International Journal of Food and Agricultural Economics
ISSN 2147-8988, E-ISSN: 2149-3766
Vol. 7, No. 4, 2019, pp. 271-291

IMPLICATIONS OF MISSING-DATA IMPUTATION FOR SURVEY DATA: AN APPLICATION TO TECHNOLOGY ADOPTION

Haluk Gedikoglu
Department of Economics, Faculty of Social Sciences and Humanities
Konya Food and Agriculture University, Konya, Turkey
Email: haluk.gedikoglu@gidatarim.edu.tr

Joseph Parcell
Department of Agricultural and Applied Economics, Division of Applied Social Sciences, University of Missouri, Missouri, USA

Abstract

Missing-data is a problem that occurs frequently in survey data. Missing-data can result in bias and reduced efficiency for regression estimates. The objective of the current study is to analyze the impact of missing-data imputation, using multiple imputation methods, on regression estimates for survey data. The results of the current study show that multiple imputation methods result in lower standard errors for regression estimates than the regression using only complete observations. Multiple imputation methods also resulted in chances in magnitude, sign, and statistical significance for some the regression coefficient estimates. Hence, ignoring the missing-data problem might lead to significant differences in the results for regression analysis and the policy recommendations based on these results.
Keywords: *Missing-data, Multiple Imputation, Bayesian Inference, Household Surveys.*
JEL Codes: *Q12, C11, C81, C83*

1. Introduction

Missing observations are a problem that occurs frequently in survey data. Every dataset that involves primary data is prone to missing-data problem, even the datasets generated by government organizations, such as the Agricultural Resource Management Survey (ARMS), which is generated by the United States Department of Agriculture (USDA). Missing-data can result in bias and reduced efficiency of regression coefficient estimates (Horton & Kleinman, 2007; Gelman et al., 2004; Gelman & Rubin, 1992). Besides causing bias and reducing efficiency, the loss of a significant number of observations can also cause estimation difficulties (Gedikoglu, 2008). The standard regression procedure employed by economists is to use only complete observations, a process that is called listwise deletion (Schafer, 1997, Raghunathan et al., 2001). Listwise deletion can lead to the loss of a large number of observations. To illustrate this concept, in the example application of this article, which uses a typical agricultural household survey, listwise deletion caused a loss of 44 percent of the observations.

Over time, different methods have been used to handle missing-data; some of these methods are placing mean, predictive mean matching, single imputation, and multiple imputation (Little and Rubin, 2002). Placing mean and predictive mean matching are ad hoc methods that preserve the sample means, but bias the estimated variances and covariances toward zero (Schafer, 1997). Among the different methods, single and multiple imputation are model-based approaches (Little and Rubin, 2002; Reiter and Raghunathan, 2007). Single imputation treats imputed values as known values in the analysis; hence, it understates the variance of the coefficient estimates, which results in confidence intervals and significance tests that are too optimistic (Little and Rubin, 2002; Schafer, 1997). On the other hand, multiple imputation addresses this problem by conducting imputations multiple times and taking into account sampling variability due to the missing-data, which is called between-imputation variability (Schafer, 1997; Schafer & Graham, 2002).

Although statistical theory has been developed for missing-data imputation, the use of such methods has been relatively rare in agricultural economics (Robbins and White, 2011). Since data collection for analyzing socio-economic factors in agricultural economics is done through household surveys, which are conducted mostly by mail, agricultural household surveys are especially prone to the missing-data problem. Also, since most of the agricultural household surveys involve questions related to agricultural production or income, some farmers are hesitant to reveal information. For these reasons, the current study uses an agricultural household survey in the application part. However, the results from this study will also provide information beneficial for household surveys applied in other fields, too.

The majority of the existing studies (e.g., Robbins & White, 2011) measured the impact of imputation techniques on the distribution of univariate missing variables using arbitrarily created missing-data patterns. Among the few studies that analyzed missing-data imputation in agricultural economics, Robbins and White (2011), Ahearn et al. (2011), and Moss and Mishra (2011) use ARMS data for their analysis on missing-data imputation. These studies can be thought of as simulation studies, as they generate missing-data randomly from a complete dataset. Robbins and White (2011) analyze how the distribution of the variable "farm commodity payments received" changes when using two different imputation methods after some of the observations for this variable have been randomly removed from the data. One of the imputation methods used in their analysis is the method used by the USDA, which is the conditional mean imputation (a non-model-based estimation method). The second method is based on Data Augmentation (DA), which is a Markow Chain Monte Carlo (MCMC) method (Robbins & White, 2011). Robbins and White (2011) applied DA to conduct single imputation, rather than multiple imputation. Their results show that the method of imputation impacts the distribution of the variable imputed. As the ARMS dataset does not include missing observations, missing-data for the study by Robbins and White (2011) was created by the random removal of observations from the dataset, as will be discussed in the missing-data patterns and mechanisms section; this is referred to as Missing Completely at Random (MCAR). This is an important limitation of the Robbins and White (2011) research, because it is likely that there might be systematic reasons why data is missing. Hence, in general it is difficult to observe MCAR in actual survey data (Schafer, 1997).

The study by Ahern et al. (2011) provides a comparison of the USDA's conditional mean imputation method with sequential regression multivariate imputation (SRMI), using the ARMS dataset. SRMI is based on imputing missing observations for each variable separately, based on the observed distribution of each variable (Ahearn et al., 2011). Missing-data for the study by Ahearn et al. (2011) was also created by random removal of observations from the ARMS dataset. Their results show that the method used by the USDA causes the distribution of some variables to more closely match the full dataset than by using the SRMI method, while the opposite is true for some other variables. Hence, no definite general conclusion can be made about which imputation method is preferable. Although the studies by Robbins and White (2011) and Ahearn et al. (2011) analyze the impact of missing-data imputation on certain variables, neither of these studies analyzed the impact of missing-data imputation on regression coefficient estimates. Lastly, the study by Moss and Mishra (2011) applies Gibbs Sampling (a MCMC method), which is different than the multiple imputation method developed by Rubin (1987), to estimate a Leontief production function using the ARMS data, for which they also used synthetically generated missing-data. They find that results using imputed data and results using only complete observations do not differ significantly for regression coefficient estimates. However, missing-data imputation results in higher standard errors than using only complete observations, which is unexpected as multiple imputation should decrease the standard errors since it uses more observations. The authors conclude that this unexpected result is due to the collinearity problem caused by the arbitrarily created missing-data pattern.

One of the major limitations of the studies reviewed was generating missing-data randomly from a complete dataset. As will be further discussed in section three below, regression coefficient estimates are unbiased when the missing-data is generated randomly, though there will be a loss in efficiency. This is because the data, even with random missingness, can still represent the population. To our knowledge, no study has analyzed the implications of missing-data imputation on estimates of regression coefficients using an agricultural household survey (primary data), which includes missing-data that is not generated randomly. Such research can provide important insights into the consequences of using coefficient estimates that are based on using only complete observations to support private and public decision making. Another important limitation of the previous studies is that they analyze missing-data on univariate variables. In general, survey data can have multiple or even all variables with missing observations. Hence, the reviewed studies do not adequately address the complexities that surveys in fact face: having multiple variables with missing observations, having both discrete (either binary or ordinal) and continuous variables with missing observations, and having correlation among variables with missing observations.

Even if the statistical theory has been developed for handling the missing-data problem, the practical issues mentioned above have not been fully analyzed in the statistical literature (Schafer, 1997; Schafer & Graham, 2002). For example, an important issue is how to conduct multiple imputation when all the variables in the dataset have missing observations, as the previously conducted studies assume that only a portion of the variables have missing observations. When there are multiple variables to be imputed, including the case where all variables in the dataset have missing observations, two techniques can be applied: imputing all the variables simultaneously using the multivariate normal (MVN) multiple imputation or imputing each variable separately, based on the distribution of the variable, using univariate multiple imputation (Little and Rubin, 2002; Rubin, 1987; Schafer, 1997). MVN multiple imputation assumes that the variables to be imputed have a joint multivariate normal (continuous) distribution (Little and Rubin, 2002; Rubin, 1987; Schafer, 1997). The advantage of using MVN multiple imputation is that it takes into account the correlation structure among variables with missing observations. On the other hand, the advantage of using univariate multiple imputation is to conduct imputation for each variable with missing observations based on the specific distribution of that variable. Although these two methods are developed in the literature, there is no specific guidance as to which method should be used when the dataset have both discrete and continuous variables with missing observations. Analyzing these issues will be beneficial especially to practitioners who conduct multiple imputation on primary data.

The main objective of this paper is to evaluate the impact of multiple imputation on regression coefficient estimates using an agricultural household survey as primary data. This paper also analyzes the impacts of MVN multiple imputation and univariate multiple imputation on the regression coefficient estimates, when there are multiple variables with missing observations and with discrete and continuous distributions in the dataset. The multiple imputation approaches will be analyzed using data collected through an agricultural household survey in Iowa and Missouri in 2011 conducted through mail by the authors of this study (see Appendix A). This study distinguishes from the similar studies in the literature in different ways. This study is not a simulation study. Hence, this study does not arbitrarily create missing-data. Also, the current study uses a data set, where all the variables have missing part. For that reason, the current study provides a more realistic situation. Next, the data set for this study involves different type of variables that are commonly seen in survey data. Hence, this study includes all the complications that are seen in survey data related to missing-data problem. Lastly, this study provides a compassion of MVN multiple imputation and univariate multiple imputation.

The remainder of this paper is organized as follows: in the next section, we provide information on data and the regression model. Then, missing-data patterns and mechanisms are presented. Next, multiple imputation methods are introduced. Finally, results are delineated and conclusions posited.

2. Data and Regression Model

Promoting adoption of new technologies is an important policy issue in agricultural economics to increase the agricultural productivity and the farmers' income (Gedikoglu et al., 2011; Feder et al., 1985; Huffman, 1980, Just & Zilberman, 1988). Since the study by Grilliches (1957), farmers' adoption of new technologies has been analyzed. In the current study we analyze the socio-economic factors that impact adoption of Roundup Ready® corn by the farmers in the Midwest Region of the United States. Roundup Ready® corn is a seed technology that allows Roundup Ready® corn seeds to have resistance to the herbicide Roundup®. When farmers apply the herbicide Roundup® to the corn plant, it destroys the weeds around the corn plant, without destroying the plant (Couvillion et al., 2000). Roundup Ready® corn allows farmers to apply one herbicide instead of multiple herbicides. Hence, using Roundup Ready® corn decreases herbicide application costs (Couvillion et al., 2000).

The data for the current study was obtained through a mail survey of 2995 farm operations in Iowa and Missouri in the spring of 2011 conducted by the authors of this paper. Hence, the data used in this paper is a primary data. The questions for the survey were also designed by the author to discover if the farmers had adopted new technologies and how farmer and farm characteristics impacted their technology adoption decisions. The survey was sent out to a test group of one hundred farmers and was revised before developing the final survey instrument (see Appendix A). The final survey was sent out with a cover letter and a postage paid return envelope. A reminder postcard was sent after two weeks. The response rate for the survey was 21 percent. Before calculating the response rate, the farmers who had stopped farming, farmers who had returned the survey due to not being the farm operator, and undeliverable surveys to farmers (due to an address change) were subtracted from the original number of surveys that were sent out.

The technology adoption decision of a farmer can be analyzed using a random utility model (Greene, 2008; Freeman et al., 2014; Gedikoglu et al., 2011). The farmer compares the utility gained from adopting the technology U_a with the utility gained from not adopting the technology U_{na}. The farmers adopts the technology if U_a is bigger than U_{na}, otherwise the farmer does not adopt the technology. As researchers we can't observe the random utility for the farmer, but we can observe the technology adoption decision as:

$$y_i = 1 \text{ (technology is adopted)} \qquad \text{if } U_a > U_{na} \tag{1}$$

$$y_i = 0 \text{ (technology is not adopted)} \qquad \text{if } U_a \leq U_{na} \tag{2}$$

Following the literature on technology adoption, the random utility from technology adoption U(.) is a function of the farmer and farm characteristics; age (Chang and Boisvert, 2005), farm size, including farm sales, total land, and total number of animals (Feder et al., 1985; Huffman, 1980; Just and Zilberman, 1988), state in which the farm is located (Moreno & Sunding, 2005), perceptions of the farmer (Upadhay et al., 2002; Hua et al., 2004), hired labor and off-farm income (Cornejo et al., 2005; Tokle and Huffman, 1991), and education of the farmer (Khanna, 2001; Wozniak, 1984). Since adoption of a new technology involves uncertainty with respect to benefit and costs (Feder & O'Mara, 1982; Hiebert, 1974), information sources and institutions are also included in the random utility function of farmers. The random utility function also has a random component ϵ, which accounts for the factors that are not measurable by the researcher.

Based on the random utility specification of technology adoption, a binary response model can be used as the econometric model for the technology adoption decision (Greene, 2008; Freeman et al., 2014; Gedikoglu et al., 2011). In the current study, the logistic regression model is used as the binary response model of technology adoption. This model can be represented as (Greene, 2008):

$$\Pr(y_i = 1|X_i) = \exp(X_i\beta_i)/[1 + \exp(X_i\beta_i)],\ \ i = 1,..,N \tag{3}$$

where $y_i = 1$ if the farmer adopts Roundup Ready® corn and $y_i = 0$ if the farmer does not adopt Roundup Ready® corn. $\boldsymbol{\beta_i}$ is the vector of coefficients to be estimated and $\boldsymbol{X_i}$ is the vector of independent variables, which includes the variables listed above that are included in the random utility function.

3. Missing-data Patterns and Mechanisms

The missing-data pattern is an important component of multiple imputation, which impacts the choice of the multiple imputation method (Enders, 2010; Schafer, 1997; Gelman et al., 2004). The missing-data can occur in different patterns (Enders, 2010). Consider a 4 x 3 data matrix *X* with four observations and three variables. An indicator matrix *R* can be formed based on R_{ij} = 1, if the variable x_{ij} is observed (complete) for observation *i* and zero otherwise. Consider the following,

$$R_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} R_2 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

where R_1 is an example of an univariate missing-data pattern and R_2 is a general multivariate missing-data pattern (Enders, 2010). The univariate missing-data pattern R_1 can be imputed using the univariate multiple imputation methods, based on the distribution of the variable with missing observations. When the data shows the general missing-data pattern R_2, the MVN multiple imputation model can be applied if the variables have a continuous distribution (Rubin, 1987). However, Schafer (1997) showed that the MVN model can be applied even when some of the variables have discrete (either binary or ordinal) distribution. Another option is to use univariate multiple imputation for each variable separately, based on its distribution (Schafer, 1997). However, it is not known whether MVN or univariate multiple imputation performs better in this situation. The data used in the current study shows the general missing-data pattern, which will be used to implement both MVN and univariate multiple imputation methods.

A missing-data mechanism defines the probability distribution for the missing-data and the reason for having missing-data (Rubin, 1987). A missing-data mechanism is used in determining the consequence of ignoring missing-data in statistical analysis. There are three missing-data mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Rubin, 1987). MCAR is the case where the probability of missing-data for an observation does not depend on its own missing value or the realized values in the dataset. MAR is the case where the probability of missing-data for an observation depends on the realized values in the dataset, but not on its own missing value. Lastly, for MNAR, the probability of having missingness for an observation depends both on its own missing values and realized values in the dataset.

If the data is generated as MCAR, then the observed data is a random subsample of the full sample (Schafer, 1997). Hence, the observed data is still a representative of the full sample. In this case, no bias would result from ignoring missing observations (Schafer, 1997). However, there will still be a loss in the efficiency of the estimates, as not all the observations will be used. On the other hand, if MAR exists in the data-set, ignoring missing observations will cause biased and inefficient estimates (Schafer, 1997). In general MAR is a better assumption than MCAR for most survey data, as MAR is a more restrictive assumption. Lastly, if data is generated as MNAR, ignoring missing-data will also results in biased and inefficient estimates. The difference between MAR and MNAR for handling missing-data is that when MNAR exists, missing-data mechanism should be also modeled to obtain unbiased estimates, such as by using models similar to Heckman's selection model (Little and Rubin, 2002). Following Schafer (1997) and Graham et al. (1994), in the current study we have adopted MAR, as this study involves a multivariate data-set.

4. Multiple Imputation Methods

Multiple imputation methods consists of three steps: (1) the imputation step, (2) the completed-data analysis step, and (3) the pooling step (Rubin, 1987). During the imputation step, *M* imputations (completed datasets) are generated under the chosen imputation model. The econometric model is performed separately on each imputation $m= 1,2,\dots,M$ in the completed-data analysis step. In the current study, a univariate logistic regression model is used to represent adoption of Roundup Ready® corn. Lastly, during the pooling step, the results obtained from *M* completed-data analyses are combined into a single multiple imputation based estimation results. Below further description for each step of multiple imputation is provided.

4.1 Imputation Step

M imputations are generated under the chosen imputation model. The imputation model can be a univariate model or a multivariate model based on the number of variables to be imputed and the correlation among the variables. In the current study, both univariate and multivariate models are used to evaluate the differences. The data set for the current study involves three types of data: binary (discrete), ordinal (discrete), and continuous. MVN multiple imputation was developed for imputing multiple continuous variables simultaneously (Schafer, 1997; Lee & Carlin, 2010). Another option would be to use a logit-based univariate multiple imputation for binary variables, an ordered logit-based univariate multiple imputation for ordinal variables, and a linear regression-based univariate multiple imputation for continuous variables. The disadvantage of this process is ignoring the correlation among imputed variables (Schafer, 1997). We provide information first on MVN multiple imputation, then on univariate multiple imputation methods.

4.2 Multivariate Normal Multiple Regression

The basic multivariate normal regression model for imputing missing variables can be represented as follows. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_\mathrm{N}$ be a random sample from a p-variate normal distribution, where p represents the number of variables with missing values for observation $i = 1,\dots,N$. The multivariate normal regression can be represented as (Schafer, 1997):

$$\mathrm{x}_i = \Theta' \mathrm{z}_i + \epsilon_i, \; i = 1,\ldots,N \tag{5}$$

where $\mathrm{x}_i = (\mathrm{x}_{i1}, \mathrm{x}_{i2}, \ldots, \mathrm{x}_{ip})$ is the vector of values of the variables to be imputed for observation i, z_i is a q x 1 vector of values of the complete (independent) variables for observation i, Θ is a q x p matrix of regression coefficients, and ϵ_i is a p x 1 vector of random errors from a p-variate normal distribution with mean zero and a p x p variance-covariance (positive definite) matrix Σ. Θ and Σ are referred to as the model parameters (Little and Rubin, 2002).

4.3 Data Augmentation

The MVN multiple regression model uses data augmentation to impute missing values (Rubin, 1987). Data Augmentation consists of two steps, an I step (imputation step) and a P step (posterior step), that are performed at each iteration $t = 0,1,\ldots,T$ (Little and Rubin, 2002). Consider the partition of $\mathrm{x}_i = \left(x_{i(0)}, x_{i(m)}\right)$ corresponding to the observed and missing values of the imputation variables in observation i. At iteration t of the I step, the missing values in x_i are replaced with draws from the conditional posterior distribution of $\mathrm{x}_{i(m)}^{(t+1)}$ given observed data and the current values of model parameters $\Theta^{(t)}$ and $\Sigma^{(t)}$ independently for each observation (Little and Rubin, 2002). Following Little and Rubin (2002), in the current study, T was set as 100. Next, during the P step, new values of the model parameters $\Theta^{(t+1)}$ and $\Sigma^{(t+1)}$ were drawn from their conditional posterior distribution given observed data and data imputed in the previous I step $\mathrm{x}_{i(m)}^{(t+1)}$.

4.4 Expectation-Maximization Algorithm

Expectation-Maximization (EM) algorithm was used to obtain the initial values $\Theta^{(0)}$and $\Sigma^{(0)}$for the Data Augmentation above (Schafer, 1997). To maximize the log-likelihood function, the EM algorithm iterates the expectation step (E step) and maximization step (M step). The iteration between the E step and the M step continues until the difference between the successive values for all the parameters is less than a specified amount (in this article it is 0.00001) (Little and Rubin, 2002).

4.5 Univariate Multiple Imputation

The current dataset have three types of variables: binary, ordinal, and continuous. Hence, we will use three univariate multiple imputation methods: logit, ordered logit, and linear regression, for the corresponding variables, separately. Univariate multiple imputation methods address the specific distribution for each variable, but ignore the correlation among the variables.

4.6 Completed-data Analysis Step (Regression Analysis Step)

The logistic regression model specified in the data and regression model section for adoption of Roundup Ready® corn is performed separately on each set of imputed data (completed data) $m = 1,\ldots,M$ (Gelman et al., 2004).

4.7 Pooling Step

The logistic regression results obtained from M completed-data analyses were combined into a single multiple-imputation-based regression estimation result (Enders, 2010). Let

$\{(\widehat{\beta_i}) \colon \mathrm{i} = 1,2, \dots, M\}$ be the completed-data estimates of $\beta$from M imputed datasets (Enders, 2010). The multiple imputation estimate of β is (Enders 2010):

$$\bar{\beta}_M = \frac{1}{M} \sum_{i=1}^{M} \widehat{\beta_i} \tag{6}$$

5. Results

Table 1 provides information about the type of data, including whether the variable is binary, ordinal, or continuous. Also, both the number and percentage of missing observations are reported in table 1. All the variables in the current dataset had some missing observations (the percent of missing-data was nonzero for all variables). Hence, a multivariate missing-data pattern existed in the current data-set. For that reason, either the MVN imputation method or univariate multiple imputation methods could be applied to the current data. Overall, the percentage of missing observations was low for most of the variables, except for the spouse's education and the spouse's off-farm income. Even though the percentage of missing observations was low for most of the variables, the percentage of complete observations, which is used in most statistical programs for the regression analysis, was 56 percent. Hence, 44 percent of the observations would not have been used in the regression analysis if no imputation methods were used. Table 1 also shows that when the missing observations were imputed, the number of complete observations that could be used in the regression analysis became 472.

To see the impact of imputation methods on the distribution of the variables, the mean and the standard deviation for the imputed variables were compared between no imputation [m= 0], the 5th imputation [m = 5] and the 10th imputation [m = 10] for MVN and univariate multiple imputation methods. Table 2 reports the results for the mean and table 3 reports the results for the standard deviation. Overall, the multiple imputation methods did not cause a significant variation for either the mean or the standard deviation of the variables. The results varied for discrete variables only in the second or the third digit after the decimal for most of the variables. Even for the spouse's education variable, which had the highest percentage of missing observations at 25 percent, both multiple imputation methods provided a mean and standard deviation that were very close to the values calculated without imputation.

Table 4 provides the comparison of the logistic regression results between the no-imputation case and the MVN multiple-imputation with M set as 10. The hypothesis that all the regression coefficients, except the constant term, are zero was rejected for both regressions with p-values of 0.000. Hence, both the no-imputation and the MVN imputation regressions were statistically significant. For the individual variables in the regression, two of the variables that were not significant in the no-imputation case became significant at 10 percent significance level in the multiple-imputation case (i.e., age and land rented). Hence, with MVN imputation regression we can conclude that older farmers are more likely to adopt Roundup Ready® corn than younger farmers. This result is in line with the results of the studies by Chang and Boisvert (2005), Upadhyay et al. (2002), and Soule et al. (2000). The other conclusion that we reach with MVN imputation regression is farmers with more rented land are more likely to adopt Roundup Ready® corn than farmers with less rented land. Farmers with rented land could be more cost concerned, so they could be more likely to adopt Roundup Ready® corn (Couvillion et. al., 2000). This result is in line with results of the studies by Rahm and Huffman (1984) and Hua et al. (2004). Overall, MVN imputation regression results for age and land rented variables are supported with the results from previous studies.

Table 1. Variable Descriptions and Number of Missing and Imputed Observations

Variables	**Type**	**Description**	**Complete Obs.**	**Missing Obs.**	**Percentage Missing**	**Imputed Obs.**	**Total Obs.**
Roundup Ready® Corn	Binary	1 if adopted, 0 otherwise	453	19	4%	19	472
Age	Continuous	Age in years	460	12	3%	12	472
Owned Land	Continuous	Number of acres	466	6	1%	6	472
Land Rented Out	Continuous	Number of acres	461	11	2%	11	472
Land Rented	Continuous	Number of acres	459	13	3%	13	472
State	Binary	1 if located in Missouri, 0 if located in Iowa	467	5	1%	5	472
Farm Sales	Ordinal	1=$10,000- $99,999; 2 =$100,000-$249,999;3=$250,000 - $499,999; 4 =$500,000 and more	456	16	3%	16	472
Non-family Labor	Binary	1 if hired non-family labor, 0 otherwise	463	9	2%	9	472
Environmental Perceptions							
I am concerned about the water quality of streams and lakes in my county[1]	Ordinal	1=Strongly disagree, 2= Disagree, 3= Neither agree nor disagree,4=Agree, 5=Strongly agree	463	9	2%	9	472
I am concerned about the air quality in my county.	Ordinal		443	29	6%	29	472
I am concerned about the global warming.	Ordinal		463	9	2%	9	472
Sources of Information / Institutions							

How much influence does each of the following have on agricultural production decisions you make? Other Farmers[2]	Ordinal	1= None, 3=Some, 5=Very much	452	20	4%	20	472
Non-farming Neighbors	Ordinal		450	22	5%	22	472
Banks	Ordinal		449	23	5%	23	472
Contractors	Ordinal		449	23	5%	23	472
University/Extension	Ordinal		453	19	4%	19	472
USDA	Ordinal		447	25	5%	25	472
Other Government Org.	Ordinal		446	26	6%	26	472
Off-farm Income							
Farm Operator[3]	Ordinal	1=None, 2=$1-$9,999; 2 =$10,000-$24,999; 3=$25,000 -$49,999; 4 =$50,000-$99,999, 5=$100,000 and more	458	14	3%	14	472
Spouse	Ordinal		373	99	21%	99	472
Education							
Farm Operator[4]	Ordinal	1=Less than high school, 2=High school, 3= Some college or vocational school, 4=Bachelor degree, 5=Graduate degree	436	36	8%	36	472
Spouse	Ordinal		352	120	25%	120	472
Total Animal Units	Continuous	Total number of animals in animal units	462	10	2%	10	472

Note: [1, 2, 3] The description is the same for the other variables underneath.
Note: [4] The description is the same for the other variables underneath.

The university/extension variable is negative and statistically significant in no-imputation regression. Hence, the more a farmer's agricultural production decisions are impacted from university/extension, the less likely that the farmer adopts Roundup Ready® corn. This result is contrary to the results of the previous studies by Huffman (1980), Abdulai et al. (2002), and Husen et al. (2017). In the MVN imputed regression the university/extension variable is not found statistically significant. This result is in line with the results of the studies by Feder et al. (2004) and Gedikoglu and McCann (2010). Hence, if we did not use MVN imputed regression, we would infer farmers' whose agricultural production decisions influenced from university/extension services are less likely to adopt Roundup Ready® corn. Hence, policy recommendations could be altered based on using MVN multiple imputation or not.

Table 2. Comparison of the Mean between No-Imputation and Multiple Imputations

Variables Multiple	**No-Imputation**	**MVN Multiple Imputation**			**Univariate Imputation**		
	m = 0	**m = 5**	**m=10**		**m =5**	**m=10**	
Roundup Ready® Corn	0.466	0.464	0.466		0.453	0.455	
Age	53	53	53		53	53	
Owned Land	235	234	234		236	234	
Land Rented Out	20	20	20		21	19	
Land Rented	170	167	166		168	167	
State	0.490	0.489	0.489		0.493	0.493	
Farm Sales	3.171	3.167	3.172		3.160	3.169	
Non-family Labor	0.283	0.282	0.284		0.283	0.282	
Environmental Perceptions							
Water Quality	3.994	3.994	3.994		3.991	3.983	
Air Quality	4.115	4.104	4.113		4.097	4.116	
Global Warming	2.544	2.541	2.541		2.542	2.534	
Sources of Information / Institutions							
Other Farmers	2.573	2.571	2.574		2.560	2.586	
Non-farming Neighbors	1.718	1.716	1.716		1.714	1.714	
Banks	1.866	1.864	1.864		1.851	1.837	
Contractors	1.490	1.490	1.490		1.498	1.482	
University/Extension	2.210	2.210	2.210		2.215	2.197	
USDA	2.145	2.145	2.145		2.158	2.151	
Other Government Org.	1.794	1.794	1.794		1.812	1.798	
Off-farm Income							
Farm Operator	2.614	2.614	2.614		2.621	2.602	
Spouse	2.842	2.842	2.842		2.821	2.777	
Education							
Farm Operator	2.489	2.502	2.498		2.491	2.472	
Spouse	2.744	2.758	2.742		2.680	2.646	
Total Animal Units	187	187	187		189	185	

Table 3. Comparison of the Standard Deviation between No-Imputation and Multiple Imputations

Variables	No-Imputation	MVN Multiple Imputation			Univariate Multiple Imputation		
	m = 0	m = 5	m = 10		m = 5	m= 10	
Roundup Ready® Corn	0.499	0.499	0.501		0.500	0.498	
Age	11	11	11		11	11	
Owned Land	256	255	255		256	256	
Land Rented Out	102	102	102		102	102	
Land Rented	339	337	339		341	342	
State	0.500	0.503	0.502		0.500	0.500	
Farm Sales	1.502	1.498	1.489		1.498	1.498	
Non-family Labor	0.451	0.449	0.452		0.450	0.451	
Environmental Perceptions							
Water Quality	1.191	1.192	1.195		1.190	1.190	
Air Quality	1.080	1.077	1.083		1.098	1.078	
Global Warming	1.357	1.361	1.358		1.358	1.358	
Sources of Information / Institutions							
Other Farmers	1.164	1.166	1.167		1.162	1.160	
Non-farming Neighbors	1.020	1.040	1.024		1.019	1.019	
Banks	1.118	1.136	1.145		1.118	1.118	
Contractors	0.869	0.865	0.879		0.869	0.869	
University Extension	1.178	1.179	1.177		1.178	1.178	
USDA	1.192	1.201	1.206		1.192	1.192	
Other Government Org.	1.040	1.045	1.033		1.040	1.040	
Off-farm Income							
Farm Operator	1.158	1.180	1.162		1.158	1.158	
Spouse	1.248	1.260	1.268		1.248	1.250	
Education							
Farm Operator	1.608	1.643	1.620		1.609	1.613	
Spouse	1.453	1.540	1.472		1.461	1.454	
Total Animal Units	362	361	362		362	362	

The state variable is found negative and statistically significant both in no-imputation and MVN multiple imputation regressions. Hence, farmers in Missouri are less likely to adopt Roundup Ready® corn than farmers in Iowa. This could be due to Missouri being more of a cattle farming state, whereas Iowa being a major corn producing state and due to cost reduction aspect of Roundup Ready® corn (Couvillion et al.,2000). Farm sales variable is positive and statistically significant in both regression models. Hence, the higher the farm sales the more likely that a farmer adopts Roundup Ready® corn, which could be due to higher costs of adoption. This result is in line with results of the studies by Qaim and de Janvry (2003) and

Barham et al. (2004). Perceptions about global warming is also found statistically significant both in no-imputation and MVN multiple imputation regressions. The more a farmer is concerned about the global warming, the less likely that the farmers adopts Roundup Ready® corn. Since Roundup Ready® corn is mainly adopted for cost reduction, these farmers could be less environmentally concerned and focusing more on the profitability. This result is in line with the results of the studies by Upadhyay et al. (2002) and (Gedikoglu, 2008). Lastly, the USDA as a source of information is found positive and statistically significant in both regressions. Hence, the more the USDA has influence on a farmer's agricultural production decisions, the more likely that the farmer adopts Roundup Ready® corn. The USDA provides funding and technical support to farmers, so the farmers who use Roundup Ready® corn, who are expected to be profitability concerned, could be more actively searching for the USDA's funding and technical support (Gedikoglu, 2008).

Table 4. Comparison of Regression Results between No-Imputation and Multivariate Normal Multiple Imputation

Variables	**No-Imputation**			**Multivariate Normal Multiple Imputation**				
	Coeff.	**S.E.**	**p-Value**	**Coeff.**	**S.E.**	**p-Value**	**Coeff.[1] (%)**	**S.E.[2] (%)**
Age	0.008	0.010	0.390	0.023	0.006	0.000	188%	-40%
Owned Land	-0.001	0.000	0.182	0.000	0.000	0.349	-100%	0%
Land Rented Out	0.001	0.001	0.451	0.000	0.001	0.820	-100%	0%
Land Rented	0.001	0.000	0.424	0.001	0.000	0.088	0%	0%
State	-0.995	0.207	0.000	-0.640	0.148	0.000	36%	-29%
Farm Sales	0.366	0.108	0.001	0.323	0.073	0.000	-12%	-32%
Non-family Labor	-0.041	0.222	0.854	-0.154	0.165	0.351	-276%	-26%
Environmental Perceptions								
Water Quality	0.052	0.104	0.614	-0.052	0.069	0.457	-200%	-34%
Air Quality	-0.093	0.105	0.377	0.022	0.080	0.786	124%	-24%
Global Warming	-0.198	0.081	0.014	-0.126	0.056	0.023	36%	-31%
Sources of Information / Institutions								
Other Farmers	0.086	0.096	0.373	-0.077	0.070	0.271	-190%	-27%
Non-farming Neighbors	0.054	0.105	0.610	0.065	0.085	0.446	20%	-19%
Banks	0.094	0.107	0.382	0.114	0.083	0.168	21%	-22%
Contractors	0.088	0.143	0.537	-0.088	0.104	0.394	-200%	-27%
University/Extension	-0.205	0.119	0.085	-0.048	0.081	0.552	77%	-32%
USDA	0.237	0.118	0.044	0.247	0.092	0.008	4%	-22%
Other Government Org.	0.179	0.119	0.131	0.054	0.089	0.545	-70%	-25%
Off-farm Income								
Farm Operator	0.038	0.114	0.739	0.008	0.081	0.920	-79%	-29%
Spouse	0.001	0.110	0.994	-0.018	0.084	0.826	-1900%	-24%
Education								
Farm Operator	0.004	0.078	0.956	-0.038	0.055	0.485	-1050%	-29%
Spouse	0.074	0.083	0.378	0.021	0.069	0.763	-72%	-17%
Total Animal Units	0.000	0.000	0.715	0.000	0.000	0.683	0%	0%
Constant	-1.910	0.877	0.029	-1.945	0.608	0.001	-2%	-31%

Note: 1,2 shows the percentage difference in coefficient estimates and standard errors between multivariate normal multiple imputation and no-imputation, respectively

Table 5. Comparison of Regression Results between No-Imputation and Univariate Multiple Imputation

Variables	No-Imputation			Univariate Multiple Imputation				
	Coeff.	S.E.	p-Value	Coeff. (%)	S.E.	p-Value	Coeff.[1] (%)	S.E.[2]
Age	0.008	0.010	0.390	0.010	0.008	0.233	25%	-20%
Owned Land	-0.001	0.000	0.182	-0.001	0.000	0.195	0%	0%
Land Rented Out	0.001	0.001	0.451	0.002	0.001	0.211	100%	0%
Land Rented	0.000	0.000	0.424	0.000	0.000	0.123	0%	0%
State	-0.995	0.207	0.000	-0.870	0.181	0.000	13%	-13%
Farm Sales	0.366	0.108	0.001	0.303	0.090	0.001	-17%	-17%
Non-family Labor	-0.041	0.222	0.854	0.012	0.192	0.950	129%	-14%
Environmental Perceptions								
Water Quality	0.052	0.104	0.614	0.059	0.089	0.506	13%	-14%
Air Quality	-0.093	0.105	0.377	-0.034	0.095	0.717	63%	-10%
Global Warming	-0.198	0.081	0.014	-0.176	0.068	0.010	11%	-16%
Sources of Information / Institutions								
Other Farmers	0.086	0.096	0.373	0.020	0.086	0.812	-77%	-10%
Non-farming Neighbors	0.054	0.105	0.610	0.051	0.096	0.598	-6%	-9%
Banks	0.094	0.107	0.382	0.149	0.093	0.111	59%	-13%
Contractors	0.088	0.143	0.537	-0.010	0.125	0.934	-111%	-13%
University/Extension	-0.205	0.119	0.085	-0.079	0.095	0.409	61%	-20%
USDA	0.237	0.118	0.044	0.232	0.103	0.024	-2%	-13%
Other Government Org.	0.179	0.119	0.131	0.091	0.102	0.372	-49%	-14%
Off-farm Income								
Farm Operator	0.038	0.114	0.739	0.062	0.097	0.523	63%	-15%
Spouse	0.001	0.110	0.994	-0.013	0.095	0.889	-1400%	-14%
Education								
Farm Operator	0.004	0.078	0.956	-0.023	0.067	0.734	-675%	-14%
Spouse	0.074	0.083	0.378	0.051	0.074	0.493	-31%	-11%
Total Animal Units	0.000	0.000	0.715	0.000	0.000	0.884	0%	0%
Constant	-1.910	0.877	0.029	-2.009	0.739	0.007	5%	-16%

Note:[1] Shows the percentage difference in coefficient estimates between univariate multiple imputation and no-imputation.
[2] Shows the percentage difference in standard errors between univariate multiple imputation and no-imputation.

In terms of the magnitude of the coefficient estimates, the MVN multiple imputation caused changes in the magnitude for most of the variables (e.g., 21 variables out of 23). MVN imputation caused a decrease in the coefficient estimates of 13 of the variables, while the coefficient estimates of 8 variables increased in comparison to regression with no-imputation. It is important to see that the majority of the coefficient estimates (e.g., 19 variables out of 23) had *lower* standard errors in the MVN imputed regression than in the no-imputation regression. The other four coefficient estimates had no change in the standard error. Hence, the MVN imputation significantly increased the efficiency of the estimates. Overall, MVN multiple imputation could alter the regression coefficient estimates and improve the precision of the estimates.

Table 6. Impact of Missing Observations on Coefficient Estimates

	MVN Imputation			**Univariate Imputation**		
	RVI[1]	**FMI**[2]	**RE**[3]	**RVI**	**FMI**	**RE**
Age	0.068	0.064	0.997	0.006	0.005	0.999
Owned Land	0.050	0.048	0.998	0.026	0.026	0.997
Land Rented Out	0.028	0.027	0.999	0.088	0.082	0.992
Land Rented	0.130	0.117	0.994	0.037	0.036	0.996
State	0.042	0.040	0.998	0.009	0.009	0.999
Farm Sales	0.076	0.071	0.996	0.010	0.010	0.999
Non-family Labor	0.047	0.045	0.998	0.012	0.012	0.999
Environmental Perceptions						
Water Quality	0.088	0.081	0.996	0.019	0.019	0.998
Air Quality	0.088	0.081	0.996	0.042	0.040	0.996
Global Warming	0.120	0.109	0.995	0.009	0.009	0.999
Sources of Information / Institutions						
Other Farmers	0.119	0.107	0.995	0.016	0.016	0.998
Non-farming Neighbors	0.104	0.095	0.995	0.017	0.017	0.998
Banks	0.162	0.141	0.993	0.023	0.022	0.998
Contractors	0.106	0.096	0.995	0.011	0.011	0.999
University/Extension	0.076	0.072	0.996	0.001	0.001	0.999
USDA	0.106	0.097	0.995	0.003	0.003	0.999
Other Government Org.	0.071	0.066	0.997	0.018	0.018	0.998
Off-farm Income						
Farm Operator	0.124	0.111	0.994	0.020	0.020	0.998
Spouse	0.244	0.200	0.990	0.030	0.029	0.997
Education						
Farm Operator	0.196	0.166	0.992	0.031	0.030	0.997
Spouse	0.358	0.269	0.987	0.034	0.033	0.997
Total Animal Units	0.023	0.023	0.999	0.017	0.017	0.998
Constant	0.088	0.081	0.996	0.006	0.006	0.999

Note:[1] Shows the relative variance increase.
[2] Shows the fraction of missing information.
[3] Shows the relative efficiency.

Table 5 provides the comparison of logistic regression results between the no-imputation case and the univariate multiple-imputation method with *M* set as 10. The hypothesis that all the regression coefficients, except the constant term, are zero was rejected for the univariate imputed regression with p-value of 0.000. Hence, in addition to no-imputation regression being statistically significant, univariate-imputed regression was also statistically significant. Similar to no-imputation regression, age and land rented variables were not statistically significant in the univariate-imputed regression at 10 percent significance level (even if they had lower p-values in the univariate imputed regression). However, the university/extension variable was not significant in the univariate-imputed regression. Hence, the univariate-imputed regression shows that the university/extension variable was overstated in the no-imputation regression. A comparison of the standard errors of the coefficient estimates shows that the majority of the coefficient estimates (e.g., 19 variables out of 23) have *lower* standard errors in the univariate imputed regression than in the no-imputation regression. Similar to the case of the MVN multiple imputation, the other four coefficient estimates had no change in their standard errors.

For the magnitude of coefficient estimates, the univariate-multiple imputation caused 9 variables to have a decrease and 11 variables to have an increase in the coefficient estimates in comparison to the no-imputation regression. The other 3 variables had no change in the magnitude of their coefficient estimates.

The comparison of the regression results between the two multiple imputation results show that age and land rented variables were statistically significant in the MVN multiple imputation regression, but not in the univariate multiple regression. The comparison of the standard errors between the MVN multiple imputation regression and the univariate multiple imputation regression showed that all 19 variables, for which standard errors were lower than for the no-imputation regression, had lower standard errors in the MVN imputation regression than in the univariate imputation regression. Hence, MVN multiple imputation regression results were more efficient than the univariate multiple imputation regression. In terms of the magnitude of the change in coefficient estimates, 16 variables had higher changes in the MVN multiple imputation regression than in the univariate multiple imputation regression. Only 4 variables had greater increases in their coefficient estimates in the univariate multiple imputation regression than in the MVN multiple imputation regression. Overall, there could be differences in regression results and policy recommendations based on using MVN multiple imputation or univariate multiple imputation. Since, MVN multiple imputation resulted in lower standard errors for 19 variables than univariate multiple imputation, MVN multiple imputation resulted in more efficient estimates in the current study. Hence, we would prefer to use the regression results from the MVN multiple imputation.

Table 6 shows the impact of missing variables on the variable estimates for the MVN multiple imputation regression and the univariate multiple imputation regression. The relative variance increase (RVI), which is the increase in the variance of a variable due to missing information (Schafer, 1997), was relatively low for all variables in both the MVN multiple imputation regression and the univariate multiple imputation regression, with the exception of spouse's off-farm income and education variables for the MVN multiple imputation regression. This was expected as these two variables had the highest percentage of missing observations. The same results were also valid for the fraction of missing information (FMI), which is the ratio of information lost due to missing-data to the total information that would be present if there were no missing-data (Schafer, 1997). Lastly, the relative efficiency in table 6 is helpful in deciding the number of total imputations, *M*, which was 10 in the current study. Relative efficiency (RE) shows the ratio of the variance of an estimator with *M* set as 10 to the variance if *M* was infinite (Rubin, 1987). The results showed that RE was very high for both the MVN and the univariate multiple imputation regressions. Hence, the choice of setting *M* as 10 in the current study was justified.

6. Conclusions

Missing-data is a problem that occurs frequently in primary data collection using survey instruments. Missing-data can result in biased estimates and reduced efficiency for regression analysis. The current study analyzed the impact of missing-data and multiple imputation methods on an agricultural household survey. The current study showed that there can be a significant amount of information lost in household surveys due to non-response. Although the individual percentages of missing-data were low for the variables of interest in the current study, overall, 44 percent of observations could not be used in a standard regression due to non-response, if imputation methods had not been used.

The current study also analyzed the impact of using the multivariate normal multiple imputation method when some of the variables had a discrete distribution and the results were compared to the univariate multiple imputation method, which provided imputation based on the distribution of each variable. Our results showed that the regression estimates had lower standard errors for multivariate normal multiple imputation regression than for the univariate multiple imputation regression. Hence, the multivariate normal method is preferred to the univariate method, even when the variables have discrete distributions. Overall, both multiple imputation methods provided regression estimates with lower standard errors than using no-imputation regression. Hence, use of multiple imputation methods can improve the efficiency of regression results.

Multiple imputation results have important policy implications. Policy makers can end up enforcing different policies based on whether they used no-imputation regression or the multiple imputation regressions. This is because multiple imputation-based regression and regression using only complete observations can have differences in the sign, magnitude, and statistical significance of coefficient estimates. Since multiple-imputation methods provide unbiased estimates and increased efficiency of the regression estimates, policy recommendations should be made using multiple imputation methods rather than using regressions with missing observations.

References

Abdulai, A., Owusu, V. , & Bakang, J. E. (2011). "Adoption of Safer Irrigation Technologies and Cropping Patterns: Evidence from Southern Ghana", *Ecological Economics*, 70, 1415-1423.

Ahearn, M., David, B., Clay, D. M. & Milkove, D. (2011). "Comparative Survey Imputation Methods for Farm Household Income." *American Journal of Agricultural Economics*, 93(2): 613-618.

Barham, B. L., Foltz, J. D., Smith, D. J. & Moon, S. (2004). "The Dynamics of Agricultural Biotechnology Adoption: Lessons from rBST Use in Wisconsin, 1994-2001." *American Journal of Agricultural Economics,* 86(1): 61-72.

Chang, H. H., & Boisvert, R. N. (2005). "Are Farmers' Decisions to Work off the Farm and Participate in the Conservation Reserve Program Independent, Joint or Sequential?" *Selected Paper for Presentation at the American Agricultural Economics Association Annual Meeting*, Providence, Rhode Island, July 24-27.

Cornejo, F. J., Hendricks,C. & Mishra, A. (2005). "Technology Adoption and Off-farm Household Income: The Case of Herbicide-Tolerant Soybeans." *Journal of Agricultural and Applied Economics*, 37(3), 549-563.

Couvillion, C. W., F. Kari, F., Hudson, D. & Allen, A. (2000). "A Preliminary Economic Assessment of Roundup Ready Soybeans in Mississippi." Mississippi State University, Research Report 2000-005.
http://ageconsearch.umn.edu/bitstream/15783/1/rr00-005.pdf

Enders, C.K. (2010). "Applied Missing-data Analysis." Gilford Press, New York.

Feder, G., & O'Mara. G. T. (1982). "On Information and Innovation Diffusion: A Bayesian Approach." *American Journal of Agricultural Economics*, 64, 141-145.

Feder, G., Just, R.E. & Zilberman, D. (1985). "Adoption of Agricultural Innovations in Developing Countries: A Survey." *Economic Development and Cultural Chang*e, 33(2), 255-298.

Feder, G., Murgai, R. & Quizon, J. B. (2004). "Sending Farmers Back to School: The Impact of Farmer Field Schools in India." *Review of Agricultural Economics*, 26(1): 4-62.

Freeman, A.M., Herriges, J. A. & Kling, C. L. (2014). The Measurement of Environmental and Resource Values: Theory and Methods. Third Edition. Resources for Future Press, New York, NY.

Gedikoglu, H. (2008). "Adoption of Nutrient Management Practices." Ph.D. Dissertation, University of Missouri.

Gedikoglu, H. & McCann, L. (2010). "What Have We Learned About Adoption of Agricultural and Conservation Technologies and How Can We Use that Knowledge to Improve Policies, Programs and Technologies?" In T. Napier, ed. *Human Dimensions of Soil and Water Conservation*, Nova Publishers.

Gedikoglu, H., McCann, L. & Artz, G. (2011). "Off-farm Work Effects on Adoption of Nutrient Management Practices." *Agricultural and Resource Economics Review*, 42 (2): 293-306.

Gelman, A., & Rubin, D. B. (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, 7: 457–472.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004). "Bayesian Data Analysis." 2nd ed. London: Chapman & Hall/CRC.

Graham, J.W., Hofer, S.M. & Piccinin, A.M. (1994). Analysis with Missing-data in Drug Prevention Research. *Advances in Data Analysis for Prevention Intervention Research,* National Institute on Drug Abuse.

Grilliches, Z. (1957). "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25(4), 501-522.

Greene, W. H. (2008). "Econometric Analysis." New York: Prentice-Hall Inc.

Hiebert, L. D. (1974). "Risk, Learning, and the Adoption of Fertilizer Responsive Seed Varieties." *American Journal of Agricultural Economics,* 56, 764-768.

Horton, N. J., & Kleinman, K. P. (2007). "Much Ado about Nothing: A Comparison of Missing-data Methods and Software to Fit Incomplete Data Regression Models." *American Statistician*, 61: 79–90.

Hua W., C. Zulauf, & Sohngen, B. (2004). "To Adopt or Not to Adopt: Conservation Decisions and Participation in Watershed Groups." *Selected Paper for Presentation at the American Agricultural Economics Association Annual Meeting,* Denver, Colorado, July 1-4.

Huffman, W. E. (1980). "Farm and Off-Farm Work Decisions: The Role of Human Capital." *The Review of Economics and Statistics*, 62(1), 14-23.

Husen, N.A., Loos, T.K. & Siddig, H.A. (2017). "Social Capital and Agricultural Technology Adoption among Ethiopian Farmers." *American Journal of Rural Development 5*, 65-72.

Just, R. E. & Zilberman, D. (1988). "The Effects of Agricultural Development Policies on Income Distribution and Technological Change in Agriculture." *Journal of Development Economics,* 28, 193-216.

Khanna, M. (2001). "Sequential Adoption of Site-Specific Technologies and Its Implications for Nitrogen Productivity: A Double Selectivity Model." *American Journal of Agricultural Economics,* 83(1), 35-51.

Lee, K.J., & Carlin, J.B. (2010). "Multiple Imputation for Missing-data: Fully Conditional Specification Versus Multivariate Normal Imputation." *American Journal of Epidemiology,* 5:624-632.

Little, R. J. A., & Rubin, D. B. (2002). "Statistical Analysis with Missing-data." 2nd ed. Hoboken, NJ: Wiley.

Moreno, G., & Sunding, D. L. (2005). "Joint Estimation of Technology Adoption and Land Allocation with Implications for the Design of Conservation Policy." *American Journal of Agricultural Economics,* 87(4), 1009-1019.

Moss, C., & Mishra, A. K. (2011). "Imputing Missing Information in the Estimation of Production Functions and Systems." *American Journal of Agricultural Economics*, 93(2): 619-626.

Qaim, M., & de Janvry, A. (2003). "Genetically Modified Crops, Corporate Pricing Strategies, and Farmers' Adoption: The Case of Bt Cotton in Argentine." *American Journal of Agricultural Economics,* 85(4), 814-828.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J. & Solenberger, P. (2001). "A Multivariate Technique for Multiply Imputing Missing Values Using a sequence of regression Models." *Survey Methodology,* 27: 85–95.
Rahm, M.R. & Huffman, W.E. (1984). "The Adoption of Reduced Tillage: The Role of Human Capital and Other Variables." *American Journal of Agricultural Economics, 66,* 405-413.
Reiter, J. P., & Raghunathan, T. E. (2007). "The Multiple adaptations of Multiple Imputation." *Journal of the American Statistical Association,* 102: 1462–1471.
Robbins, M.W., & White, T.K. (2011). "Farm Commodity Payments and Imputation in the Agricultural Resource Management Survey." *American Journal of Agricultural Economics*, 93(2): 606-612.
Rubin, D. B. (1987). "Multiple Imputation for Nonresponse in Surveys. New York: Wiley.
Schafer, J. L. 1997. "Analysis of Incomplete Multivariate Data." Boca Raton, FL: Chapman & Hall/CRC.
Schafer, J. L., & Graham, J. W. (2002). "Missing-data: Our View of the State of the Art." *Psychological Methods,* 7:147–177.
Soule, M. J., Tegene,A. & Wiebe., K. D. (2000). "Land Tenure and the Adoption of Conservation Practices." *American Journal of Agricultural Economics,* 82(4), 993-1005.
Tokle, J. G., & Huffman, W. E. (1991). "Local Economic Conditions and Wage Labor Decisions of Farm and Rural Nonfarm Couples." *American Journal of Agricultural Economics,* 73, 52-70.
Upadhyay, B. M., Young, D. L., Wang, H. H. & Wandschneider, P. (2002). "How Do Farmers Who Adopt Multiple Conservation Practices Differ from Their Neighbors?" *Selected Paper for Presentation at the American Agricultural Economics Association Annual Meeting,* Long Beach, California, July 28-31.
van Buuren, S. (2007). "Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification." *Statistical Methods in Medical Research,* 16: 219–242.
Wozniak, G. D. (1984). "The Adoption of Interrelated Innovations: A Human Capital Approach." *The Review of Economics and Statistics,* 66(1), 70-79.

Acknowledgements

This research was supported in part by the USDA-NIFA Evans-Allen Research Program project #2011-004, the USDA-NIFA AFRI Bio-Energy Grants Program project #2012-00839, and in part by TUBİTAK-BİDEB-2232 Program project #117C017. An earlier version of this paper was presented at the 2012 AAEA Annul Meeting. We thank participants of the 2012 AAEA Annual Meeting and Kevin J. Boyle of the Virginia Tech University for their very helpful comments.

Appendix A. Survey Questions

1. In 2010, did you use Roundup Ready® Corn?
☐ Yes ☐ No
2. What year were you born? ________________
3. How many acres of land did you own in January 2010? (Please write the number of acres on the line.)
________ Acres

4. How many acres of land did you rent to other farmers in 2010?
________ Acres
5. How many acres of land did you rent from others in 2010?
________ Acres

6. Please list the county state(s) where your farm is located.
______________________ State (or states)
7. What amount of gross farm sales did you have in 2010?
a. Between $1 and $9,999
b. Between $10,000 and $49,999
c. Between $50,000 and $99,999
d. Between $100,000 and $249,999
e. Between $250,000 and $499,999
f. $500,000 or more
8. Did you hire non-family farm labor in 2010?
☐ Yes ☐ No
9. To what extent do you agree or disagree with the following statements? Please circle the number that best corresponds to your answer.

	Strongly Disagree		Neither Agree nor Disagree		Strongly Agree
a. I am concerned about the water quality of streams and lakes in my county.	1	2	3	4	5
b. I am concerned about the air quality in my county.	1	2	3	4	5
c. I am concerned about the global warming.	1	2	3	4	5

10. How much influence does each of the following have on agricultural production decisions you make? (Please circle the number that best indicates the amount of influence.)

	None		Some		Very Much
a. Other farmers	1	2	3	4	5
b. Non-farming neighbors	1	2	3	4	5
c. Banks	1	2	3	4	5
d. Contractors	1	2	3	4	5
e. University/Extension	1	2	3	4	5
f. USDA	1	2	3	4	5
g. Other government organizations	1	2	3	4	5

11. What was your annual off-farm gross income in 2010?
a. No off-farm income
b. Between $1 and $9,999
c. Between $10,000 and $24,999
d. Between $25,000 and $49,999
e. Between $50,000 and $99,999
f. $100,000 or more
12. What was your annual off-farm gross income in 2010?
a. No off-farm income
b. Between $1 and $9,999
c. Between $10,000 and $24,999
d. Between $25,000 and $49,999
e. Between $50,000 and $99,999
f. $100,000 or more

13. What is your highest level of education?
a. Less than High School
b. High School
c. Some college or vocational school
d. Bachelor’s degree
e. Graduate degree, such as Master’s
14. What is your spouse’s highest level of education?
a. Less than High School
b. High School
c. Some college or vocational school
d. Bachelor’s degree
e. Graduate degree, such as Master’s
15. How many of the following livestock animals of all ages did you have on your farm at one time in 2010?
________ Dairy cattle
________ Beef cattle on feed for slaughter market
________ Beef cows
________ Swine 55 lbs. or less
________ Swine more than 55 lbs.
________ Broilers
________ Turkeys
________ Sheep or goats
________ Other livestock (please list) ___________________