,(fowner) + \beta_{10}(powner) + \beta_{11}\,(crppayment) + \beta_{12}\,(vprod1) + \beta_{13}\,(insur) + \beta_{14}(entropy) + \beta_{15}(metro1) + \epsilon \qquad (4.1)$$

Where *i* represent operator and spouses. In addition, *G* denotes the link function. In my model, *G* denotes a probit function. The variables retain the definitions provided in table 4.1. Here the variable ofw_i represents the dummy variable for the decision to work off-farm.

As mentioned in chapter three, a GAM is utilized for estimation in this study. Generally, the choice of a variable in two parts (parametric and nonparametric) comes from economic theory. For example, in the study of Environmental Kuznets curve, covariate income is considered as a semiparametric covariate (Paudel et al, 2005)**.** Likewise, in this study, I consider continuous variable, household net worth and the value of agricultural production as a nonparametric component in the semiparametric model. However, I did not find strong evidence to consider them in nonparametric variables. Poudel, Paudel, and Bhattarai (2009) used methods suggested by Blundell and Duncan (1998) to verify that forestry was a nonparametric component in the study regarding Environmental Kuznets curve. Following Poudel, Paudel, and Bhattarai's (2009) idea and the method as suggested by Blundell and Duncan (1998), the same approach is applied in selecting the nonparametric covariate. The null and alternative hypothesis in the test statistics are:

H_0: covariate is nonparametric

H_1: covariate is not nonparametric

Hence, if the test statistic is significant at a given level of significance, the variables are a parametric component otherwise; it is a nonparametric covariate. The results of the test statistics are shown in Table 4.2.

Generally, continuous variables are used for smoothing, so I apply the Blundell and Ducan (1998) approach to check for the variable characteristics in the semiparametric model. Table 4.2 shows that all continuous variables are significant at the 1% level of significance except for variable *vprod1* (the value of production). Therefore, *vprod1* is a nonparametric covariate in our semiparametric analysis. The econometric framework of semiparametric probit model is represented by the following equation:

Table 4.2. Variable Selection for Semiparametric Model Using Blundell and Duncan (1998)

Variables	Rho	Z-value	P Value	Type of Variables
hoffop	-	-	-	Dependent
hoffsp	-	-	-	Dependent
op.age	-26.12746	-5.67509	<0001	Parametric
sp.age	-14.23304	-4.48124	<0001	Parametric
op.educ	614.22730	5.14686	<0001	Parametric
sp.educ	-28.78883	-3.20613	<0001	Parametric
ophthins	-	-	-	Categorical Variable
sphthins	-	-	-	Categorical Variable
hhsize06	-1482.78800	-5.02485	<0001	Parametric
hhsize13	116.47640	5.21200	<0001	Parametric
hhnw1	0.00567	3.77395	<0001	Parametric
direct	0.00357	5.64055	<0001	Parametric
indirect	0.00222	5.23363	<0001	Parametric
fowner	-	-	-	Categorical Variable
powner	-	-	-	Categorical Variable
crppayment	-0.08888	-3.64255	<0001	Parametric
vprod1	0.00229	0.30579	0.680	Non Parametric
insur	-	-	-	Categorical Variable
entropy	-1446.17200	-4.91314	<0001	Parametric
metro1	-	-	-	Categorical Variable

$$Ofw_i = G(\beta_0 + \beta_1 (age)_i + \beta_1 (age)_i^2 + \beta_2 (op.educ)_i + \beta_3 (phthins)_i + \beta_4 (hhsize06) + \beta_5 (hhsize13) + \beta_6 (hhnw1) + \beta_7 (direct) + \beta_8 (indirect) + \beta_9 (fowner) + \beta_{10}(powner) + \beta_{11} (crppayment) + f\ (vprod1) + \beta_{13} (insur) + \beta_{14}(entropy) + \beta_{15}(metro1) + \epsilon \qquad (4.2)$$

Where *f(.)* represent the nonparametric component in the semiparametric regression.

CHAPTER 5

RESULTS AND DISCUSSION

The chapter three focused on the parametric and semiparametric estimation procedures that are applicable to analyze off-farm labor supply in the United States. In this chapter, both parametric and semiparametric estimation methods are used to analyze data from the Agriculture Resource Management Survey (ARMS) 2006 on labor allocation. The results obtained from parametric and semiparametric procedure are compared using Hong and White (1995) and Likelihood Ratio (LR) test at the end of this chapter.

5.1. Probit Estimation Results

The probit estimators described in chapter 3 are applied to equation (4.2) using the ARMS data. The result of parametric model (4.1) and the marginal effect of the operator and spouses are presented in table 5.3 and 5.4 respectively. The notation (a) in the tables indicates the inconsistent significance result for operator and spouse. The results are based on the robust standard errors (RSS) method as RSS can remove heterogeneity among the dependent variables. The positive sign on the estimated *age* coefficient and the negative sign on estimated coefficient on the age square (*opagesq*) variable implies that there is a parabolic relationship between age and labor allocation to the off-farm labor decision. This indicates that the probability of off-farm work increases with age, but at a decreasing rate. In particular, the marginal estimates imply that an additional year increases the probability of off-farm employment of the operator by 2.4%, but at a decreasing rate (table 5.1). Similarly, results were obtained for the spouse labor supply. Results indicate that the probability of off-farm employment by spouse increases by 2.7 percent, but at decreasing rate. In addition, the probability of off-farm labor supply by operator starts to decrease at 43 years, and the probability starts to decrease at 33 years in case of spouse.

The coefficient of education level (*opeduc*) is positive and significant at the 1 percent level of significance. In case of operator's education, findings support that farm operators with higher levels of education are more likely to participate in off-farm employment. In particular, the marginal effect indicates that an additional year of schooling increases the likelihood of off-farm work by 1.7 percent (table 5.1). Likewise, results in table 5.2 show a positive and significant effect of education on off-farm labor supply by spouses. The marginal effect reveals that an additional year of schooling increases the probability of working on off-farm by 4.8 percent (table 5.2). The likelihood of off-farm participation by spouses is approximately twice that of farm operators, which might indicate that spouses are more likely to work off-farm work than operators are, ceteris paribus. Findings from this study are in agreement with the results of Huffman (1980).

Some non-farm jobs provide health insurance to their employees, which likely attract farm operators and their spouses to non-farm employment. To assess the impact of health insurance, operator's and their spouse's health insurance status/availability is included in the regression. The estimated parameter of off-farm work health insurance (*ophthins*) is positive and significant at the 1 percent level of significance (table 5.1), indicating that health insurance also plays an important role in labor allocation. Moreover, the higher marginal effect (=0.49) of health insurance indicates that if an operator receives health insurance from off-farm work, the probability of working off-farm increases by 49 percent. Consistent with the decisions of farm operators, our results suggest that the chance of the spouse working off the farm is almost 60 percent more, compared to the spouses who does not receive health insurance from off-farm employment.

Table 5.1. Parameter Estimates and Summary Statistics for Probit Model of Off-Farm Labor Supply for Operator

Variables	Estimate		z-Value	∂y/∂x	
Intercept	-2.73439	***	-6.36000		
opage	0.07510	***	5.09000	0.02424	***
opagesq	-0.00088	***	-6.59000	-0.00028	***
opeduc	0.05512	***	4.96000	0.01779	***
ophthins	1.35583	***	25.60000	0.49073	***
hhsize06[(a)]	0.03416		0.76000	0.01103	
hhsize13[(a)]	-0.03501		-1.59000	-0.01130	
hhnw1	-0.00003	**	-2.48000	-0.00001	**
fowner[(a)]	0.20781	***	2.69000	0.06788	***
powner	0.10984		1.48000	0.03546	
vprod1	-0.00023	***	-2.94000	-0.00007	***
crppayment[(a)]	0.00001	*	1.83000	0.00000	*
direct[(a)]	0.00001	*	-1.81000	0.00000	*
indirect	-0.00001	**	-2.47000	0.00000	**
insur[(a)]	-0.23988	***	-4.55000	-0.07509	***
entropy[(a)]	-0.03071		-0.17000	-0.00991	
metro1[(a)]	-0.01446		-0.32000	-0.00466	
Pseudo R^2			0.2543		
Wald Chisquare =			1154.66		
Log pseudolikelihood =			-2362.37		

Note: * indicates statistical significance at a= 0.1 level
* *indicates statistical significance at a= 0.05 level
*** indicates statistical significance at a= 0.01 level
a indicates inconsistent significance result between operator and spouse

Table 5.2. Parameter Estimates and Summary Statistics for Probit Model of Off-Farm Labor Supply for Spouse

Variables	**Estimate**		**z-Value**	**∂y/∂x**	
Intercept	-2.50541	***	-5.59000		
spage	0.06831	***	4.13000	0.02720	***
spagesq	-0.00103	***	-6.44000	-0.00041	***
speduc	0.12131	***	11.61000	0.04830	***
sphthins	1.74967	***	28.22000	0.58033	***
hhsize06[(a)]	-0.30980	***	-6.66000	-0.12333	***
hhsize13[(a)]	-0.05313	**	-2.44000	-0.02115	**
hhnw1	-0.00002	**	-2.35000	-0.00001	**
fowner[(a)]	-0.00532		-0.07000	-0.00212	
powner	-0.01239		-0.18000	-0.00493	
vprod1	-0.00009	***	-2.75000	-0.00003	***
crppayment[(a)]	0.00001		0.50000	0.00000	
direct[(a)]	0.00001		-0.83000	0.00000	
indirect	0.00001	***	-2.64000	0.00000	***
insur[(a)]	-0.02481		-0.49000	-0.00987	
entropy[(a)]	0.40100	**	2.38000	0.15964	**
metro1[(a)]	-0.08577	*	-1.94000	-0.03410	*
Pseudo R^2			0.3108		
Wald Chi-square = 1256.55			1256.55		
Log pseudo likelihood =			-2447.53		

Note: * indicates statistical significance at a= 0.1 level
* * indicates statistical significance at a= 0.05 level
*** indicates statistical significance at a= 0.01 level
a indicates inconsistent significance result between operator and spouse

The presence of children in the households would limit the time available for working off-farm; because, women have traditionally devoted more time taking care of children and performing general homemaking duties. As expected, the coefficient of variables that represents the number of children under age 6 (*hhsize06*) for the spouse is negative and highly significant (table 5.2). This result shows that the probability of off-farm work by spouses decreases with each additional child under the age of 6. The marginal effect (-0.12) imply that an additional child under 6 in the number decreases the probability to working off-farm for spouses by 12 percent. In the case of the number of children between age 6 and 17, the coefficient is also negative and highly significant. This result reveals existence of negative correlation between the number of children and off-farm labor supply by spouses. In contrast, I did not find any significance between the number of children in the household and off-farm labor supply by farm operators.

In practice, a financially, well established (measured by household net worth) farm may have less incentive to work off the farm. Therefore, our expectation is that the allocation of labor hours in off-farm work is inversely related to household net worth. The coefficient of household net worth (*hhnw1*) is negative and significant at the 1% level of significance (table 5.1). The relatively small marginal effect reveals that farm operators who have higher net worth are less likely to work off-farm (table 5.1). Similarly, the probability of a spouse off-farm work decreases with increase in household net worth. Again the effect is very small (table 5.2).

Table 5.1 shows that the coefficient on full owner (*fowner*) is positive and significant at the one percent level of significance. This result suggests that full owners are more likely to work off-farm compared to farms operated by tenants (table 5.1). The marginal effect (=0.067) of full-ownership suggests that full owners are about 7% more likely to work off-farm compared to tenants. In contrast, I found the opposite, but not significant relationship between full ownership and off-farm work by spouses.

In my analysis, I included value of agricultural production (*vprod1*) as a measure of the farm size to assess the impact of farm size on off-farm labor supply by farm operators and their spouses. In labor economics literature, some economists (Mishra and Goodwin, 1997) argue that operators whose farm size is large are less likely to work off the farm. The coefficient of *vprod1* is negative and statistically significant at the 1% level of significance for both operators and spouses (table 5.1 and 5.2), indicating that as farm size increase, the probability of off-farm working by operators and their spouses decreases, but the marginal effect is very small. This conclusion is consistent with the finding of Fernandez-Cornejo, Hendricks and Mishra, 2005; Mishra and Holthausen, 2002; Sumner, 1982, Lass and Gempesaw, 1992; El-Osta, Mishra and Ahearn, 2004.

Theil's entropy index (*entropy*) is incorporated in my research to measure the impact of farm diversification on labor allocation. The *entropy* takes a value 1 for a diversified farm and 0 for specialized farms. The parameter estimate of *entropy* is negative for operators, but not significant. In contrast, the coefficient of *entropy* is positive and significant at the one percent level of significance in case of their spouse (table 5.2). This result indicates that the possibility of spouses working off the farm increases if the farm is diversified. The marginal effect (0.15) of entropy suggests that spouses from diversified farms are 15% more likely to work off the farm.

Widely studied, the literature of government farm program payments and their impacts on time allocation have shown that both direct and indirect farm payments were significantly and positively correlated with less off-farm work by farm operators (Chang and Mishra, 2008). In addition, Mishra and Sandretto (2001) point out that farm program payment stabilize total household income and hence lessen the need to work off the farm. As expected, results show that operators who received indirect farm program payments are less likely to work off the farm. This finding is consistent with the finding of El-Osta, Mishra and Ahearn (2004). The relatively less marginal effect of the payments reveal that farmers who receive more indirect payments are less likely to work off of the farm (table 5.1). Thus, this findings are consistent with the result of Dewbre and Mishra 2007; Ahearn, El-Osta, and Dewbre, 2006; El-Osta, Mishra and Ahearn, 2004. In contrast, the spouses are more likely to work off-farm with increase in farm program payments.

The aim of farmers is to maximize their profit, which depends upon the production of their agricultural products. Thus, they are likely to insure their production. To assess the impact of crop insurance, I include a dummy variable on purchase of crop insurance (*insur*). The estimated coefficient of crop insurance (*insur*) is negative and significant at 1% level of significance. This result indicates that the probability of working off-farm by farm operators who have crop insurance decrease compared to the farm operators who do not have crop insurance. The marginal effect (=-0.075) of crop insurance indicates that insured farm operators are 7.5% less likely to work on off of the farm compared to operators without crop insurance. This result is consistent with the findings of Key, Roberts and O'Donoghue (2009).

5.2. Semiparametric Results

The results of semiparametric probit model for operators and their spouses are shown in table 5.3 and table 5.4 respectively. The inconsistent result for parametric and semiparametric results are indicated by (b). Semiparametric results imply that the coefficient of age (*opage*) is positive and the coefficient of age square (*opagesq*) is negative for both operator and spouse, which are consistent with the findings of parametric estimation. These findings are statistically significant at the one percent level of significance. This model implies that an additional year in operator age increases the probability of off-farm labor supply by 2.28 %. Similarly, the off-farm employment of a spouse increase by 2% but with a decreasing rate. Furthermore, results show that the probability of off-farm labor supply by operator starts to decrease at 44 years, and the probability starts to decrease at 34 years in case of spouse. These results are nearly equivalent to the findings from parametric model.

The semiparametric results also in agree with parametric estimates for education. Results indicate that an additional year of schooling by farm operators increases the probability of working off-farm work by 1.33%, ceteris paribus (table 5.3). Compared to farm operators, spouses are more likely to work off-farm if they are more educated. The marginal effect of spouse education is 3.25, which means that the probability of spouses working off-farm increases by 3.25 % for each additional year of schooling (table 5.4). Our results agree with the findings of El-Osta, Mishra and Morehart (2008) that labor allocation to off-farm job is positively related with educational attainment of the operator and spouse.

The semiparametric model also supports the findings of parametric probit model that the probability of working off the farm is positively related with the health insurance provided by off-farm jobs. The marginal effect (=0.30) of health insurance for operators suggests that the probability of working off-farm is 30 percent more likely if operators receive health insurance from off-farm employment. Hence, our result suggests a significant impact of health insurance on labor allocation. Not only are farm operators more likely to work off the farm if they receive health benefits but also to are their spouses. Moreover, the probability of labor allocation to nonfarm employment is approximately 46% for the spouses if they receive health insurance from off-farm work.

Table 5.3. Semiparametric Estimates and Summary Statistics for Probit Model of Off-Farm Labor Supply for Operator

Variables	Estimate		z-Value	∂y/∂x	
Intercept	-3.31700	***	-8.00600	-0.81515	***
opage	0.09290	***	6.55900	0.02283	***
opagesq	-0.00107	***	-8.37700	-0.00026	***
opeduc	0.05422	***	5.07900	0.01333	***
ophthins	1.24700	***	24.83700	0.30637	***
hhsize06[(a)]	0.05456		1.20400	0.01341	
hhsize13[(a)]	-0.02172		-0.99700	-0.00534	
hhnw1[(a,b)]	-0.00001		-1.09300	0.00000	
fowner[(a)]	0.14020	*	1.89100	0.03445	
powner	0.09635		1.36100	0.02368	
crppayment[(a)]	0.00001	*	2.08200	0.00000	
direct[(b)]	0.00001		-0.39200	0.00000	
indirect[(a,b)]	0.00001		-1.59700	0.00000	
insur[(a)]	-0.17160	***	-3.41700	-0.04218	**
entropy	-0.22560		-1.38800	-0.05544	
metro1	0.00181		0.04200	0.00044	
Pseudo R^2			0.35000		
LR Test p-value			0.00000		
Log pseudo likelihood =			-2243.97900		

Note: * indicates statistical significance at a= 0.1 level
* * indicates statistical significance at a= 0.05 level
*** indicates statistical significance at a= 0.01 level
a indicates inconsistent significance result between operator and spouse
b indicates inconsistence significance results between parametric and semiparametric model

Table 5.4. Semiparametric Estimates and Summary Statistics for Probit Model of Off-Farm Labor Supply for Spouse

Variables	Estimate		z-Value	∂y/∂x	
Intercept	-2.80100	***	-6.72100	-0.73307	***
spage	0.07939	***	5.29700	0.02078	***
spagesq	-0.00117	***	-8.19200	-0.00031	***
speduc	0.12450	***	11.69200	0.03260	***
sphthins	1.75300	***	28.53900	0.45873	***
hhsize06[(a)]	-0.30790	***	-6.41300	-0.08060	***
hhsize13[(a)]	-0.04920	*	-2.24500	-0.01288	
hhnw1[(a)]	-0.00001	**	-2.64800	0.00000	*
fowner[(a)]	-0.09490		-1.30400	-0.02484	
powner	-0.03889		-0.56600	-0.01018	
crppayment[(a)]	0.00001		0.60500	0.00000	
direct	0.00001		0.11900	0.00000	
indirect[(a)]	0.00001	*	-2.04400	0.00000	
insur[(a)]	0.05337		1.08600	0.01397	
entropy[(b)]	0.26530		1.60100	0.06945	
metro1[(b)]	-0.06467		-1.50400	-0.01693	
Pseudo R^2			0.39200		
LR Test p-value			0.00000		
Log pseudo likelihood =			-2374.51900		

Note: * indicates statistical significance at a= 0.1 level
* * indicates statistical significance at a= 0.05 level
*** indicates statistical significance at a= 0.01 level
a indicates inconsistent significance result between operator and spouse
b indicates inconsistence significance results between parametric and semiparametric model

Similar result are found in the case of the number of children in the farm household. The coefficient for number of children under age 6 (*hhsize06*) is positive and coefficient of children between the ages of 13 and 17 (*hhsize13*) is negative (table 5.3), but only significant in this case for spouse (table 5.4). This finding suggests that increase in number of children under age seventeen, decreases the likelihood of off-farm work by spouses. Moreover, the probability of working off-farm by farm spouses decreases by 8% for an additional child under age 6. Likewise, our finding also suggests that the likelihood of working off-farm by spouses decreases by 1.28% per additional family member between age 13 and 17(table 5.4).

As mentioned in our discussion on parametric results, the negative estimated parameter in semiparametric model result of household net worth shows that a wealthy and well-established household is less likely to engage time in off-farm work. Similar impacts are found on spouse's labor allocation. The marginal effect of household net worth (*hhnw1*) for spouses is relatively very small, which implies that wealthy spouses are less likely to work off the farm.

Similar to the parametric estimates, the semiparametric estimates for the variable *fowner* positive and significant. This is also consistent with our finding from parametric model results. This estimate shows that the probability of full owner operator to work off-farm increase by 3% compared to tenants. In our semiparametric model, the positive and significant result for conservation reserve payments (CRP) indicates that an additional dollar CRP increases the probability of off-farm labor supply for farm operators; however, the marginal effect is very small. The estimated coefficients of direct payments (*direct*) are not significant for either operator or their spouses. But, indirect payments (*indirect*) has a significant negative effect on the spouse's off-farm labor allocation decision (table 5.4). These findings revel that spouses who received indirect government payments are less likely to work off-farm. However, the marginal effects imply that the probability of spouses working off-farm decreases minutely.

Similar to parametric probit estimation, the semiparametric estimation shows that crop insurance has negative and significant effect in off-farm labor supply. Moreover, the probability of working off-farm by farm operators decreased by less than 4.2% compared to operators who did not buy crop insurance. I find a contradictory result for spouses as indicated by positive sign of the estimated parameter (table 5.4). As expected, using a semiparametric additive model I found that the entropy measure of diversification has negative effect on operators off-farm labor supply, but not significant. However, for spouses the sign of coefficient (*entropy*) is positive and significant at the 1% level of significance. Results indicate that spouses from diversified farm are more likely to work off-farm compared to the spouses from specialized farm. Moreover, the probability of off-farm work by spouse's who are from diversified farm are 5.5% more likely to work off the farm compared to spouses from specialized farm.

5.5. Specification Tests

The univariate parametric probit model is compared to a semiparametric specification using Hong and White's test. The estimated T_n statistics and p-values are reported in Table 5.5. The parametric probit and semiparametric probit model is compared for both operator and spouses. It is found that semiparametric model is significant at the 1% level of significance. The effect of farm size on the decision of off-farm labor supply by farm operator and spouse are shown in figure 5.1 and 5.2 respectively. In the operator situation, the parametric model gives prediction above the semiparametric model within the range of farm size with value ranges from $375,000-2,800,000. Beyond 2.8 that range, two model have similar prediction. For spouses, the parametric predicted probability curve is above the semiparametric predicted probability curve in the given range, which tells that the parametric curve is over fitted to semiparametric curve. Hence, one can conclude that the semiparametric model is more appropriate estimation procedure compare to parametric model to analyze off-farm labor supply. I also performed the likelihood ratio (LR) test to specify correct model. The LR test p value given in Table 5.3 and Table 5.4 provides comparisons between parametric and semiparametric models. GAM fit is superior to parametric fit because the chi-square test statistics is 236 on 6.5 degree of freedom, which is highly significant. In the case of spouses, the chi-square test statistics is 146 at 7.75 degrees of freedom, which is also highly significant. Similarly, figure 5.2 shows that the parametric curve is above the semiparametric curve which implies that the semiparametric model is better than the parametric model. In summary, our findings strongly support a semiparametric model.

Table 5.5. Model Specification tests, Off-Farm Labor Supply in the United States *

Model	**T_n Value**	**P-Value**
Operator	41.74268 ***	0.000
Spouse	23.42644 ***	0.000

* Univariate Probit model is assumed as the null parametric model. T_n is the Hong -White specification tests which is asymptotically N(0,1) under H_0.

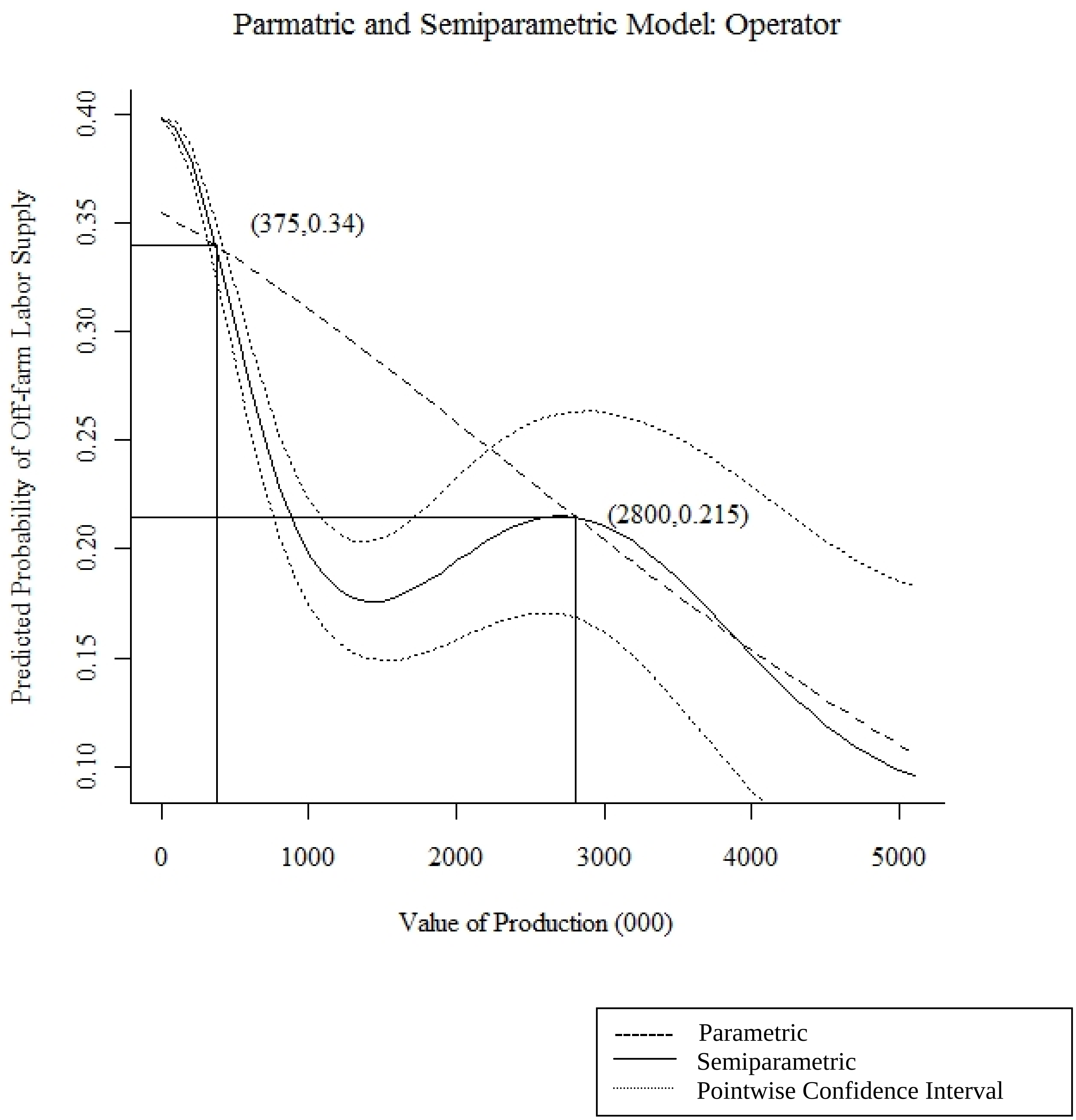

Figure 5.1. Comparision between Semiparametric and Parametric Specification for Operator

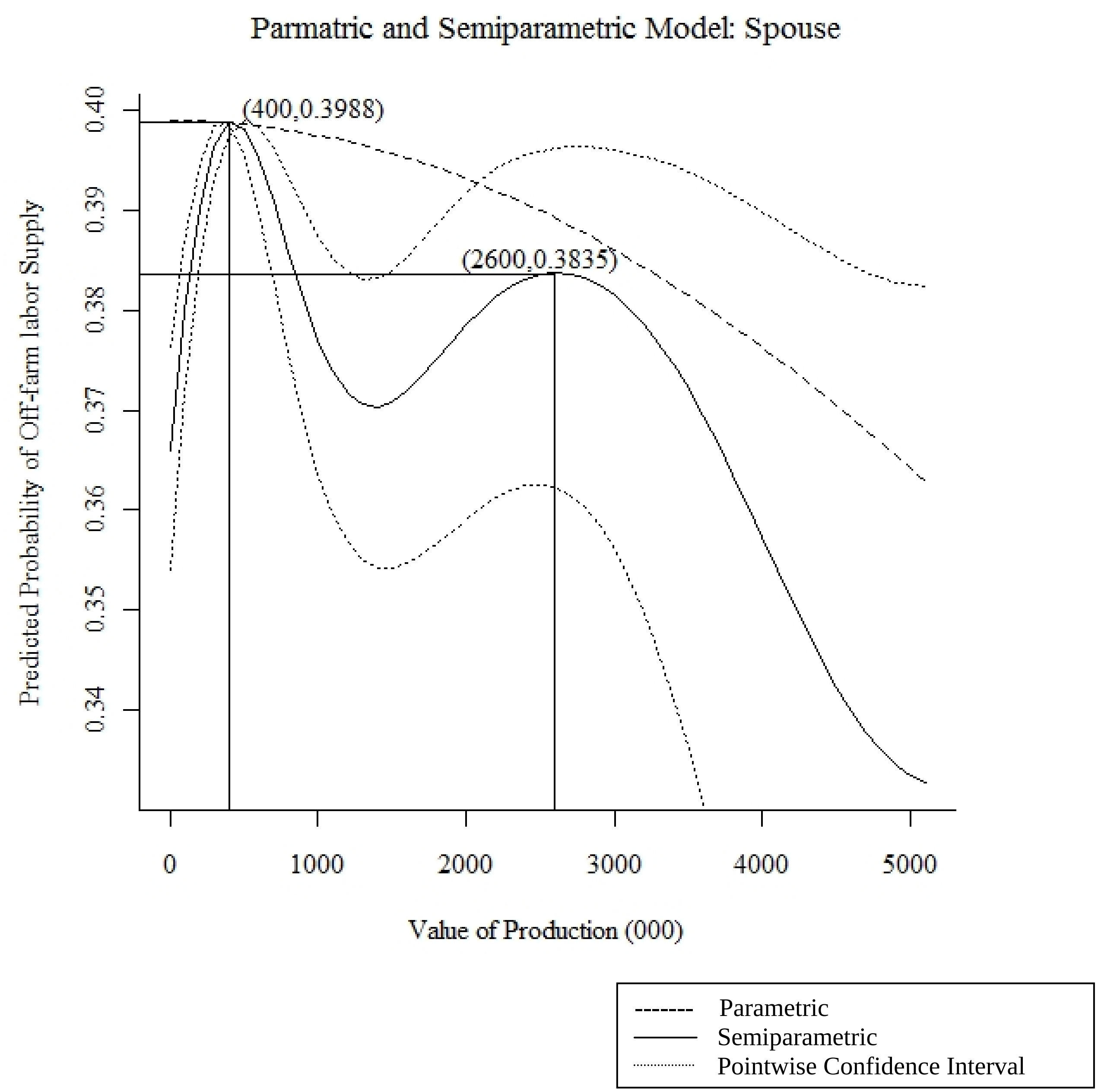

Figure 5.2. Comparision between Semiparametric and Parametric Specification for Spouse

CHAPTER 6

SUMMARY AND CONCLUSIONS

The objective of this study was to estimate parametric and semiparametric model in decisions of off-farm work for farm operators and their spouses in the United States. In particular, I am interested on comparing the results obtained for both parametric and semiparametric models. Using ARMS 2006 data, off-farm labor allocation equations are estimated for both operator and spouses using parametric and semiparametric generalized additive model. Result from the parametric and semiparametric model were compared using Hong and White (1995) specification test. Although, my results show that more variables are significant in the parametric probit model than in the semiparametric additive model, the specification test shows that the semiparametric additive model is more consistent than the probit model. In addition, the graphical comparisons of fitted lines of parametric and semiparametric models show that the probit model is over fitted. Therefore, our results predict that the semiparametric model is more informative and is better suited for estimating off-farm labor supply model.

In this study, I evaluated the role of operators, spouses characteristics, farm characteristics, and family characteristics in off-farm labor supply decisions by farm operators and spouses. The results confirm that operators and spouses characteristics, like age and education are positively correlated with off-farm labor supply. This means that the more educated farm operators and spouses are, more likely to work off the farm. Like previous studies, I found that the number of children in a household also help to determine off-farm labor supply for spouses. Findings imply that women are less likely to participate in off-farm employment, if small children are present in the household. As expected, I found that the wealthy farm operators and their spouses are less likely to work in off-farm work.

Farm operators who received government payments are less likely to participate in off-farm employment. I analyzed three types of payments: direct payments, indirect payments, and conservation reservation payments and their influence in the decision to work off-farm for both operators and spouses. I found that operators are significantly less likely to participate in off-farm employment if they received indirect payments; however, their spouse's labor allocation does not depend upon direct payments, as the estimated coefficients are less significant. Crop insurance is another important factor in off labor allocation decision by operators and spouses. Research here shows that farmers who have crop insurance are less likely to work off the farm. The finding might revels that farmers are confident to receive expected benefit from crop insurance in case of crop loss. Finally, results from this study also indicate spouses of diversified farm are more likely to work off the farm. In addition, I found that if the farms are in metro area the spouse is less likely to work off the farm.

Parametric and semiparametric estimate are evaluated and specification tests were used to judge appropriateness of these model. The sign of coefficient for both parametric and semiparametric is quite similar except for few variables. To illustrate, I found that the sign of direct payment (*direct)* and crop insurance (*insur*) is positive in parametric model for spouses; however, the coefficients are negative in semiparametric model. Further the number of significant variable differ in parametric and semiparametric model. For example, direct and indirect payments are significant in parametric model; however, they are insignificant in semiparametric model for the operator. These results imply the existence of nonlinearity in off-farm labor supply model, and semiparametric model captures the nonlinearity of labor supply model. The prediction power of the off-farm labor supply model is less compare to the semiparametric model. Semiparametric model shows substantive effect on the inferences I make from this model. Finally, the evidence from a likelihood ratio test and Hong and White (1995) test show the importance of semiparametric model. Hence, the semiparametric model is a better modeling approach in the study of off-farm labor supply in the United States. This research introduces an alternative approach to study functional form of labor supply model, which might be the further step of my research.

REFERENCES

Ahearn, M. C., H. S. El-Osta, and J. Dewbre (2002), “The Impact of Government Subsidies on the Off-farm Labor Supply of Farm Operators” Selected paper for presentation at the 2002 AAEA meetings, Long Beach, California, July 28–31.

Ahearn, M. C., H. El-Osta, and J. Dewbre (2006), “The Impact of Coupled and Decoupled Government Subsidies on Off-farm Labour Participation of U.S. Farm Operators.” *American Journal of Agricultural Economics* 88: 393-408.

Ahearn, M., J.E. Perry, and H. S. El-Osta (1993), "The Economic Well-Being of Farm Operator Households, 1988-90." U.S Department of Agriculture, Economic Research Service, AER-666.

Barlett, P. F. (1986), "Part-time Farming: Saving the Farm or Saving the life-Style? “ *Rural Sociology* 51: 289-313.

Bickel, P., Y. Ritov, and T. M. Stoker (2006), "Tailor-Made Tests for Goodness of Fit to Semiparametric Hypothesis." *The Annals of Statistics* 34:721-741.

Blundell, R., A. Duncan, and K. Pendakur (1998), "Semiparametric Estimation and Consumer Demand." *Journal of Applied Econometrics* 13:435-461.

Blundell, R. and A. Duncan (1998), "Kernel Regression in Empirical Microeconomics." *The Journal of Human Resources* 33: 62-87.

Blundell, R. and J. L. Powell (2004), "Endogeneity in Semiparametric Binary Response Models." *The Review of Economic Studies* 71:655-679.

Bollman, R. D. (1979), "Off-farm Work by Farmers: An Application of the Kinked Demand Curve for Labor." *Canadian Journal of Agriculture Economics* 27:37-60.

Chang, H. H., and A. K. Mishra (2008), “Impact of Off-farm Labor Supply on Food Expenditures of the Farm Household.” *Food Policy* 33:657-664.

Chen, X. and Y. Fan (2006), "Estimation of Coupla-based Semiparametric Time Series Models." *Journal of Econometrics* 130:307-335.

Craven, P. and G. Wahba (1979), “Smoothing Noisy Data with Spline Function.” *Numerische Mathematik* 31:377-403.

Dewbre, J. and A.K. Mishra (2007), "Impact of Program Payments on Time Allocation and Farm Household Income." *Journal of Agricultural and Applied Economics* 39:489-505.

Dimitri, C., A. Effland, and N. Conklin (2005), "The 20th Century Transformation of U.S. Agriculture and Farm Policy." U.S. Department of Agriculture, Economic Research Service, Economic Information Bulletin, No. 3.

El-Osta, H. S., A. K. Mishra, and M. C. Ahearn (2004), "Labor Supply by Farm Operators Under "Decoupled" Farm Program Payments." *Review of Economics of the Household* 2:367-385.

El-Osta, H. S., A. K. Mishra, and M. J. Morehart (2008), "Off-Farm Labor Participation Decisions of Married Farm Couples and the Role of Government Payments." *Review of Agricultural Economics* 30:311-332.

El-Osta, H. S. and C. Ahearn (1996), "Estimating the Opportunity Cost of Unpaid Farm Labor for US Farm Operators." U. S. Department of Agriculture, Economic Research Survey Technical Bulletin No. 1848.

Fan, J. and I. Gijbels (1992), "Variable Bandwidth and Local Linear Regression Smoothers." *The Annals of Statistics* 20:2008-2036.

Fernandez-Cornejo, F. (2007), "Off-Farm Income, Technology Adoption, and Farm Economic Performance." U.S. Department of Agriculture, Economic Research Report Number 36.

Fernandez-Cornejo, J., C. Hendricks, and A. K. Mishra (2005), "Technology Adoption and Off-farm Household Income: The Case of Herbicide-Tolerant Soybeans." *Journal of Agricultural and Applied Economics* 37:549-563.

Fuguitt, G. V. (1959), "Part-Time Farming and the Push-Pull Hypothesis." *The American Journal of Sociology* 64:375-379.

Furtan, W. H., G. C. Van Kooten, and S. J. Thompson (1985), "The Estimation of Off-farm Supply Function in Saskatchewan." *Journal of Agricultural Economics* 36:211-220.

Gardner, B. L. (2002), *American Agriculture in the Twentieth Century: How it Flourished and What it Cost*. Harvard University Press, Cambridge.

Gebremedhin, T. G. and R. D. Christy (1996), "Structural Changes in U.S. Agriculture: Implications for Small Farms." *Journal of Agricultural and Applied Economics* 28:57-66.

Goodwin, B. K. and M. T. Holt (2002), "Parametric and Semiparametric Modeling of the Off-Farm Labor Supply of Agrarian Households in Transition Bulgaria." *American Journal of Agricultural Economics* 84:184-209.

Gould, B. W. and W. E. Saupe (1989), "Off-Farm Labor Market Entry and Exit." *American Journal of Agricultural Economics* 71:960-969.

Hastie, T. J. (1990), *Generalized Additive Models*. Chapman & Hall, New York.

Hardle, W., and E. Mammen (1993), "Comparing Nonparametric Versus Parametric Regression Fits." *Annals of Statistics* 21:1926-1947.

Hennessy, T. C. and T. Rehman (2008), "Assessing the Impact of the 'Decoupling' Reform of the Common Agricultural Policy on Irish Farmers' Off-farm Labour Market Participation Decisions." *Journal of Agricultural Economics* 59:41-56.

Hong, Y. and H. White (1995), "Consistent Specification Testing via Nonparametric Series Regression." *Econometrica* 63:1133-1995.

Hsiao, C., Q. Li, and J. Racine (2007), "A consistent Model Specification Test with Mixed Discrete and Continuous data." *Journal of Econometrics* 140: 802-826.

Huffman, W. E. (1980), "Farm and Off-Farm Work Decisions: The Role of Human Capital." *The Review of Economics and Statistics* 62:14-23.

Huffman, W. E., and M. D. Lange (1989), "Off-farm Work Decisions of Husbands and Wives: Joining Decisions Making." *Review of Economics and Statistics* 71:471-480.

Huffman, W. E. and H. El-Osta (1997), "Off-Farm Work Participation, Off-Farm Labor Supply and Off-Farm Labor Demand of U.S. Farm Operators." Selected paper for presentation at the 1998 AAEA meetings, Salt Lake City, Utah, August 2-5.

Keele, L. (2008), *Semiparametric Regression for the Social Sciences,* John Wiley and Sons.

Key, N., M. J. Roberts, and E. O'Donoghue (2006), "Risk and Farm Operator labor Supply." *Applied Economics* 38:573-586.

Key, N. and M .J. Roberts (2009), "Nonpecuniary Benefits to Farming: Implication for Supply Response to Decoupled Payments." *American Journal of Agricultural Economics* 91:1-18.

Kilkenny, M. (1993), "Rural/urban Effects of Terminating Farm Subsidies." *American Journal of Agricultural Economics* 75:968-980.

Kimhi, A. (1994), "Participation of farm owners in farm and off-farm work including the option of full-time off-farm work." *Journal of Agricultural Economics* 45:232-239.

Kimhi, A. (2004), "FamilyComposition and Off-Farm Participation Decisions in Israeli Farm Households." *American Journal of Agricultural Economics* 86:502-512.

Kimhi, A. and M. J. Lee (1996), "Off-Farm Work Decisions of Farm Couples: Estimating Structural Simultaneous Equations with Ordered Categorical Dependent Variables." *American Journal Of Agricultural Economics* 78:687-698.

Korb, P. (1999), "Choosing to Work Off-farm." *Rural Development Perspectives* 14: 44-49.

Kwon, C. W., P. F. Orazem, and D. M. Otto (2006), "Off-Farm Labor Supply Responses to Permanent and Transitory Farm Income." *Agricultural Economics* 34:59-67.

Larson D. W. and H. Yu Hu, (1977), “Factors Affecting the Supply of Off-Farm Labor among Small Farmers in Taiwan.” *American Journal of Agricultural Economics*, 59:549-553.

Lass, D. A. and C. M. Gempsesaw (1992), "The Supply of Off-Farm Labor: A Random Coefficients Approach." *American Journal of Agricultural Economics* 74:400-411.

Lass, D. A., J. L. Findeis, and M. C. Hallberg (1991), "Factors Affecting the Supply of Off-Farm Labor: A Review of the Empirical Evidence." In Multiple Job Holding Among Farm Families, M.C. Hallberg, J.L Findeis, and D.A. Lass eds. Iowa State University Press, Ames.

Lee, K. S. (2001), “Essays on Semiparametric Estimation.” PhD Dissertation, Louisiana State University, Baton Rouge.

Leistriz, F. L., H. G. Vreugdenhill, B. L. Ekstorm, and A. G. Leholm (1985), “Off-Farm Income and Employment of North Dakota Farm Families.” Agricultural Economics Miscellaneous Report No. 88. Agricultural Experiment Station, North Dakota State University, Fargo, ND.

Lewbel, A, and O. Linton (2002), "Nonparametric Censored and Truncated Regression.” *Econometrica* 70:765-779.

Li, Q. and J. S. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, New Jersey.

Li, Q. and S. Wang (1998), “A Simple Consistent Bootstrap Test for a Parametric Regression Function." *Journal of Econometrics* 87:145-165.

Manski, C. F. (1975), “The Maximum Score Estimation of the Stochastic Utility Model of Choice.” *Journal of Econometrics* 3:205-228.

Mishra, A. K.,and D. M. Holthausen (2002), "Effect of Farm Income and Off-Farm Wage Variability on Off-farm Labor Supply." *Agriculture and Resource Economic Review* 31:187-199.

Mishra, A. K. and C. L. Sandretto (2002), "Stability of Farm Income and the Role of Nonfarm Income in U.S. Agriculture." *Review of Agricultural Economics* 24:208-221.

Mishra, A. K. and H. S. El-Olsta (2001) "A Temporal Comparison of Source of Variability in Farm Household Income." *Agriculture Finance Review* 61:181-198.

Mishra, A.K. (1996), “Farm Income Variability and the Off-Farm Labor Supply of Farm Operators and Their Spouses.” PhD Dissertation, North Carolina State University.

Mishra, A.K., and B.K. Goodwin (1997), "Farm Income Variability and the Supply of Off-Farm Labor." *American Journal of Agricultural Economics* 79:880-887.

Mishra, A. K., H. S. El-Osta, M. J. Morehart, J. D. Johnson, and J. W. Hopkins(2002), “Income Wealth and the Economic Well being of Farm Households.” ERS Agricultural Economic Report No. AER812.

Poudel, B. N., K. P. Paudel, and K. Bhattarai (2009)," Searching for an Environmental Kuznets Curve in Carbon Dioxide Pollutant in Latin American Countries." *Journal of Agricultural and Applied Economics* 41:13-27.

Phimister, E. and D. Roberts (2006), “The effect of Off-farm Work on the Intensity of Agricultural Production.” *Environmental & Resource Economics* 34:493-515.

Racine, J. S. and Q. Li (2004), “Nonparametric Estimation of Regression Functions with both Categorical and Continuous Data.” *Journal of Econometrics* 119:99–130.

Roberts M. J. and N. Key (2008), “Agricultural Payments and Land Concentration: A Semiparametric Spatial Regression Analysis.” *American Journal of Agricultural Economics* 90:627-643.

Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression." *Econometrica* 56:931-954.

Sexton, R.N. (1975), Determinants of Multiple Job-holdings by Farm Operators. PhD Dissertation, North Carolina State University, Raleigh.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Simar, L. and P. W. Wilson (2007), "Estimation and Inference in Two Stage, Semiparametric models of Productive Efficiency." *Journal of Econometrics* 136:31–64.

Sumner, D. A. (1982), "The Off-Farm Labor Supply of Farmers." *American Journal of Agricultural Economics* 64:499-509.

Theil, H. (1971), *Principles of Econometrics*. John Wiley and Sons, Inc., New York.

Thompson, S. (1985) "A Model of Off-farm Employment." Forum No. 10. West Germany: Kieler Wissenschaftsverlag.

Wood, S. N. (2006), *Generalized Additive Models: an Introduction with R*. Chapman & Hall/CRC London.

Yang, D. T. (1997), "Education and Off-Farm Work." *Economic Development and Cultural Change* 45:613-32.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge.

Zheng, J. X. (1996), "A consistent Test of Functional From via Nonparametric Estimation Techniques", *Journal of Econometrics* 75:263-289.

APPENDIX 1

R CODE FOR DATA IMPORT AND VARIABLE CREATION

```
arm <- read.table("C:/Users/mahesh/Documents/Thesis/labor.dat", header=TRUE, sep="\t",
na.strings="NA", dec=".", strip.white=TRUE)
#House hold information are available in  VERSION 1 and HHCLS 1in arms data
arm.v1<-subset(arm,VERSION==1 & HHCLS==1)
attach(arm.v1)
#House hold net worth.
hhnw1<-ifelse(HHNW<0,0,HHNW)/1000
vprod1<-VPRODTOT/1000
#Government payment
govtpmt<-ifelse(IGOVT>0,1,0)
#Operator Education
op.educ<-
ifelse(OP_EDUC==1,10,ifelse(OP_EDUC==2,12,ifelse(OP_EDUC==3,14,ifelse(OP_EDUC==4,16,0))))
#Spouse Education
sp.educ<-
ifelse(SP_EDUC==1,10,ifelse(SP_EDUC==2,12,ifelse(SP_EDUC==3,14,ifelse(SP_EDUC==4,16,0))))
#operator age#
op.age<-OP_AGE
op.age.sq<-op.age^2
sp.age<-SP_AGE_V1
#Tenure class#
fowner<-ifelse(TENURCL2==1,1,0)
powner<-ifelse(TENURCL2==2,1,0)
tenant<-ifelse(TENURCL2==3,1,0)
#Farm Organization
indiv<-ifelse(P1201==1,1,0)
partner<-ifelse(P1201==2,1,0)
crop<-ifelse(P1201==3||P1201==4,1,0)
otheror<-ifelse(P1201==5,1,0)
#Health Insurance
ophthins<-ifelse(R1264>0,1,0)
sphthins<-ifelse(R1265>0,1,0)
#Work to off-farm jobs
opmiles<-ifelse(R941<=0,0,R941)
spmiles<-ifelse(R942<=0,0,R942)
#Government payment
crppayment<-P538+P539
crpacres<-ifelse(P28==0,0,crppayment/P28)
#Direct and Indirect Payment
direct<-IGOVDP
#direct<-P520*P525/100
```

```
indirect<-P522*P525/100+P528+P531+P536+P565
indpyacre<-ifelse(P20==0,0,indirect/P20)
dispayment<-ifelse(P20==0,0,P537/P200)
#Market value of land
landvalue<-P854
landacre<-ifelse(P20==0,0,P854/P20)
#variable for number of acree
noacres<-ifelse(P20==0,0,P20)
#P39 is acres in federal crop insurance
insurincome<-P552
inincacre<-ifelse(P39==0,0,P552/P39)
insurexp<-R732
#Net income from insurance
net.insureince<-(insurincome-insurexp)
net.inssq<-net.insureince*net.insureince
#Metro classification
nonmetro<-ifelse(ERS_FM==0,1,0)
farmcounty<-ifelse(ERS_FM==1,1,0)
metro<-ifelse(ERS_FM==8,1,0)
neometro<-ifelse(farmcounty==1|nonmetro==1,1,0)
metro1<-ifelse(metro==1&neometro==0,1,0)
#Government payment and crop insurance
gp<-ifelse(IGOVT>0,1,0)
insur<-ifelse(R732>0,1,0)
gpinsur<-gp*insur
ownedacre<-P20/P26
#Hours of working in non-farm sectors
hoffop<-P488+P489+P490+P491
hoffsp<-P588+P589+P590+P591
##Household Size
hhsize06<-ifelse(HH_SIZE06_V1<0,0,HH_SIZE06_V1)
hhsize13<-ifelse(HH_SIZE13_V1<0,0,HH_SIZE13_V1)+ifelse(HH_SIZE17_V1<0,0,HH_SIZE17_V1)
# Entropy
entropy<-ENTROPY
#Final Data
labor<-
na.exclude(data.frame(hoffop,hoffsp,op.age,sp.age,op.educ,sp.educ,ophthins,sphthins,hhsize06,hhsize13,
hhnw1,fowner,powner,vprod1,crppayment,direct,indirect,insur,entropy,metro1))
dim(labor)
#Save selected data in a file
write.table(labor,file="C://Users//mahesh//Desktop//Thesis I//Data//mydata.txt",col.names=TRUE)
```

APPENDIX 2

RCODE FOR DATA ANALYSIS (PROBIT MODEL)

```
la<-read.table(file="C://Users//mahesh//Desktop//Thesis I//Data//mydata.txt", header=T)
#model variable creation
la<-labor
hoffop<-la$hoffop
op.age<-la$op.age
op.educ<-la$op.educ
ophthins<-factor(la$ophthins)
hhsize06<-la$hhsize06
hhsize13<-la$hhsize13
hhnw1<-la$hhnw1
fowner<-factor(la$fowner)
powner<-factor(la$powner)
vprod1<-la$vprod1
crppayment<-la$crppayment
direct<-la$direct
indirect<-la$indirect
insur<-factor(la$insur)
entropy<-la$entropy
metro1<-factor(la$metro1)
hoffsp<-la$hoffsp
sp.age<-la$sp.age
sp.educ<-la$sp.educ
sphthins<-la$sphthins
#final data
k<-
data.frame(hoffop,hoffsp,op.age,sp.age,op.educ,sp.educ,ophthins,sphthins,hhsize06,hhsize13,hhnw1,fow
ner,powner,vprod1,crppayment,direct,indirect,insur,entropy,metro1)
#check for variable
str(k)
#######################################################
#Descriptive statistics
library(psych)
disc<-describe(k);disc
################################################
#Parametric estimation
library(AER)
########################################################
#probit model/Operator
########################################################
```

```r
#change off-farm labor supply in bivariate variable as
of_op<-ifelse(hoffop>0,1,0)
#Probit model
op_probit <- glm(of_op ~ op.age +I(op.age^2)+ op.educ + ophthins + hhsize06+hhsize13 +
hhnw1+fowner+powner+vprod1+crppayment+direct+indirect+insur+entropy+metro1,data = k, family =
binomial(link = "probit")); summary(op_probit)
#marginal effect of probit model(Average of sample marginal effect)
fav<-mean(dnorm(predict(op_probit,type="link")))
me_probit<-fav*coef(op_probit);me_probit
#McFadden's pseudo R^2
op_probit0<-update(op_probit,formula=.~1)
pseudo_Rsq<-1-as.vector(logLik(op_probit)/logLik(op_probit0));pseudo_Rsq
#Prediction
pred<-round(fitted(op_probit))
table(true=of_op,pred)
#Visualition
library("ROCR")
pred<-prediction(fitted(op_probit),+of_op)
plot(performance(pred,"acc"))
plot(performance(pred,"tpr","fpr"))
abline(0,1,lty=2)
############################################################
#probit model/Spouse
############################################################
#change off-farm labor supply in bivariate variable as
of_sp<-ifelse(hoffsp>0,1,0)
#Probit model
sp_probit <- glm(of_sp ~ sp.age +I(sp.age^2)+ sp.educ + sphthins + hhsize06+hhsize13 +
hhnw1+fowner+powner+vprod1+crppayment+direct+indirect+insur+entropy+metro1,data = k, family =
binomial(link = "probit"));summary(sp_probit)
#marginal effect of probit model(Average of sample marginal effect)
fav<-mean(dnorm(predict(sp_probit,type="link")))
me_probit<-fav*coef(sp_probit);me_probit
#McFadden's pseudo R^2
sp_probit0<-update(sp_probit,formula=.~1)
pseudo_Rsq<-1-as.vector(logLik(sp_probit)/logLik(sp_probit0));pseudo_Rsq
#Prediction
pred<-round(fitted(sp_probit))
table(true=of_sp,pred)
#Visualition
pred<-prediction(fitted(sp_probit),+of_sp)
plot(performance(pred,"acc"))
plot(performance(pred,"tpr","fpr"))
abline(0,1,lty=2)
```

APPENDIX 3

VARIABLE SELECTION PROCEDURE IN SEMIPARAMETRIC MODEL

```
#Using Blundell And Duncan  (1998) approach
library(np)
##Test for vprod1
#Introduce Instrument variable "hhnw1"
#model x=w*pi+uhat
lm.vprod1<-lm(vprod1~hhnw1)
uhat<-resid(lm.vprod1)
#plm
bw.vprod1<-npplregbw(hoffop~uhat|vprod1)
np_vprod1<-npplreg(bw.vprod1,residuals=TRUE)
t_test<-np_vprod1$xcoef/np_vprod1$xcoeferr
np_vprod1$xcoef
t_test
##Test for hhnw1
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.hhnw1<-lm(hhnw1~vprod1)
uhat.hhnw1<-resid(lm.hhnw1)
#plm
bw.hhnw1<-npplregbw(hoffop~uhat.hhnw1|hhnw1)
np_hhnw1<-npplreg(bw.hhnw1,residuals=TRUE)
t_hhnw1<-np_hhnw1$xcoef/np_hhnw1$xcoeferr
np_hhnw1$xcoef
t_hhnw1
##Test for hhnw1
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.hhnw1<-lm(hhnw1~vprod1)
uhat.hhnw1<-resid(lm.hhnw1)
bw.hhnw1<-npplregbw(hoffop~uhat.hhnw1|hhnw1)
np_hhnw1<-npplreg(bw.hhnw1,residuals=TRUE)
t_hhnw1<-np_hhnw1$xcoef/np_hhnw1$xcoeferr
np_hhnw1$xcoef
t_hhnw1
##Test for op.age
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.op.age<-lm(op.age~vprod1)
uhat.op.age<-resid(lm.op.age)
bw.op.age<-npplregbw(hoffop~uhat.op.age|op.age)
```

```
np_op.age<-npplreg(bw.op.age,residuals=TRUE)
t_op.age<-np_op.age$xcoef/np_op.age$xcoeferr
np_op.age$xcoef
t_op.age
##Test for opmiles
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.opmiles<-lm(opmiles~vprod1)
uhat.opmiles<-resid(lm.opmiles)
bw.opmiles<-npplregbw(hoffop~uhat.opmiles|opmiles)
np_opmiles<-npplreg(bw.opmiles,residuals=TRUE)
t_opmiles<-np_opmiles$xcoef/np_opmiles$xcoeferr
np_opmiles$xcoef
t_opmiles
##Test for direct
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.direct<-lm(direct~vprod1)
uhat.direct<-resid(lm.direct)
bw.direct<-npplregbw(hoffop~uhat.direct|direct)
np_direct<-npplreg(bw.direct,residuals=TRUE)
t_direct<-np_direct$xcoef/np_direct$xcoeferr
np_direct$xcoef
t_direct
##Test for indirect
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.indirect<-lm(indirect~vprod1)
uhat.indirect<-resid(lm.indirect)
bw.indirect<-npplregbw(hoffop~uhat.indirect|indirect)
np_indirect<-npplreg(bw.indirect,residuals=TRUE)
t_indirect<-np_indirect$xcoef/np_indirect$xcoeferr
np_indirect$xcoef
t_indirect
#Test for entropy
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.entropy<-lm(entropy~vprod1)
uhat.entropy<-resid(lm.entropy)
bw.entropy<-npplregbw(hoffop~uhat.entropy|entropy)
np_entropy<-npplreg(bw.entropy,residuals=TRUE)
t_entropy<-np_entropy$xcoef/np_entropy$xcoeferr
np_entropy$xcoef
t_entropy
```

```
##Test for hhsize06
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.hhsize06<-lm(hhsize06~vprod1)
uhat.hhsize06<-resid(lm.hhsize06)
bw.hhsize06<-npplregbw(hoffop~uhat.hhsize06|hhsize06)
np_hhsize06<-npplreg(bw.hhsize06,residuals=TRUE)
t_hhsize06<-np_hhsize06$xcoef/np_hhsize06$xcoeferr
np_hhsize06$xcoef
t_hhsize06
##Test for hhsize13
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.hhsize13<-lm(hhsize13~vprod1)
uhat.hhsize13<-resid(lm.hhsize13)
bw.hhsize13<-npplregbw(hoffop~uhat.hhsize13|hhsize13)
np_hhsize13<-npplreg(bw.hhsize13,residuals=TRUE)
t_hhsize13<-np_hhsize13$xcoef/np_hhsize13$xcoeferr
np_hhsize13$xcoef
t_hhsize13
##Test for op.educ
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.op.educ<-lm(op.educ~vprod1)
uhat.op.educ<-resid(lm.op.educ)
bw.op.educ<-npplregbw(hoffop~uhat.op.educ|op.educ)
np_op.educ<-npplreg(bw.op.educ,residuals=TRUE)
t_op.educ<-np_op.educ$xcoef/np_op.educ$xcoeferr
np_op.educ$xcoef
t_op.educ
##Test for crppayment
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.crppayment<-lm(crppayment~vprod1)
uhat.crppayment<-resid(lm.crppayment)
bw.crppayment<-npplregbw(hoffop~uhat.crppayment|crppayment)
np_crppayment<-npplreg(bw.crppayment,residuals=TRUE)
t_crppayment<-np_crppayment$xcoef/np_crppayment$xcoeferr
np_crppayment$xcoef
t_crppayment
##Test for sp.age
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.sp.age<-lm(sp.age~vprod1)
```

```
uhat.sp.age<-resid(lm.sp.age)
#plm
bw.sp.age<-npplregbw(hoffsp~uhat.sp.age|sp.age)
np_sp.age<-npplreg(bw.sp.age,residuals=TRUE)
t_sp.age<-np_sp.age$xcoef/np_sp.age$xcoeferr
np_sp.age$xcoef
t_sp.age
##Test for sp.educ
#Introduce Instrument variable "vprod1"
#model x=w*pi+uhat
lm.sp.educ<-lm(sp.educ~vprod1)
uhat.sp.educ<-resid(lm.sp.educ)
#plm
bw.sp.educ<-npplregbw(hoffsp~uhat.sp.educ|sp.educ)
np_sp.educ<-npplreg(bw.sp.educ,residuals=TRUE)
t_sp.educ<-np_sp.educ$xcoef/np_sp.educ$xcoeferr
np_sp.educ$xcoef
t_sp.educ
```

APPENDIX 4

R CODE FOR DATA ANAYSIS (GENERALIZED ADDITIVE MODEL)

#Semiparametric model/ generalized additive model/Operator

```
################################################################
library(mgcv)
#Semiparametric probit
op_add<-gam(of_op~op.age +I(op.age^2)+ op.educ + ophthins + hhsize06+hhsize13 +
hhnw1+fowner+powner+s(vprod1)+crppayment+direct+indirect+insur+entropy+metro1,omit.missing =
TRUE,method="GCV.Cp",family=binomial(link="probit"))
summary(op_add)
#marginal effect of probit model
fav_op<-mean(dnorm(predict(op_add,type="link")))
me_probit1<-fav_op*coef(op_add)
me_probit1
#Log likelihood value
logLik.gam(op_add)
#plot(op_add,residuals=TRUE,all.terms=TRUE,shade=TRUE,shade.col=2)
##Likelihood Ratio test for model comparision
anova(op_probit, op_add, test="Chisq")
################################################################
```

#Semiparametric model/ generalized additative model/Spouse

```
################################################################
#Semiparametric probit
sp_add<-gam(of_sp~sp.age +I(sp.age^2)+ sp.educ + sphthins + hhsize06+hhsize13 +
hhnw1+fowner+powner+s(vprod1)+crppayment+direct+indirect+insur+entropy+metro1,omit.missing =
TRUE,gcv=TRUE, method="GCV.Cp ,family=binomial(link="probit"))
summary(sp_add)
#marginal effect of probit model
fav_sp<-mean(dnorm(predict(sp_add,type="link")))
me_probit1<-fav_sp*coef(sp_add)
me_probit1
#Log likelihood value
logLik.gam(sp_add)
#plot(sp_add,residuals=TRUE,all.terms=TRUE,shade=TRUE,shade.col=2)
##Likelihood Ratio test for model comparision
anova(sp_probit, sp_add, test="Chisq")
####################################################
```

APPENDIX 5

MODEL COMPARISONS

#Specification test(Hong and White 1995)

```
##################################
#Operator
n<-nrow(k)
uhat_op<-resid(op_probit)
uhat_sp<-resid(sp_probit)
vhat_op<-resid(op_add)
vhat_sp<-resid(sp_add)
s_op<-var(uhat_op)
s_sp<-var(uhat_sp)
#Mn_op<-1/n*sum(uhat_op,vhat_op)
#Mn_op<-cov(uhat_op,vhat_op)
Mn_op<-1/length(uhat_op)*(sum(uhat_op^2)-sum(vhat_op^2))
Mn_sp<-1/length(uhat_sp)*(sum(uhat_sp^2)-sum(vhat_sp^2))
Pn_op<-length(op_probit$coef)
Pn_sp<-length(sp_probit$coef)
Tn_op<-(2*Pn_op)^(-1/2)*((n*Mn_op/s_op)-Pn_op)
Tn_sp<-(2*Pn_sp)^(-1/2)*((n*Mn_sp/s_sp)-Pn_sp)
Tn_op
Tn_sp
Pval <- function(z){abz<-abs(z);norm<-pnorm(abz);p<-2*(1-norm);p}
Pval(Tn_op)
Pval(Tn_sp)
##########################
#Plotting parametric fitting.
#predicted_op <-predict(op_probit,type="response")
#predicted_sp<-predict(op_probit,type="response")
#dim(predicted_op)
#plot(vprod1,predicted_op)
#Create New data
vp<-seq(1,length(vprod1),100)
#Average over all data
mean_op<-mean(data.frame(1,op.age,I(op.age^2), op.educ , ophthins , hhsize06,hhsize13,
hhnw1,fowner,powner,vprod1,crppayment,direct,indirect,insur,entropy,metro1))
mean_sp<-mean(data.frame(1,sp.age,I(sp.age^2), sp.educ , sphthins , hhsize06,hhsize13,
hhnw1,fowner,powner,vprod1,crppayment,direct,indirect,insur,entropy,metro1))
opcoef_op<-op_probit$coef
spcoef_sp<-sp_probit$coef
#
#Parametric fit for operator
```

```
fitted_op<-opcoef_op[1]+mean_op[2]*opcoef_op[2]+
mean_op[3]*opcoef_op[3]+mean_op[4]*opcoef_op[4]+
mean_op[5]*opcoef_op[5]+mean_op[6]*opcoef_op[6]+
mean_op[7]*opcoef_op[7]+mean_op[8]*opcoef_op[8]+
mean_op[9]*opcoef_op[9]+mean_op[10]*opcoef_op[10]+
vp*opcoef_op[11]+vp*opcoef_op[12]+
mean_op[13]*opcoef_op[13]+mean_op[14]*opcoef_op[14]+
mean_op[15]*opcoef_op[15]+mean_op[16]*opcoef_op[16]+mean_op[17]*opcoef_op[17]
fit_probit<-dnorm(fitted_op)
#plot(vp,fit_probit)
#
#Parametric fit for spouse
fitted_sp<-spcoef_sp[1]+mean_sp[2]*spcoef_sp[2]+
mean_sp[3]*spcoef_sp[3]+mean_sp[4]*spcoef_sp[4]+
mean_sp[5]*spcoef_sp[5]+mean_sp[6]*spcoef_sp[6]+
mean_sp[7]*spcoef_sp[7]+mean_sp[8]*spcoef_sp[8]+
mean_sp[9]*spcoef_sp[9]+mean_sp[10]*spcoef_sp[10]+
vp*spcoef_sp[11]+vp*spcoef_sp[12]+
mean_sp[13]*spcoef_sp[13]+mean_sp[14]*spcoef_sp[14]+
mean_sp[15]*spcoef_sp[15]+mean_sp[16]*spcoef_sp[16]+mean_sp[17]*spcoef_sp[17]
fit_probit_sp<-dnorm(fitted_sp)
########################
probit<-function(xb){dnorm(xb)}
newdata<-
data.frame(expand.grid(list(op.age=mean(op.age),op.agesq=mean(op.age^2),op.educ=mean(op.educ),oph
thins=mean(ophthins),hhsize06=mean(hhsize06),hhsize13=mean(hhsize13),hhnw1=mean(hhnw1),fowner
=mean(fowner),powner=mean(powner),vprod1=seq(1,length(vprod1),100),crppayment=mean(crppaymen
t),direct=mean(direct),indirect=mean(indirect),insur=mean(insur),entropy=mean(entropy),metro1=mean(
metro1))))
newdata_sp<-
data.frame(expand.grid(list(sp.age=mean(sp.age),sp.agesq=mean(sp.age^2),sp.educ=mean(sp.educ),sphth
ins=mean(sphthins),hhsize06=mean(hhsize06),hhsize13=mean(hhsize13),hhnw1=mean(hhnw1),fowner=
mean(fowner),powner=mean(powner),vprod1=seq(1,length(vprod1),100),crppayment=mean(crppayment)
,direct=mean(direct),indirect=mean(indirect),insur=mean(insur),entropy=mean(entropy),metro1=mean(m
etro1))))
#Transform into Probabilities
predict.fit <- predict.gam(op_add, newdata = newdata, se.fit=TRUE)
predict.fit_sp <- predict.gam(sp_add, newdata = newdata_sp, se.fit=TRUE)
mu.fit.op <- probit(predict.fit$fit)
mu.fit.sp <- probit(predict.fit_sp$fit)
###########################################
#Plot Operator
par(mfrow=c(1,1))
```

```
plot(vp,mu.fit.op,type="l",ylab="Predicted Probability of Off-farm Labor Supply", xlab="Vprod1",
bty="l")
title("Parmatric and Semiparametric Model: Operator",font.main= 6)
lines(vp,fit_probit,lty=2)
#Confidence Bands
lines(newdata$vp,probit(predict.fit$fit-2*predict.fit$se.fit), lty=3)
lines(newdata$vp, probit(predict.fit$fit+2*predict.fit$se.fit), lty=3)
legend(3000,0.40, c("Parametric", "Semiparametric", "Confidence Bands"),lty = c(2, 1,3), merge =
TRUE)
#Plot Spouse
plot(vp,mu.fit.sp,type="l",ylab="Predicted Probability of Off-farm labor Supply", xlab="Vprod1",
bty="l")
title("Parmatric and Semiparametric Model: Spouse") #font.main= 6)
lines(vp,fit_probit_sp,lty=2)
#Confidence Bands
lines(newdata_sp$vp,probit(predict.fit_sp$fit-2*predict.fit_sp$se.fit), lty=3)
lines(newdata_sp$vp, probit(predict.fit_sp$fit+2*predict.fit_sp$se.fit), lty=3)
legend(100,0.35, c("Parametric", "Semiparametric", "Confidence Bands"),lty = c(2, 1,3), merge = TRUE)
```

APPENDIX 6

STATA CODE FOR PARAMETRIC PROBIT MODEL

```
*Improt data
odbc load, dialog(complete) dsn("Excel Files") table("Sheet1$")
*generate dummy variable for operator
gen ofop=0
replace ofop =1 if   hoffop>0
gen ofsp=0
replace ofsp =1 if   hoffop>0
*generate age square variable
gen opagesq=opage^2
gen spagesq=spage^2
*Univariate probit model for operator
probit  ofop opage opagesq opeduc ophthins hhsize06 hhsize13 hhnw1 fowner powner vprod1
crppayment direct indirect insur entropy metro1, vce(robust)
*marginal effect
mfx
*Univariate probit model for spouse
probit  ofsp spage spagesq speduc sphthins hhsize06 hhsize13 hhnw1 fowner powner vprod1
crppayment direct indirect insur entropy metro1, vce(robust)

*Mariginal effect
mfx
```

VITA

Mahesh Pandit was born in Sundarbazar-2, Lamjung, Nepal. He received his high school education at Shree Adarsha Bal Secondary School, Sundarbazar-4, Lamjung, Nepal. The author graduated from Tribhuvan University, Nepal, where he received a Bachelor of Science in statistics in 2004 and Master of Science in statistics in 2007. In August 2008, he entered the master's program in the Department of Economics and Agribusiness at Louisiana State University. He is currently a candidate for the master's degree which will be awarded in August 2010.