



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Do Agricultural Subsidies Affect the Labor Allocation Decision? Comparing Parametric and Semiparametric Methods

Mahesh Pandit, Krishna P. Paudel, and Ashok K. Mishra

This study estimates off-farm labor supply from farm operators and their spouses using two different estimation procedures and data from the 2006 Agricultural Resource Management Survey. A semiparametric model was found to be better specified to study off-farm labor supply from operators and spouses than the parametric model. Contrary to previous findings, results found using the semiparametric model indicate that neither direct nor indirect government payments have any impact on the off-farm labor supply of farm operators. These findings indicate that existing literature may overstate the impact of farm payments on the economic well-being of farm households.

Key words: farm households, government payments, off-farm labor supply, semiparametric

Introduction

Proper model specification is an essential aspect of economic policy analysis, if economists want an accurate representation of the economic activity they are modeling. Historically speaking, most studies of off-farm labor supply have used parametric methods. During the 1970s and 1980s, ordinary least squares (OLS) was preferred by economists studying off-farm labor supply (Larson and Hu, 1977; Sumner, 1982; Gould and Saupe, 1989). Some researchers (Huffman, 1980; Gould and Saupe, 1989) began using a maximum likelihood estimation procedure through probit and logit models to study off-farm labor supply. Mishra and Goodwin (1997) used a Tobit model to study the effect of farm-income variability on off-farm labor supply among farm operators and their spouses. While these types of models were commonly used (El-Osta, Mishra, and Ahearn, 2004; Phimister and Roberts, 2006), researchers also began to use other models such as bivariate probit models (Ahearn, El-Osta, and Dewbre, 2006) and multinomial logit models (El-Osta, Mishra, and Morehart, 2008) to study off-farm labor allocation decisions. El-Osta and Ahearn (1996), Mishra and Goodwin (1997), and Ahearn, El-Osta, and Dewbre (2006) established an inverse relationship between government payments and off-farm labor supply using Tobit models.

Over the last ten years, government payments have accounted for nearly 30% of farm net income on average. During this period, the federal government has distributed an average of \$18.2 billion annually to farmers in the form of direct government payments (U.S. Department of Agriculture, 2010). These payments include direct payments for commodity programs, countercyclical payments, marketing loan benefits, emergency or disaster payments, tobacco transition payments, and conservation program payments (see Monke, 2004, for details on farm-commodity programs).

Mahesh Pandit is a Ph.D. student, Krishna P. Paudel is an associate professor, and Ashok K. Mishra is a professor in the Department of Agricultural Economics and Agribusiness at Louisiana State University and Louisiana State University Agricultural Center.

The authors thank the co-editor, Greg Galinato, and two anonymous reviewers for their comments, which helped to improve the quality of this journal article.

Review coordinated by Gregmar I. Galinato.

Our study is timely and relevant to current policy discussions, because the federal government has been trying to find ways to reduce unemployment. One way to reduce unemployment is to create incentives using fiscal policies. Recent studies using parametric analysis have shown that higher agricultural subsidies would reduce off-farm labor supply and thus reduce unemployment because quantity supplied for workers trying to find jobs in off-farm markets are lower. Findings from this research will help determine whether spending on agricultural subsidies (specifically, direct and indirect farm payments) reduces unemployment.

Multiple studies address the factors influencing a farm family's decision to participate in off-farm labor. Of greatest relevance to this article is the literature on how government payments affect the off-farm labor-supply decision. Using Agricultural Resource Management Survey (ARMS) data, Ahearn, El-Osta, and Dewbre (2006) and El-Osta, Mishra, and Morehart (2008) found that government payments tended to increase the number of hours operators work on the farm and decrease the hours devoted to off-farm labor, regardless of the payment type (coupled or decoupled). The study further found that government payments had a positive effect on the total number of hours worked. Ahearn, El-Osta, and Dewbre (2006) also showed that government payments have a negative effect on off-farm labor participation among farm operators and their spouses. Using data from Kansas farm households (more homogenous and local in nature), Mishra and Goodwin (1997) found that government payments were negatively related to off-farm labor participation.

The above studies, which used parametric methods to estimate their empirical models, have some known weaknesses. For example, a parametric method requires strong assumptions regarding the functional forms and is subject to misspecification, which may lead to poor results (Keele, 2008). Could the impact of government payments on off-farm labor supply be spurious as a result? Mishra et al. (2002) noted that government payments are skewed toward large farms and may be nonlinear in nature. Therefore, any conclusions based on a parametric method could lead to biased results and flawed policy design.

This study estimates off-farm labor supply among farm operators and their spouses using both parametric and semiparametric methods, test parametric versus semiparametric model specification,¹ and assesses the impact of government payments on labor allocation. We use a spline-based semiparametric model after identifying variables entering the model nonparametrically. Finally, we use the Hong and White (1995) test to identify appropriate model specification.

Conceptual Model

We assume that individuals allocate time to work on farm labor, off-farm labor, and leisure in such a way that the optimal allocation is achieved when the net marginal values of the time devoted to the activities are equal. In our case, the farm operator household is assumed to maximize utility:

$$(1a) \quad U = U(I, L^o, L^s, C^o, C^s, \tau),$$

subject to time constraints of the operator:

$$(1b) \quad T^o = L^o + y^o + F^o,$$

time constraints of the operator's spouse:

$$(1c) \quad T^s = L^s + y^s + F^s,$$

¹ While the semiparametric method is used frequently in statistics and general economics literature, it is used less frequently in agricultural economics research related to labor supply. Goodwin and Holt (2002) used a semiparametric (single-index) model to study farm-labor allocation in Bulgaria. They used both Hausman (1978) and Bera, Jarque, and Lee (1984) tests for their specification search. They found that there are no significant differences in the parameters estimated by the probit and the semiparametric single-index models.

the farm production function:

$$(1d) \quad Y_f = f(F^o, F^s, X_f, C^o, C^s, R),$$

and income:

$$(1e) \quad I = w^o y^o + w^s y^s + P_f Y_f - r_f X_f + V,$$

where superscripts o and s indicate “operator” and “spouse,” I is total income, L is time allocated to leisure, C is human capital, and τ denotes other factors such as life stage, number of children, farm tenure, and access to health insurance. In constraint equations, T is total time endowment, y is time allocated to off-farm work, F is time allocated to farm work, Y_f are farm outputs, X_f are inputs used in farm, R represents location and farm-specific factors (such as distance to city, diversification, and government farm program payments), w is the off-farm wage rate, P_f are farm output prices, r_f are farm input prices, and V signifies other household nonlabor income.

We also assume that the utility function and the production function are concave, continuous, and twice-differentiable (Mishra and Goodwin, 1997). The first-order conditions from the maximization problem provide many useful results, including the optimality conditions for off-farm labor supply. The optimal decision to work off farm (y^*) is obtained by substituting the optimal values of leisure and farm work hours derived from the first-order conditions from equations (1a)–(1e):

$$(1f) \quad y^* = T - L^* - F^* = f(\mathbf{w}, P_f, r_f, V, C, \tau, R, I),$$

where L^* represents optimal leisure and F^* represents optimal allocations of farm work hours.

Estimation Methods

A farm operator’s decision to work off farm can be expressed as a discrete choice model.² Let y denote the decision of a farm operator to work off farm, which is 1 if the farm operator decides to work off farm and 0 otherwise; X denotes the independent variables listed in equation (1f).

Parametric Method

A probit model is commonly used for the off-farm labor supply decision, which can be presented as:

$$(2) \quad \mathbf{y}^* = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \mathbf{y} = 1 \text{ if } \mathbf{y}^* > 0, \quad 0 \text{ otherwise;}$$

$$(2a) \quad E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0;$$

$$(2b) \quad \text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2;$$

where $\boldsymbol{\beta}$ represents the coefficients associated with the explanatory variables.

Let $\phi(\cdot)$ denote the probability density function of the normal distribution and $\Phi(\cdot)$ represent the cumulative distribution function of the normal distribution. The parameters are estimated by a maximum likelihood estimation procedure. The log likelihood function for a probit density function is represented as:

$$(3) \quad \ln(l_\beta) = \sum_{i=1}^n \{y_i \ln \Phi(\mathbf{X}'_i \boldsymbol{\beta}) + (1 - y_i) \ln(1 - \Phi(\mathbf{X}'_i \boldsymbol{\beta}))\}.$$

² Here, we present the case for farm operators, but one can easily substitute the same arguments for spouses.

The estimation of parameter β is equivalent to maximizing the log likelihood function with respect to the β parameter. The marginal effect of a continuous variable in a probit model is given by:

$$(4) \quad ME = \frac{\partial E[y|\mathbf{X}]}{\partial \mathbf{X}} = \phi(\mathbf{X}'\beta)\beta,$$

where the marginal effects are calculated by averaging across all observation. The marginal effects for a binary independent (dummy) variable are:

$$(5) \quad ME = \text{Prob}[y = 1|d = 1] - \text{Prob}[y = 1|d = 0].$$

We calculate robust standard errors (RSE) for all parameter estimated in the model.

Semiparametric Method

A semiparametric method can correct the weaknesses of the parametric and nonparametric methods because it balances the pros and cons of the parametric and nonparametric methods. One can think of semiparametric method as being a hybrid form of the parametric and nonparametric methods (Lee, 2001). The nonparametric components in a semiparametric method are distribution free, so a strong assumption of the functional form is not required. The semiparametric method also captures nonlinearity in the data. On the other hand, a semiparametric model reduces the number of variables entering nonparametrically and helps mitigate the problems associated with the curse of dimensionality. Using a semiparametric model typically avoids the problem of misspecification. Finally, the estimated parameters in the semiparametric model are asymptotically efficient.

The semiparametric regression model is often referred to as an additive or generalized additive model (GAM). A semiparametric model prevents overfitting (when relationships appear more statistically significant than they actually are) and provides the best mean squared error fit, as it is adjusted by penalty factors. A smoothing spline estimation procedure for a simple linear equation is provided in Appendix A.

We cannot use a linear equation directly on off-farm labor allocation decisions among farm operators and spouses because the decision variable is binary. We need to set up a semiparametric additive model with nonparametric and parametric terms as a Penalized Generalized Linear Model (PGLM), which takes the following form:

$$(6) \quad g\{E(y_i)\} = \mathbf{X}_i^* \beta + \sum_{j=1}^J f_j(z_{ji}),$$

where $y_i \sim$ follows a distribution from the exponential family and g is a known link function.³ Link function could be “probit,” “logit,” or any other common link functions. J is the number of variables entering the semiparametric model nonparametrically. As in the parametric model, the link functions play an important role in addressing the problems of the linear probability model (i.e., constant marginal effects and predicted probability exceeding beyond the range $[0, 1]$). Parametric model matrix \mathbf{X}_i^* also includes a column of one for intercept variable, and β is a parameter vector and f_j serves as a smoothing function for covariates z_{ji} .

In our case, y_i is a binary variable (1 = yes, 0 = no) for the off-farm labor allocation decision and \mathbf{X}_i^* denotes independent variables such as age, education, access to health insurance, number of children in the household, household net worth, farm ownership, farm size, government program payments (direct, indirect, and conservation reserve payments), crop insurance, entropy

³ A distribution belongs to the exponential family of distribution if the probability density or probability mass function can be written as $f_{\psi}(y) = \exp[\frac{\{y\psi - b(\psi)\}}{a(\phi)} + c(y, \phi)]$. Here b , a , and c are arbitrary functions, ϕ is an arbitrary ‘scale’ parameter, ψ is a canonical parameter of the distribution and depends on the model parameter in GLM.

Table 1. Definition and Summary Statistics (N=5,121)

| Variable | Definition | Mean | Std Dev | Min | Max |
|-------------------------------------|--|--------|---------|--------|---------|
| Dependent Variables | | | | | |
| <i>ofop</i> | =1 if the operator worked off farm, 0 otherwise | 0.307 | 0.461 | 0.000 | 1.000 |
| <i>ofsp</i> | =1 if the spouse worked off farm, 0 otherwise | 0.464 | 0.499 | 0.000 | 1.000 |
| Operator and Spouse Characteristics | | | | | |
| <i>opage</i> | Age of operator in years | 55.372 | 12.027 | 19.000 | 92.000 |
| <i>spage</i> | Age of spouse in years | 52.782 | 11.870 | 17.000 | 92.000 |
| <i>opeduc</i> | Years of formal education, operator | 13.459 | 1.913 | 10.000 | 16.000 |
| <i>speduc</i> | Years of formal education, spouse | 13.581 | 2.195 | 0.000 | 16.000 |
| <i>ophthins</i> | =1 if the farm operator received health insurance through off-farm work, 0 otherwise | 0.192 | 0.394 | 0.000 | 1.000 |
| <i>spthins</i> | =1 if the farm spouse received health insurance through off-farm work, 0 otherwise | 0.232 | 0.422 | 0.000 | 1.000 |
| Family Characteristics | | | | | |
| <i>hhsiz06</i> | Number of household members under the age of six | 0.151 | 0.501 | 0.000 | 6.000 |
| <i>hhsiz13</i> | Number of household members between ages six and seventeen | 0.545 | 0.997 | 0.000 | 7.000 |
| <i>hwnw1</i> | Household net worth (\$1000000) | 1.974 | 3.098 | 0.000 | 43.402 |
| Farm Characteristics | | | | | |
| <i>direct</i> | Direct farm program payments (\$1000) | 8.155 | 20.983 | 0.000 | 237.000 |
| <i>indirect</i> | Indirect farm program payments (\$1000) | 8.485 | 23.656 | 0.000 | 362.986 |
| <i>fowner</i> | = 1 if the farm is fully owned , 0 otherwise | 0.402 | 0.490 | 0.000 | 1.000 |
| <i>powner</i> | = 1 if the farm is partially owned, 0 otherwise | 0.490 | 0.500 | 0.000 | 1.000 |
| <i>crppayment</i> | Conservation reserve payments (\$1000) | 0.609 | 3.703 | 0.000 | 70.000 |
| <i>vprod1</i> | Farm size, value of agricultural output sold (\$1000000) | 0.710 | 1.630 | 0.000 | 27.000 |
| <i>insur</i> | = 1 if the farm has crop insurance, 0 otherwise | 0.317 | 0.465 | 0.000 | 1.000 |
| <i>entropy</i> | Entropy measure of farm diversification | 0.144 | 0.138 | 0.000 | 0.582 |
| Local Economic Condition | | | | | |
| <i>metro1</i> | = 1 if the farm is located in a metro county, 0 otherwise | 0.341 | 0.474 | 0.000 | 1.000 |

(index of diversification), and farm location. Table 1 includes summary statistics. The vector \mathbf{z}_j represents variables whose functional form cannot be specified. These variables enter the model nonparametrically.

In equation (6), the covariates \mathbf{X}^* are assumed to have a linear effect. The vector \mathbf{z}_j is nonlinear and fitted using 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 contain only continuous covariates.⁴ This model can be solved by using a penalized likelihood maximization procedure. The details of this procedure are available in Appendix B.

Variable Selection Procedure

Before estimating a model using a semiparametric method, it is essential to identify which variables should be entered in parametrically and which should be entered nonparametrically. Although a variable entering nonparametrically can be identified using established economic theories, these theories sometimes fail to appropriately place variables either parametrically or nonparametrically.

⁴ The kernel-based nonparametric model is available for dummy or multiple categorical independent variables (Racine and Li, 2004). The recently developed *crs* package in R can take both continuous and categorical variables even with the Spline method (Nie and Racine, 2012).

For this reason, variables must be categorized by the way in which they enter a model before the semiparametric model can be estimated.

Research into corresponding hypothesis tests is somewhat scant. We use a method suggested by Ruppert, Wand, and Carroll (2003, p. 168) to test the linearity of a variable entering the model nonparametrically. Consider the following models:

$$(7) \quad \begin{aligned} \text{Model 1: } y &= \alpha + \beta_1 x_1 + f_2(x_2) + \epsilon \\ \text{Model 2: } y &= \alpha + \beta_1 x_1 + \beta_2(x_2) + \epsilon \end{aligned}$$

where y is a dependent variable and x_1 and x_2 are independent variables. In Model 1, x_2 is entered nonparametrically and in Model 2 it is entered parametrically. Therefore, the testing hypothesis is:

H_0 : x_2 enters parametrically

H_1 : x_2 enters nonparametrically

The log likelihood ratio (LR) test or contrasting deviance statistic is then:⁵

$$(8) \quad LR = -2(\text{LogLikelihood}_0 - \text{LogLikelihood}_1),$$

where LogLikelihood_0 is the log likelihood for the restricted model (Model 2) and LogLikelihood_1 is the log likelihood for the unrestricted model (Model 1). The test statistics under the null hypothesis follow an approximate chi-square distribution, and the degrees of freedom equal the difference in the number of parameters across the two models. If observed LR is in the upper tail of its null distribution, then we conclude that the null hypothesis of linearity (parametric form) should be rejected (Ruppert, Wand, and Carroll, 2003).

Specification Test

Comparisons of the parametric and semiparametric results are another aspect of a semiparametric analysis. Hong and White (1995), Zheng (1996), Li and Wang (1998), and Hsiao, Li, and Racine (2007) provide some examples of specification tests. Hong and White (1995) introduced a consistent test of functional form using nonparametric techniques. The Hong and White test is based on the covariance between the residual from the parametric and discrepancy between the parametric and nonparametric fitted values, so it depends on model specification. The null hypothesis is that the parametric specification is correct against the semiparametric specification. The test statistics \hat{T}_n are given by:

$$(9) \quad \hat{T}_n = (n\tilde{m}_n/\hat{\sigma}_n^2 - P_n)/(2P_n)^{1/2};$$

$$(10) \quad \tilde{m}_n = n^{-1} \sum_{t=1}^n \hat{\epsilon}_{nt}^2 - n^{-1} \sum_{t=1}^n \hat{\eta}_{nt}^2;$$

where $\hat{\sigma}_n^2$ estimator for the variance of the error term under H_0 , P_n is dimension of parameter for parametric covariates, $\hat{\epsilon}_{nt}^2$ regression error from parametric estimation procedure, and $\hat{\eta}_{nt}$ is the residual from nonparametric estimation. Hong and White prove that the test statistics converge to a normal distribution under the correct specification but grow to infinity faster than the parametric rate under misspecification. That is, as $n \rightarrow \infty$, $T_n \xrightarrow{d} N(0, 1)$ under H_0 . The hypothesis H_0 is rejected for large values of T_n .

The likelihood ratio or contrasting deviance test can also be employed for model specification (that is, to compare the specification of parametric and semiparametric models as described in the previous section). The hypothesis can be tested as:

⁵ The deviance for a model is simply -2 times the log likelihood, so it also follows a chi-square distribution with the same degree of freedom of likelihood ratio test.

H_0 : Parametric Model

H_1 : Semiparametric Model

$$(11) \quad LR = -2(\text{LogLikelihood}_0 - \text{LogLikelihood}_1)$$

If the observed LR value falls within the upper tail of a chi-square distribution, then we conclude that the null hypothesis of the parametric model specification should be rejected.

Data

The empirical analysis uses 5,121 observations from the 2006 Agricultural Resource Management Survey (ARMS) collected by the United State Department of Agriculture/Economic Research Service. ARMS is a large national data set containing detailed information on the U.S. farm production sector, including (but not limited to) household labor activities, years of formal education, household health insurance status, family characteristics (i.e., the number of children present in the household), farm program payments, income and expenses, and farm type. We choose a set of variables from ARMS to represent the variables shown in equation (1f). We use education as a proxy measure for human capital (C) and wage (w). We use CRP, direct, and indirect payments as proxies for nonfarm income (V) because they are subsidies provided by the government to farmers. Age, number of children in the household, and health and crop insurance are used as a proxy measure for τ . Metro dummy, farm ownership (full owned or partially own), and entropy are proxy measures of R , which represents farm characteristics. Household net worth is used as a proxy measure for income, I . The value of production is used as a proxy measure for P_f and r_f .

Table 1 presents descriptive statistics of the variables used in the analysis. The data show that 31% of farm operators and 46% of their spouses work off farm. We are interested in the off-farm labor allocation decision, so we create a new dummy variable for both operators and spouses based on whether they supply labor to off-farm work. In our analysis, a value of 1 is assigned if the operator (spouse) works off farm and a 0 is assigned otherwise.

The literature addressing off-farm labor supply (Huffman, 1980; Mishra and Goodwin, 1997) suggests that off-farm work experience is an important factor affecting off-farm labor allocation. Unfortunately, the 2006 ARMS data do not contain any information on the number of years of off-farm work experience. Fringe benefits from off-farm employment, such as health insurance, may induce operators and spouses to work off farm. In our analysis, a value of 1 is assigned if the operator (spouse) receives health insurance from off-farm work and a value of 0 is assigned otherwise. The number of children in a household is divided into two categories: under the age of six and between ages six and seventeen. Higher levels of education provide better off-farm working opportunities, so this variable is included in the model as the number of years spent in a formal school setting. Following previous research, household net worth is used as a measure of the financial wealth of a household. Financially, well-established (measured by household net worth) farm operators may have less incentive to work off farm. Government payments also play an important role in farm operators' labor allocation decisions (Mishra and Goodwin, 1997; Dewbre and Mishra, 2007). Mishra and Sandretto (2002) point out that farm program payments stabilize total household income, thereby lessening the need to work off farm. Accordingly, we include information regarding different farm program payments such as direct, indirect, and conservation reserve payments in our analysis. Table 1 shows that for the year 2006, farms received an average \$8,155 in direct payments, \$8,485 in indirect payments, and \$609 in conservation reserve payments.

Farm size plays an important role in the labor allocation decision.⁶ Operators of small farms typically participate more in off-farm employment activities, work more hours off farm, and have a higher off-farm income than do operators of larger farms (Fernandez-Cornejo, Hendricks, and

⁶ One of the reviewers suspected that farm size and health insurance received from off-farm work might be endogenous. Our results show that there is no problem of endogeneity. Test results are available upon request from the authors.

Mishra, 2005). Mishra and Goodwin (1997) argue that operators, whose farm size is large, are less likely to work off farm because they must spend more time on farm work. We consider the value of agricultural output as a proxy for farm size. Farm operators who purchase crop insurance are less likely to work off farm because, a farmer receives indemnity payments in case of crop failure. These payments restore lost income while also reducing farm income variability. The type of county—metro or nonmetro—was included in the model to assess the impact of farm location on off-farm labor force participation among operators and spouses. Such locational variables have been included in earlier studies of the farm labor supply decision (El-Osta, Mishra, and Ahearn, 2004; Ahearn, El-Osta, and Dewbre, 2006; El-Osta, Mishra, and Morehart, 2008). We assume farms located in closer proximity to metro areas are more likely to have operators (spouses) who work off farm, given that it takes less travel time and offers more employment opportunities relative to a farm in a nonmetro area.

Results and Discussion

We first test the jointness in the decision making between farm operators and their spouses in the 2006 ARMS data and find it to be nonsignificant.⁷ Results from the copula test (Clayton copula functional form, dependence parameter = 1.6078; standard deviation = 1.0446) reject jointness in labor supply decisions. This allows us to estimate operators' and spouses' off-farm labor supply decision equations separately. Similar to our results, Mishra and Goodwin (1997); Lass, Findeis, and Hallberg (1989); Lass and Gempshaw (1992); Ahearn, El-Osta, and Dewbre (2006); and El-Osta, Mishra, and Morehart (2008) did not find any evidences of jointness in labor supply decisions.

The test statistics for categorizing whether a variable enters parametrically or nonparametrically are determined based on the likelihood ratio test described above. Test statistics are provided in table 2. We find different sets of variables entering nonparametrically in the semiparametric model for farm operator and spouse. For operators, deviance for age (*opage*) is significant, but the age-squared (*opagesq*) variable is not significant. This means that age squared captures nonlinearity of age and entered parametrically in the semiparametric model. Deviance for farm size (*vprod1*) is significant at a 5% level, an indication that farm size is a nonparametric covariate in operators' labor allocation decisions. For spouses, the deviance is significant for age (*spage*), household net worth (*hhnw1*), farm size (*vprod1*), direct payment (*direct*), indirect payment (*indirect*), and entropy (*entropy*). All variables entering nonparametrically are significant at the 5% level for both operators and spouses. These variables have significant effect on off-farm labor supply decision.

Table 3 provides information on coefficients and marginal effects related to operators' off-farm labor decisions under parametric and semiparametric models. Similarly, table 4 provides information on coefficients and marginal effects related to spouses' off-farm labor decisions under parametric and semiparametric models. The positive and significant coefficient on operator age (*opage*) and the negative and significant coefficient on operator age squared (*opagesq*) imply that age has an inverted-U-shape (quadratic) relationship with predicted probability of working off farm. Similar results hold for spouses. In the semiparametric model for spouses, the curve has a plateau then decreases as age increases, as shown in figure 2a. In particular, in the operator model, the marginal estimates from the parametric models imply that a unit change (additional year of age) decreases the probability of off-farm employment by 0.005, whereas it increases the probability of off-farm employment by 0.023 in the semiparametric model (table 3). The probability of an operator working off farm starts decreasing at age forty-three in the parametric model and forty-four in the semiparametric model. In the case of spouses, the probability of working off farm increases by 0.010 in the parametric model. The peak

⁷ This jointness test is based on a copula. Suppose $F(y_1, y_2) = C(F_1(y_1), F_2(y_2); \alpha)$ represents the joint distribution of farm-labor allocation decisions of farm operators (y_1) and their spouses (y_2). Here, α represents dependence between marginal distributions $F_1(y_1)$ and $F_2(y_2)$; $C(\cdot)$ is a copula function. If $\alpha = 0$, then the marginal distributions are independent. Details on the copula method used to test jointness can be found in Genest and Rémillard (2004) and Yan (2007).

Table 2. Variable Selection for Labor Allocation Model

| Variable | Operator | | | Spouse | | |
|------------------------|----------|----------|----------|--------|----------|----------|
| | DF | Deviance | P-value | DF | Deviance | P-value |
| <i>opage/spage</i> | 6.6395 | 72.9160 | 0.0000** | 3.0081 | 79.9710 | 0.0000** |
| <i>opagesq/spagesq</i> | 3.7450 | 4.5840 | 0.2987 | 1.5286 | 6.3780 | 0.0245** |
| <i>hhnw1</i> | 1.5267 | 2.1282 | 0.2469 | 5.4575 | 23.6130 | 0.0004** |
| <i>vprod1</i> | 3.8622 | 185.3800 | 0.0000** | 4.6466 | 88.5560 | 0.0000** |
| <i>crppayment</i> | 6.0814 | 8.9842 | 0.1804 | 0.5949 | 1.6675 | 0.1070 |
| <i>direct</i> | 3.7458 | 4.5842 | 0.2988 | 3.3025 | 12.2770 | 0.0086** |
| <i>indirect</i> | 3.0162 | 6.6453 | 0.1851 | 0.7567 | 3.3674 | 0.0451** |
| <i>entropy</i> | 5.7705 | 8.0165 | 0.2169 | 0.5678 | 958.2500 | 0.0000** |

Notes: Deviance is -2 times the difference between the log likelihood value from linear and nonlinear regression. In the spline-based regression model, penalties shrink degrees of freedom, so it could be noninteger. The effective degrees of freedom (DF) is $tr[(X'X + S)]^{-1}X'X$ (Ruppert, Wand, and Carroll, 2003; Wood, 2006). All programming was done using R 2.15.0 and package mgcv (see www.r-project.org; R, 2012). We use contrasting deviance or likelihood ratio test using code: `anova(mod.2, mod.1, test="Chisq")`. Double asterisks (**) indicate that the corresponding variables to those p-values are significant at the 5% level.

age for off-from employment among spouses is thirty-three in the parametric model. The findings support the life-cycle hypothesis in off-farm labor supply for both operators and their spouses.

The coefficient on educational attainment for both operator and spouses (*opeduc/speduc*) is positive and significant for both the parametric and semiparametric models. Our results confirm previous research that suggests both farm operators and their spouses with higher levels of education are more likely to work off farm (Mishra and Goodwin, 1997). In particular, the marginal effect for operators (table 3) indicates that an additional year of schooling increases the likelihood of off-farm work by 0.014 in the parametric model and 0.013 in the semiparametric model. The difference in the marginal effects is due to smoothing of the farm-size variable in the semiparametric model for operators. Similarly, the marginal effect for spouses reveals that an additional year of schooling increases the probability of off-farm work by 0.033 in both the parametric and semiparametric models (table 4). The likelihood of off-farm participation among spouses is nearly twice that of operators, *ceteris paribus*.

Often, nonfarm jobs provide fringe benefits such as access to health insurance, a benefit that is likely to attract farm operators and their spouses to off-farm employment. Our result shows that health insurance plays a positive and significant role in the off-farm labor allocation decision for both operators and spouses. For example, if an operator receives health insurance, then he or she has a 45% and 31% higher probability of working off farm in the parametric and semiparametric models, respectively. Again, the difference in the marginal effects is due to the smoothing of the farm-size variable in the semiparametric model. Consistent with the decision of farm operators, our results suggest that the probability of spouses working off farm when receiving health insurance from their off-farm job is 50% higher in the parametric model. As with operators, the probability of working off farm for spouses also declines (50% vs. 45%) when moving from the parametric model to the semiparametric model.

As expected, the coefficient on the number of children under age six (*hhsiz06*) for spouses is negative and significant for both models (table 4). The marginal effects imply that an additional child under the age of six decreases the spouses' probability of working off farm by 8% for both the parametric and semiparametric models. In the case of the number of children between age six and seventeen (*hhsiz13*), the coefficient is also negative and highly significant. Presence of children in the household limits the time available for off-farm work among spouses, especially for farm households, where women have traditionally devoted more time to caring for children. These results support the findings of Mishra and Goodwin (1997); Goodwin and Holt (2002); and El-Osta, Mishra, and Morehart (2008).

The coefficient on household net worth (*hhnw1*) reveals that farm operators and their spouses with higher net worth are less likely to work off farm, indicating an income effect. Table 3 shows that

Table 3. Parameter Estimates and Marginal Effects Parametric and Semiparametric Probit Model: Operator

| Variable | Parametric | | Semiparametric | |
|------------------------|--------------------------|--------------------------|------------------------------|--------------------------|
| | Coefficients | Marginal Effect | Coefficients | Marginal Effect |
| Parametric estimate | | | | |
| <i>opage</i> | 0.07641*** (0.01482) | −0.00504*** (0.00056) | 0.09355*** (0.01516) | 0.02305*** (0.00487) |
| <i>opagesq</i> | −0.00088 (0.00013) | | −0.00107*** (0.00014) | |
| <i>opeduc</i> | 0.05481*** (0.01113) | 0.01428*** (0.00290) | 0.05339*** (0.01144) | 0.01316*** (0.00369) |
| <i>sphthins</i> | 1.355161*** (0.05297) | 0.45154*** (0.01921) | 1.24717*** (0.05360) | 0.30734*** (0.01853) |
| <i>hhsz06</i> | 0.03416 (0.04484) | 0.0089 (0.01169) | 0.05458 (0.04843) | 0.01345 (0.01563) |
| <i>hhsz13</i> | −0.03611 (0.02200) | −0.0094 (0.00574) | −0.02219 (0.02329) | −0.00547 (0.00751) |
| <i>hhnw1</i> | −0.03184** (0.01233) | −0.00829** (0.00323) | −0.00329 (0.00903) | −0.00081 (0.00291) |
| <i>fowner</i> | 0.20705*** (0.07738) | 0.05476*** (0.02077) | 0.13649* (0.07957) | 0.03364 (0.02568) |
| <i>powner</i> | 0.10994 (0.07421) | 0.0285513 (0.01921) | 0.09247 (0.07606) | 0.02279 (0.02455) |
| <i>vprod1</i> | −0.22706*** (0.07778) | −0.05917*** (0.01984) | | |
| <i>crppayment</i> | 0.00629 (0.00556) | 0.00164 (0.00145) | 0.00574 (0.00555) | 0.00142 (0.00179) |
| <i>direct</i> | −0.00289* (0.00171) | −0.00075* (0.00044) | −0.00037 (0.00155) | −0.00009 (0.00050) |
| <i>indirect</i> | −0.00527** (0.00211) | −0.00137** (0.00055) | −0.00218 (0.00145) | −0.00054 (0.00047) |
| <i>insur</i> | −0.23932*** (0.05273) | −0.06179*** (0.01348) | −0.17335*** (0.05386) | −0.04271*** (0.01738) |
| <i>entropy</i> | −0.019803 (0.18092) | −0.00516 (0.04714) | −0.22179 (0.17397) | −0.05465 (0.05615) |
| <i>metro1</i> | −0.011387 (0.04512) | −0.00296 (0.01174) | 0.00184 (0.04600) | 0.00045 (0.01485) |
| Nonparametric estimate | | | | |
| <i>vprod1</i> | | | $df = 8.086, \chi^2 = 313.5$ | |

Notes: Hong and White's test statistic value is 41.3064, which is significant at the 1% level. The LR test statistic of the semiparametric model against parametric model is 235.01 with 6.5 degrees of freedom, which is significant at the 1% level. Single, double, and triple asterisks (*, **, ***) indicate significance at the 10%, 5%, 1% levels. Values in parenthesis are standard errors.

the coefficient on full owner (*fowner*) is positive and significant at the 1% level for operators in both parametric and semiparametric models, suggesting that full owners are more likely to work off farm compared to farms operated by tenants (table 3). The marginal effect (0.054) of full ownership in the parametric model suggests the probability of a full owner working off farm is 5.4% higher compared to tenants. The value-of-agricultural-production variable (*vprod1*), a proxy for farm size that entered nonparametrically, is negative and statistically significant at the 1% level for both operators and spouses (tables 3 and 4) in the parametric model. This result suggests that as farm size increases, the probability of operators and their spouses working off farm decreases, which is consistent with the

Table 4. Parameter Estimates and Marginal Effects Parametric and Semiparametric Probit Model: Spouse

| Variable | Parametric | | Semiparametric | |
|-------------------------|--------------------------|--------------------------|--------------------------------|--------------------------|
| | Coefficients | Marginal Effect | Coefficients | Marginal Effect |
| Parametric variables | | | | |
| <i>spage</i> | 0.07124*** (0.01676) | 0.01001*** (0.00057) | | |
| <i>spagesq</i> | −0.00106*** (0.00016) | | | |
| <i>speduc</i> | 0.12502*** (0.01046) | 0.03344*** (0.00266) | 0.12902*** (0.01126) | 0.03348*** (0.00448) |
| <i>spthins</i> | 1.73148*** (0.06216) | 0.50237*** (0.01347) | 1.74221*** (0.06410) | 0.45209*** (0.02581) |
| <i>hhsize06</i> | −0.3151*** (0.04705) | −0.08429*** (0.01247) | −0.3028*** (0.05089) | −0.07857*** (0.02005) |
| <i>hhsize13</i> | −0.06089*** (0.02170) | −0.01629*** (0.00579) | −0.06415** (0.02403) | −0.01665* (0.00940) |
| <i>hhnw1</i> | −0.03729*** (0.01081) | −0.00998*** (0.00287) | | |
| <i>fowner</i> | −0.02084 (0.07478) | −0.00558 (0.02001) | −0.02672 (0.07794) | −0.00693 (0.03092) |
| <i>powner</i> | −0.01208 (0.07044) | −0.00323 (0.01883) | 0.031 (0.07373) | 0.00804 (0.02925) |
| <i>vprod1</i> | −0.11215*** (0.02161) | −0.03000*** (0.00571) | | |
| <i>crppayment</i> | 0.00925* (0.00550) | 0.00247* (0.00147) | 0.00764 (0.00590) | 0.00198 (0.00235) |
| <i>direct</i> | −0.0008 (0.00122) | −0.00021 (0.00033) | | |
| <i>indirect</i> | −0.00481*** (0.00119) | −0.00129*** (0.00032) | | |
| <i>insur</i> | −0.00692 (0.05082) | −0.00185 (0.01359) | 0.03646 (0.05372) | 0.00946 (0.02124) |
| <i>entropy</i> | 0.42075** (0.16986) | 0.11255** (0.04536) | | |
| <i>metro1</i> | −0.07543* (0.04474) | −0.02017* (0.01194) | −0.04246 (0.04531) | −0.01102 (0.01803) |
| Nonparametric variables | | | | |
| <i>spage</i> | | | $df = 3.461, \chi^2 = 379.087$ | |
| <i>hhnw1</i> | | | $df = 4.099, \chi^2 = 23.197$ | |
| <i>vprod1</i> | | | $df = 8.476, \chi^2 = 124.822$ | |
| <i>direct</i> | | | $df = 3.139, \chi^2 = 11.557$ | |
| <i>indirect</i> | | | $df = 1.001, \chi^2 = 9.550$ | |
| <i>entropy</i> | | | $df = 1.001, \chi^2 = 2.842$ | |

Notes: Hong and White's test statistic value is 25.55, which is significant at the 1% level. The LR test statistic of the semiparametric model against the parametric model is 156.66 with 14.17 degrees of freedom, which is significant at the 1% level. Single, double, and triple asterisks (*, **, ***) indicate significance at the 10%, 5%, 1% levels. Values in parenthesis are standard errors.

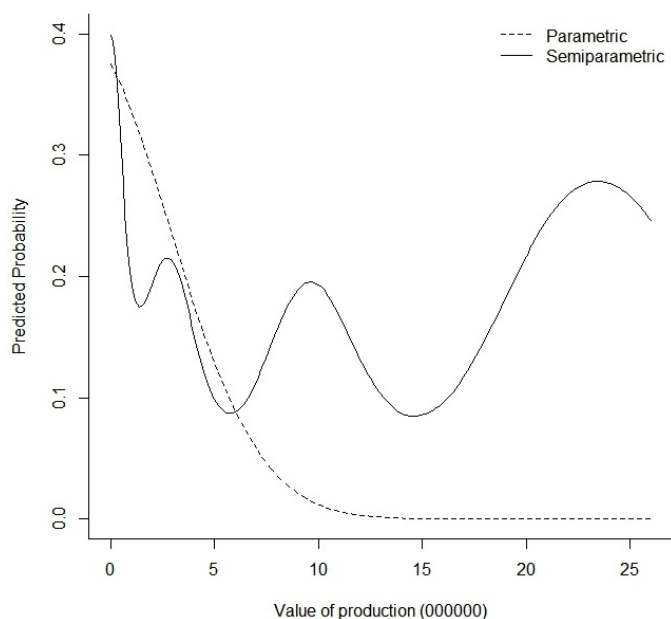


Figure 1. Parametric and Semiparametric Partial Regression Plots of the “Value of Production” Variable in the Operator Model

findings of Sumner (1982); Lass and Gempshaw (1992); Mishra and Holthausen (2002); and El-Osta, Mishra, and Ahearn (2004).

As noted earlier, the farm-size variable (*vprod1*) enters nonparametrically in the semiparametric model. The effect of farm size on the decision of off-farm labor supply among farm operators is shown in figure 1. The probability of off-farm work decreases as farm size increases up to \$6 million in production value for operator and spouse. We find no distinct pattern of relationship for higher values of production. In fact, one observes bumpy fitted curves for both operator and spouse if the value of production is higher than around \$6 million (number of observations represented by small number—1.54% of total observations or seventy-nine observations in total).⁸ Figure 2b shows that the probability of off-farm work decreases for farms exceeding \$20 million in production value in the semiparametric model describing spouses’ behavior.

We also find that the coefficient of conservation reserve payments (*crppayment*) is positive and significantly correlated with spouses’ off-farm labor supply. Spouses are more likely to seek off-farm employment as conservation payments increase. As expected, results for the parametric model show that operators who receive direct payments (*direct*) and indirect payments (*indirect*) are less likely to work off farm. This finding is consistent with findings from El-Osta, Mishra, and Ahearn (2004). For spouses, only the coefficient on indirect payments in the parametric model is significant, indicating an income effect. This finding is consistent with results from El-Osta, Mishra, and Ahearn (2004); Ahearn, El-Osta, and Dewbre (2006); and Dewbre and Mishra (2007). When examining the semiparametric model, results show that spouses are less likely to work off farm with an increase in both direct and indirect farm program payments (figures 2d and 2e). When using the semiparametric model, direct and indirect payments are no longer significant variables in explaining off-farm labor supply among farm operators.

⁸ Although it is possible to reduce the bumpy fitted curve of the variable *vprod1* by log transformation, for both the operator and spouse semiparametric regressions, we did not pursue this approach, as the transformation impacts the magnitude and significance of other parameters in the model. We thank an anonymous reviewer for pointing out the potential benefits of a log transformation to get smoother curves.

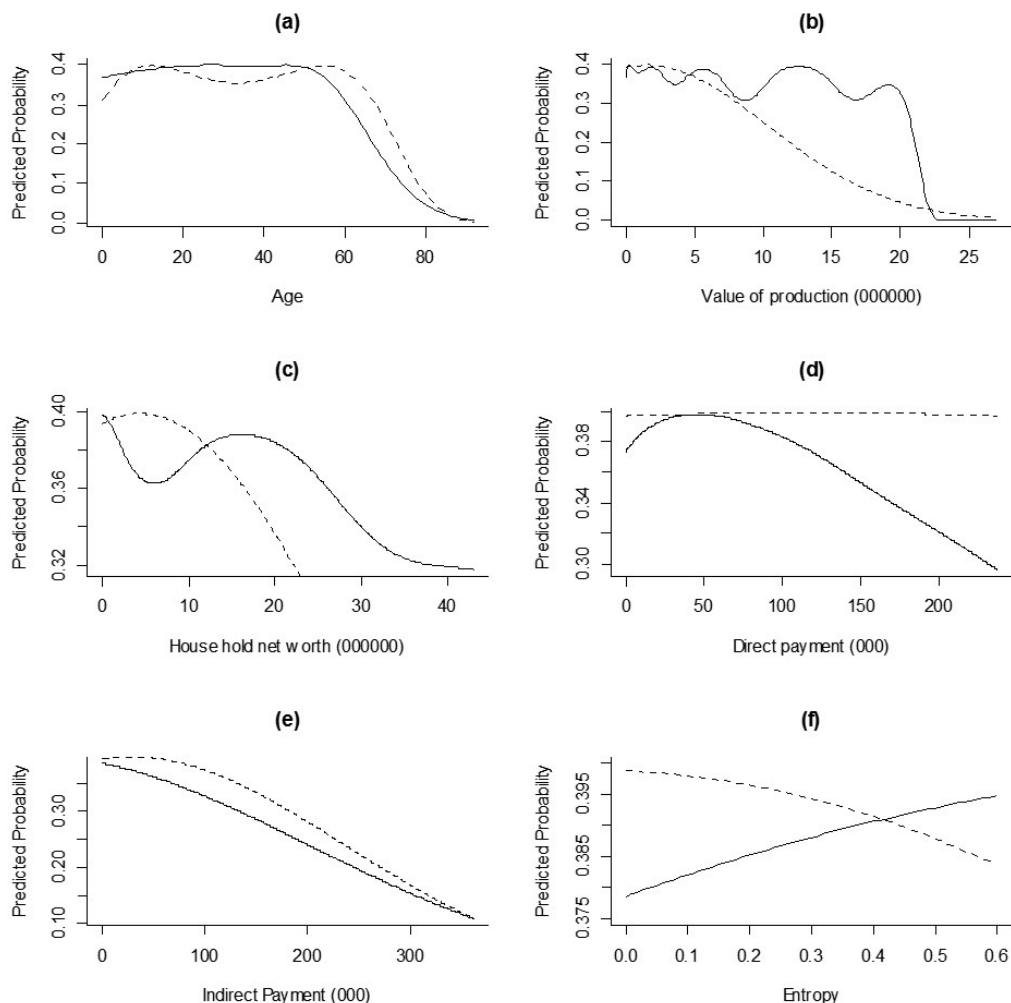


Figure 2. Parametric and Semiparametric Partial Regression Plots of Spouse Model Variables

Notes: Variables entering nonparametrically are (a) age, (b) value of production, (c) household net worth, (d) direct payments, (e) indirect payments, and (f) entropy.

To assess the impact of crop insurance (*insur*) on off-farm labor supply, we include a dummy variable (1 if the farm has crop insurance, 0 otherwise). The estimated coefficient on purchase of crop insurance is negative and significant at the 1% level for operators in both the parametric and semiparametric models (table 3). Results indicate that the probability of working off farm among farm operators who have purchased crop insurance is 6.1% and 4.2% lower in parametric models and semiparametric models, respectively, compared to operators without crop insurance. A possible explanation for this result is that farmers who buy crop insurance operate large farms that specialize in the production of program crops (e.g., corn, cotton, soybeans, wheat). Finally, we incorporate Theil's entropy index (*entropy*) to measure the impact of farm diversification on labor allocation. The coefficient of *entropy* is positive and significant at the 5% level for spouses (table 4). The marginal effect (0.11) of entropy suggests that as farms specialize, the probability of spouses working off farm increases by 0.11 (parametric model). This is supported by nonparametric estimate in the semiparametric model (figure 2f). Our result also indicates that spouses are less likely to work off farm if the farm is located in a metro county. The marginal effect is 0.02 for this variable.

The parametric probit specification is compared to a semiparametric specification using Hong and White's test for both operators and spouses (Hong and White, 1995). The estimated \hat{T}_n statistics and p-values are reported in notes at the bottom of tables 3 and 4. Results show that the semiparametric model is significant at a 1% level. Hence, one can conclude that the semiparametric model is a more appropriate estimation procedure to analyze off-farm labor supply than the parametric model. A likelihood ratio (LR) test was also performed to assess model specification. In particular, a semiparametric model is superior to a parametric model for operators, as indicated by likelihood ratio test ($df = 6.5$, $\chi^2 = 235.01$). The superiority of the semiparametric model also holds in the case of spouses' labor supply decisions ($df = 14.17$, $\chi^2 = 156.66$). Given the specification test results and figures 1 and 2(a–f), it is possible to say that a parametric model over/under predicts more than the semiparametric model. Our results support the need to use a semiparametric model when modeling off-farm labor supply decisions.

Conclusions

We estimate a parametric and spline-based semiparametric model of off-farm labor supply for farm operators and their spouses. Results from the parametric and semiparametric models were compared using the likelihood and Hong and White 1995 tests for model specification. Although our results show that more variables are significant in the parametric probit model than in the semiparametric additive probit model, the specification tests clearly indicate that the semiparametric model is better specified. Results indicate estimated parametric- and semiparametric-regression coefficients are different in terms of value and significance for both operator and spouse. These results imply the existence of nonlinearity in the off-farm labor supply model, caused by the value of farm production, age, household net worth, direct payments, indirect payments, and entropy, which can only be captured using a semiparametric model (figures 1, 2a–f).

Results from this study indicate that researchers need to be careful when modeling not only off-farm labor supply but also any dependent variable that could be influenced by both linear and nonlinear independent variables. Consequently, attention should be given to model specification. In particular, researchers should perform tests that categorize variables as entering a model parametrically or nonparametrically, which will aid in the selection of the appropriate estimation procedure. For example, in this study (and in contrast to previous findings), results indicate that the value of production (a proxy for farm size) entered the model nonparametrically. As a result, the model should be estimated using the semiparametric method.

Our analysis shows that direct and indirect government payments do not have an impact on operators' off-farm labor allocation decision in the semiparametric model. This research is important due to the looming budget deficits and the need for reduced government spending in coming decades. Policymakers should not increase government spending in the form of agricultural subsidies to reduce unemployment in the agricultural sector. Findings suggest that the impact of government policy on the labor allocation decision may not be what previous studies have found. The existing literature may be overstating the impact of farm payments on economic well-being of farm households. Without this new information, policymakers may believe a greater level of harm may be done through changes in farm program payments.

[Received June 2012; final revision received January 2013.]

References

- Ahearn, M. C., H. S. El-Osta, and J. Dewbre. "The Impact of Coupled and Decoupled Government Subsidies on Off-Farm Labor Participation of U.S. Farm Operators." *American Journal of Agricultural Economics* 88(2006):393–408.
- Bera, A. K., C. M. Jarque, and L. F. Lee. "Testing the Normality Assumption in Limited Dependent Variable Models." *International Economic Review* 25(1984):563–578.
- Craven, P., and G. Wahba. "Smoothing Noisy Data with Spline Function." *Numerische Mathematik* 31(1978):377–403.
- Dewbre, J., and A. K. Mishra. "Impact of Program Payments on Time Allocation and Farm Household Income." *Journal of Agricultural and Applied Economics* 39(2007):489–505.
- El-Osta, H. S., and M. C. Ahearn. "Estimating the Opportunity Cost of Unpaid Farm Labor for U.S. Farm Operators." Economic Research Service Technical Bulletin 1848, U.S. Department of Agriculture, 1996.
- El-Osta, H. S., A. K. Mishra, and M. C. Ahearn. "Labor Supply by Farm Operators under 'Decoupled' Farm Program Payments." *Review of Economics of the Household* 2(2004):367–385.
- El-Osta, H. S., A. K. Mishra, and M. J. Morehart. "Off-Farm Labor Participation Decisions of Married Farm Couples and the Role of Government Payments." *Review of Agricultural Economics* 30(2008):311–332.
- Fernandez-Cornejo, J., C. Hendricks, and A. K. Mishra. "Technology Adoption and Off-Farm Household Income: The Case of Herbicide-Tolerant Soybeans." *Journal of Agricultural and Applied Economics* 37(2005):549–563.
- Genest, C., and B. Rémillard. "Test of Independence and Randomness Based on the Empirical Copula Process." *Test* 13(2004):335–369.
- Goodwin, B. K., and M. T. Holt. "Parametric and Semiparametric Modeling of the Off-Farm Labor Supply of Agrarian Households in Transition Bulgaria." *American Journal of Agricultural Economics* 84(2002):184–209.
- Gould, B. W., and W. E. Saupe. "Off-Farm Labor Market Entry and Exit." *American Journal of Agricultural Economics* 71(1989):960–969.
- Hastie, T., and R. Tibshirani. *Generalized Additive Models*. London: Chapman and Hall, 1990.
- Hausman, J. A. "Specification Tests in Econometrics." *Econometrica* 46(1978):1251–1271.
- Hong, Y., and H. White. "Consistent Specification Testing Via Nonparametric Series Regression." *Econometrica* 63(1995):1133–1159.
- Hsiao, C., Q. Li, and J. S. Racine. "A Consistent Model Specification Test with Mixed Discrete and Continuous Data." *Journal of Econometrics* 140(2007):802–826.
- Huffman, W. E. "Farm and Off-Farm Work Decisions: The Role of Human Capital." *Review of Economics and Statistics* 62(1980):14–23.
- Keele, L. *Semiparametric Regression for the Social Sciences*. Chichester: Wiley, 2008.
- Larson, D. W., and H. Y. Hu. "Factors Affecting the Supply of Off-Farm Labor among Small Farmers in Taiwan." *American Journal of Agricultural Economics* 59(1977):549–553.
- Lass, D. A., J. L. Findeis, and M. C. Hallberg. "Off-Farm Employment Decisions By Massachusetts Farm Households." *Northeastern Journal of Agricultural and Resource Economics* 18(1989):149–159.
- Lass, D. A., and C. M. Gempshaw. "The Supply of Off-Farm Labor: A Random Coefficients Approach." *American Journal of Agricultural Economics* 74(1992):400–411.
- Lee, K.-S. *Essays on Semiparametric Estimation*. PhD dissertation, Louisiana State University Agricultural & Mechanical College, Department of Economics, Baton Rouge, LA, 2001.
- Li, Q., and S. Wang. "A Simple Consistent Bootstrap Test for a Parametric Regression Function." *Journal of Econometrics* 87(1998):145–165.

- Mishra, A. K., H. S. El-Osta, M. J. Morehart, J. D. Johnson, and J. W. Hopkins. "Income, Wealth, and the Economic Well-Being of Farm Households." Agricultural Economic Report 812, U.S. Department of Agriculture, Economic Research Service, Resource Economics Division, Washington, DC, 2002.
- Mishra, A. K., and B. K. Goodwin. "Farm Income Variability and the Supply of Off-Farm Labor." *American Journal of Agricultural Economics* 79(1997):880–887.
- Mishra, A. K., and D. M. Holthausen. "Effect of Farm Income and Off-Farm Wage Variability on Off-Farm Labor Supply." *Agricultural and Resource Economics Review* 31(2002):187–199.
- Mishra, A. K., and C. L. Sandretto. "Stability of Farm Income and the Role of Nonfarm Income in U.S. Agriculture." *Review of Agricultural Economics* 24(2002):208–221.
- Monke, J. "Farm Commodity Programs: Direct Payments, Counter-Cyclical Payments, and Marketing Loans." CRS Report for Congress RS21779, U.S. Department of Agriculture, Economic Research Service, Resources, Science, and Industry Division, Washington, DC, 2004. Available online at <http://www.nationalaglawcenter.org/assets/crs/RS21779.pdf>.
- Nie, Z., and J. S. Racine. "The crs Package: Nonparametric Regression Splines for Continuous and Categorical Predictors." *The R Journal* 12(2012):48–56.
- Phimister, E., and D. Roberts. "The Effect of Off-Farm Work on the Intensity of Agricultural Production." *Environmental and Resource Economics* 34(2006):493–515.
- Racine, J., and Q. Li. "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data." *Journal of Econometrics* 119(2004):99–130.
- Ruppert, D., M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge: Cambridge University Press, 2003.
- Sumner, D. A. "The Off-Farm Labor Supply of Farmers." *American Journal of Agricultural Economics* 64(1982):499–509.
- U.S. Department of Agriculture. 2010. Available online at <http://www.usda.gov/wps/portal/usda/usdahome>.
- Wood, S. N. *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: Chapman & Hall/CRC, 2006.
- Yan, J. "Enjoy the Joy of Copulas: With a Package Copula." *Journal of Statistical Software* 21(2007):1–21.
- Zheng, J. X. "A Consistent Test of Functional from via Nonparametric Estimation Techniques." *Journal of Econometrics* 75(1996):263–289.

Appendix A: Nonparametric Estimation Procedure

Let us consider a simple nonparametric model— $\mathbf{y} = f(\mathbf{z}) + \epsilon$ —where \mathbf{y} is a response variable, f is a smoothing function, and \mathbf{z} is a variable entering the model nonparametrically. The spline-smoothing method depends on minimizing the residual sum of squares (RSS) between a response variable \mathbf{y} and the nonparametric estimate $f(\mathbf{z})$. The RSS for a variable is given by:

$$(A1) \quad RSS(f) = \sum [\mathbf{y} - f(\mathbf{z})]^2.$$

The estimate of f that minimizes equation (A1) may use too many parameters, so the spline-smoothing method requires a penalization factor. Consequently, the minimization of RSS is subject to a penalty based on the number of local parameters used for spline smoothing (Keele, 2008). Suppose the penalty for a penalized regression spline method is $\lambda \int_{z_1}^{z_n} [f''(\mathbf{z})]^2 dz$ (Wood, 2006). This term is known as the roughness penalty constraint. The first term (λ) is the smoothing parameter, and the second term (integrated term) consists of the second derivative of $f(\mathbf{z})$, which measures the function's rate of change. Specifically, the second derivative measures the amount of curvature around the maximum of the likelihood function (Keele, 2008). We add a penalty term in equation (A1), so a spline estimate is given by the minimization of:

$$(A2) \quad RSS(f, \lambda) = \sum [\mathbf{y} - f(\mathbf{z})]^2 + \lambda \int_{z_1}^{z_n} [f''(\mathbf{z})]^2 dz,$$

where spline smoothing is used to minimize the sum of squares between \mathbf{y} and the nonparametric estimate. A very small value of λ gives overfitting close to the data and a large λ value produces a fit similar to the least-square method. To find an appropriate smoothing value that fits the semiparametric regression model, we select the smoothing parameter that minimizes an estimate of the expected mean square error. When the scale parameter of the distribution is known, the minimization of expected mean square error is equivalent to Mallows' C_p unbiased risk estimator (Craven and Wahba, 1978). For an unknown scale parameter, one would use the Generalized Cross Validation Score (GCVS) as suggested by Hastie and Tibshirani (1990). Wood (2006, pp. 172–174) provides a detailed explanation of the estimation procedure for the smoothing parameter.

Let $b_j(\mathbf{z})$ be the j^{th} basis function and γ_j be the local smoothing parameter, then the smoothing function f is represented as:

$$(A3) \quad f(\mathbf{z}) = \sum_i^q b_j(\mathbf{z}) \gamma_j.$$

Following Ruppert, Wand, and Carroll (2003) and Wood (2006, pp. 133–135), we can write the penalty in a matrix form as:

$$(A4) \quad \int_{z_1}^{z_n} [f''(\mathbf{z})]^2 dx = \mathbf{\Gamma}' \mathbf{S} \mathbf{\Gamma},$$

where $\mathbf{\Gamma} = (\gamma_1, \gamma_2, \dots, \gamma_q)$ is the smoothing parameter vector, $\mathbf{S} = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times q} \\ 0_{q \times 2} & 1_{q \times q} \end{bmatrix}$, with q denoting the number of knots. Equation (A2) can then be written in matrix form as:

$$(A5) \quad RSS(f, \lambda) = \|\mathbf{y} - \mathbf{Z} \mathbf{\Gamma}\|^2 + \mathbf{\Gamma}' \mathbf{S} \mathbf{\Gamma},$$

where $\mathbf{Z} = (b_1(\mathbf{z}), b_2(\mathbf{z}), \dots, b_q(\mathbf{z}))$. Ruppert, Wand, and Carroll (2003) and Wood (2006) have shown that the penalized least square estimator that minimizes equation (A5) is:

$$(A6) \quad \hat{\mathbf{\gamma}} = (\mathbf{Z}' \mathbf{Z} + \lambda \mathbf{S})^{-1} \mathbf{Z}' \mathbf{y}.$$

For a given value of λ and a set of basis functions, the prediction is given as:

$$(A7) \quad \hat{\mathbf{y}} = \mathbf{Z}(\mathbf{Z}' \mathbf{Z} + \lambda \mathbf{S})^{-1} \mathbf{Z}' \mathbf{y};$$

$$\hat{\mathbf{y}} = \mathbf{A} \mathbf{y};$$

where $\mathbf{A} = \mathbf{Z}(\mathbf{Z}' \mathbf{Z} + \lambda \mathbf{S})^{-1} \mathbf{Z}'$ is the hat matrix for the penalized spline. The prediction from the above equation equals the penalized spline prediction, which can be plotted to interpret the effects of \mathbf{z} on \mathbf{y} .

Appendix B: Procedure for Penalized GLMs

Let us consider the semiparametric model in equation (6) in terms of dependent variable \mathbf{y} . To estimate such a model, we are required to specify coefficients for the smooth and basis for each function f_j . Suppose $\mathbf{\Gamma}_j = (\gamma_{j1}, \gamma_{j2}, \dots, \gamma_{jq_j})'$ represents the vectors for the coefficient of the smooth term and $\mathbf{Z}_j = (b_{j1}(\mathbf{z}_j), b_{j2}(\mathbf{z}_j), \dots, b_{jq_j}(\mathbf{z}_j))$ is a set of basis functions chosen for j^{th} variables entering nonparametrically. The smoothness function can be represented in a matrix form as:

$$(B1) \quad f_j = \mathbf{Z}_j \mathbf{\Gamma}_j, \quad j = 1, 2, \dots, J.$$

Equation (6) can then be written as:

$$(B2) \quad g\{E(y_i)\} = \mathbf{X}_i \boldsymbol{\theta},$$

where $\mathbf{X} = [\mathbf{X}^* : \mathbf{Z}_1 : \mathbf{Z}_2 : \dots : \mathbf{Z}_J]$ and $\boldsymbol{\theta}' = [\boldsymbol{\beta}', \mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \dots, \mathbf{\Gamma}_J]$. Equation (B2) is similar to a GLM model with likelihood function $l(\boldsymbol{\theta})$. In general, the likelihood function can be expressed as an exponential family likelihood function:

$$(B3) \quad l(\boldsymbol{\theta}) = \sum_i^n \log[f_{\psi_i}(y_i)] = \sum_i^n \frac{\{y_i \psi_i - b_i(\psi_i)\}}{a_i(\phi)} + c_i(\phi, y_i),$$

where ψ_i depends on the GLM model parameters ($\boldsymbol{\theta}$).

If S is a penalty matrix, then the penalized likelihood function in equation (B2) takes the form of:

$$(B4) \quad l_p(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - \frac{1}{2} \sum_j \lambda_j \boldsymbol{\theta}' \mathbf{S}_j \boldsymbol{\theta} = l(\boldsymbol{\theta}) - \frac{1}{2} \boldsymbol{\theta}' \mathbf{S} \boldsymbol{\theta},$$

where $\mathbf{S} = \sum_{j=1}^J \lambda_j \mathbf{S}_j$, λ_j is a smoothing parameter that manipulates the tradeoff between the model's goodness of fit and smoothness, and \mathbf{S}_j is a matrix of known coefficients. Given the values of λ_j , the penalized likelihood function is maximized to find $\hat{\boldsymbol{\theta}}$. The value of λ_j is estimated using a cross-validation method (see Wood, 2006, p. 173) for details on the cross-validation method).