Direct Payments, Cash Rents, Land Values, and the Effects of Imputation in U.S. Farm-level Data

Michael W. Robbins and T. Kirk White

Research using the Agricultural Resource Management Survey (ARMS) and other data shows that direct government payments to farmers increase rents and the price of land. However, some ARMS data is imputed and does not account for relationships between payments and other variables. We investigate various imputation methods and benefits gained from a method with a wide scope rather than a parsimonious range of variables. Using our method, we estimate that an additional dollar of direct payment increases land value about $2.69 more per acre than ARMS imputation methods and that our imputations (using an exhaustive iterative sequential regression) outperform other methods and/or smaller models.

Key Words: Agricultural Resource Management Survey, cash rents, direct payments, farm subsidies, land values, missing data, multiple imputation, robust regression

Agricultural economists and policymakers have long been interested in the effect of federal farm program payments on the value of agricultural land to which the payments are attached (Floyd 1965, Gardner 1992, Kuchler and Tegene 1993, Barnard et al. 2001). Several recent studies have analyzed the effects of such farm subsidies on farm land rents and the value of the farm land. Using data from the U.S. Department of Agriculture's (USDA's) *Census of Agriculture* for 1992 and 1997 and various econometric models, Roberts, Kirwan, and Hopkins (2003) found that an additional dollar of government payment results in an increase of between $0.21 and $2.31 in land rents per dollar of payment. Using their preferred model, the authors concluded that this value fell between $0.34 and $0.41. In more recent research using the same data, Kirwan (2009) found that landlords captured roughly 25 percent of each additional dollar of government payment to farmers in the form of higher cash rents. Using data from USDA's Agricultural Resource Management Survey (ARMS) for 1998 through 2001, Goodwin, Mishra, and Ortalo-Magné (2011) found that an additional dollar of expected loan deficiency payment appeared to add $27.00 to the value of the land. These researchers also found that an additional dollar per acre of direct payment (or production flexibility contract payment, as they were called prior to 2002) raised cash rents by $0.72 per acre.

Michael W. Robbins is an associate statistician with the RAND Corporation in Pittsburgh, Pennsylvania, and T. Kirk White is an economist with the U.S. Census Bureau's Center for Economic Studies. Correspondence: *T. Kirk White ▪ Center for Economic Studies ▪ 4600 Silver Hill Road ▪ Washington, DC 20233 ▪ Phone 301.763.1879 ▪ Email thomas.kirk.white@census.gov*.

Michael Robbins acknowledges partial funding from the University of Missouri Research Board. The authors thank Sujit Ghosh for helpful comments. The views expressed are those of the authors and do not necessarily represent the views of the Economic Research Service, National Agricultural Statistics Service, U.S. Department of Agriculture, RAND, or the U.S. Census Bureau.

Several of the variables included in the ARMS and pertinent to the aforementioned models involve imputed data. Like most surveys, the data generated by the ARMS suffers from item non-response, and the National Agricultural Statistics Service (NASS), which conducts the ARMS, uses imputed data for missing values in about 150 variables for which the missingness rates range from just 1 percent to 43 percent (Robbins et al. 2011a). For example, in the 2008 ARMS, imputed farm-level data made up 23 percent of *LDP Payments for Target Commodity*, 31 percent of *Countercyclical Payments for Target Crop*, 43 percent of *Value of Commodity Certificates*, and 31 percent of *Wetland Reserve Program Payments*. For relationships between direct payments, cash rents, and land values, one can compare estimates from the ARMS data to results from other data sources, such as panel data used by Ifft, Kuethe, and Morehart (2013).[1] However, for many questions of interest to researchers and agricultural policymakers, ARMS is the only nationally representative source of data. Given the key role ARMS data thus plays in agricultural research and policy discussions, it is important for researchers and policymakers to be aware of the potential effects of imputed data on regression results.

Both NASS and USDA's Economic Research Service (ERS) impute for missing items using conditional means.[2] Since the deficiencies of these methods have been demonstrated (Miller, Robbins, and Habiger 2010), our analysis includes imputations generated using the recently developed iterative sequential regression (ISR) procedure for imputation (Robbins, Ghosh, and Habiger 2013).[3] ISR is a regression-based Markov chain Monte Carlo (MCMC) algorithm, and unlike the NASS and ERS methods, it includes the flexibility to greatly expand the scope of data incorporated into the imputation procedure, thereby tasking the imputer with selection of an imputation model. We compare the imputation methods generally used by NASS and ERS with three types of ISR imputations—exhaustive, parsimonious, and deficient—and evaluate the biases attributable to each method in estimates of incidence of farm subsidies on farm land values and cash rents, focusing on the effect of the method and model on the presence and magnitude of such biases. Since ISR is computationally intensive, our discussion of imputation models focuses on depth of input in terms of the number of variables used with the hope that a parsimonious model that minimizes the computing burden is satisfactory. However, the preferred ISR method is the exhaustive imputation because it incorporates the most variables.

It is well-established in the literature that the choice of imputation method can have a profound effect on point estimates pertinent to econometric analyses (Robbins and White 2011, Robbins, Ghosh, and Habiger 2013). For example, when using the official USDA imputations, we find that one dollar of direct payment per acre increases the per-acre value of land by \$16.78, whereas the value shifts to \$19.47 when calculated using our preferred imputations (ISR with an exhaustive imputation model). The higher value is more in line with recent research involving relatively comprehensive

[1] The ability to compare results from ARMS to other data sources is one of our reasons for focusing on the effects of imputation on direct payments, cash rents, and land values.

[2] For NASS imputations, the conditioning variables are farm type, farm sales class, and farm region.

[3] This method and study were developed as part of a two-year cooperative agreement between NASS and the National Institute for Statistical Sciences (NISS) to improve the imputation methods NASS uses for Phase III of the ARMS.

field-level panel data (Ifft, Kuethe, and Morehart 2013). Those results suggest that estimates based on official USDA imputations in the ARMS miss about 16 percent of the increase in land value for the average acre of program crop land. In 2008 (the year of our sample), the average direct payment per acre was about \$19.57 (Ifft et al. 2012). Thus, estimates using our preferred imputation method point to an additional \$52.64 in land value per acre (payment of \$19.57 × \$2.69 additional increase in land value) associated with direct payments. In 2008 there were about 260 million base acres enrolled in the Direct Payments program (Ifft et al. 2012). A back-of-the-envelope calculation therefore suggests that using exhaustive ISR imputations instead of the official NASS imputations would increase the total land valuation associated with direct payments by about \$13.7 billion. The magnitude of the increase in per-acre value of land drops from \$19.47 to \$18.13 when using ISR with a parsimonious (seemingly sufficient) imputation model.

We address another avenue of analysis of imputed econometric data that has been largely untouched in the agricultural economic literature to date: the effect of imputation on standard errors. In addition to influencing the value of survey indicators, imputations can induce bias in the standard errors of such indicators. Further, one must adjust confidence intervals to incorporate error contained within the imputations. A popular statistical procedure for making such adjustments is called multiple imputation (MI) (Rubin 1987), but MI is considered to be inappropriate for use with data that have characteristics that are common in agricultural surveys (Kott 1995). Nonetheless, by applying MI, we demonstrate that an exhaustive imputation model can increase the accuracy of the imputations and thereby decrease the width of confidence intervals of the resulting econometric estimations.

To validate and expound upon conclusions drawn from our empirical analysis, we conduct a simulation study using complete cases from the ARMS data. We randomly "poke holes" in the complete cases and replace those data with imputations calculated using various methods and models. The simulation study verifies the quality of our preferred imputations in two ways. First, the simulations show that our preferred method reduces bias in the estimates of subsidy incidence and in standard errors. Interestingly, ISR with a parsimonious imputation model produced biased point estimates of key regression coefficients. Second, the simulations demonstrate that the precision of respective interval estimates may increase under the exhaustive imputation model. In addition, the simulation study illustrates the potential utility of MI in analyses of complex economic survey data.

Although we focus on the effects of imputation in USDA's ARMS, our findings are broadly applicable. Economic surveys generally suffer from item non-response, and the imputation methods used by most statistical agencies are not well-suited to microdata analysis. Agencies that generate imputations in economic surveys such as the ARMS will be interested in the benefits of the broad imputation model we outline here, and researchers who are analyzing imputed data should be aware of the abilities and limitations of the imputation methods and models used. Researchers must be particularly careful when a relatively large percentage of the estimation sample includes data imputed by the statistical agency that collected it and should consider alternative imputation methodologies that incorporate relevant explanatory variables if the original imputation model fails to do so. In addition, our conclusions regarding the performance of specific statistical machineries (the handful of algorithms, including ISR) provide insight into the efficacy of similar algorithms used by many statistical agencies.

Direct Payments, Cash Rents, and the Value of Land

Farmers in the United States receive several types of federal payments, including subsidies for individual commodities, emergency and disaster relief, conservation program payments, and crop-specific program payments (e.g., the peanut quota buyout). We focus solely on the Direct Payments commodity program because direct payments do not depend on market prices or current production.[4] In 2008 (the year of our sample), total direct payments were about $5.2 billion in 2009 dollars (White and Hoppe 2012) while the average direct payment was about $19.57 per acre (Ifft et al. 2012). At the time, the U.S. average cash rental rate for crop land was about $96 per acre and the average value of crop land was $2,970 per acre (USDA 2008).

The direct payments are made annually and are based on the producer's historical number of acres (the so-called "base acreage") and the yields of the program crops in prior years. A farmer is allowed to determine the base acreage in several ways, but the simplest one is to use the average number of acres planted to that crop in the historical years (1998 through 2001 under current legislation). The payment to a given farmer is calculated as the product of a percentage of the base acreage (83.3 percent under the 2008 Farm Act), the farm's gross income from selling the historical yield of that commodity, and the direct payment rate for the commodity. Landlords who share-rent their land to farmers participating in the program are eligible to receive direct payments while landlords who cash-rent the land are not.[5] The farmer's direct payment does not depend on the farm's current acreage or yield for the crop although farm-level production tends to be highly correlated over time. Since the direct payment can be calculated in advance, economic theory suggests that landlords who cash-rent can extract higher rents for crop land that is associated with greater direct payments.

Following Roberts, Kirwan, and Hopkins (2003), we hypothesize that rent, r_i, received for a unit of land i is a function of expected revenue (including government payments) associated with the land net of variable costs:

$$r_i = E\left[\Sigma_k(p_k + c_k)q_{ki} - \Sigma_j w_j x_{ji}\right] + \Sigma_k DP_{ki} \tag{1}$$

where E is the expectation operator, p_k is the market price received for commodity k, c_k is the government payment (excluding the direct payment) received per unit of production of commodity k, q_{ki} is the quantity of commodity k produced on land unit i, x_{ji} is the quantity of input j (other than land) used on unit i, w_j is the marginal cost of that input, and DP_{ki} is the direct payment received on land unit i for commodity k. Commodity payments that

[4] Because the value of such payments is known in advance, it is easier to measure the incidence of these subsidies relative to other types of subsidies since there is essentially no difference between observed payments and the payments recipients expected to receive when they negotiated prices and cash rental agreements for land. Several other studies have done the same (Roberts, Kirwan, and Hopkins 2003, Goodwin, Mishra, and Ortalo-Magné 2011, Ifft, Kuethe, and Morehart 2013). Thus, our focus on direct payments also facilitates comparisons with the existing literature.

[5] Under a share-rent arrangement, the tenant gives the landlord a share of the crop as payment for use of the land. Under a cash-rent arrangement, the tenant pays the landlord a specified amount of cash per acre.

are tied to current production of the commodity can induce a greater supply, which may lead to a lower price for the commodity. Greater production may also induce an increase in the price of inputs other than land, w_j. Thus, by equation 1, if commodity payment c_k is increased, some of the increase may be captured by landlords and by other market participants if the commodity price, p_k, falls and the cost of inputs, w_j, rises (see Roberts, Kirwan, and Hopkins (2003) for a more thorough discussion). Since direct payments are not tied to current production, under equation 1 landlords will capture any increase in the direct payment by charging greater cash rents. In practice, direct payments may induce a supply response if, for example, farmers expect to be allowed to update the amount of base acreage in the future, as they were in 2002. In that case, landlords may not be able to capture 100 percent of an increase in direct payments.

We operationalize equation 1 by estimating the following equation, which is similar to equation 2 in Roberts, Kirwan, and Hopkins (2003):

$$(2) \qquad CR_i / AR_i = \alpha_0 + \alpha_n(NFI_i / A_i) + \alpha_d(DP_i / A_i) + \mathbf{X}_i\boldsymbol{\alpha}_X + u_i$$

where CR_i is the farm's cash rents for unit i, NFI_i is net farm income excluding direct payments, DP_i is direct payments, $\mathbf{X}_i$ is a vector of categorical variables, and u_i is an error term. In this formulation, A_i and AR_i are acreage variables (described in greater detail under sample selection and data for regression analysis) and α_d is the coefficient of interest. As noted by Roberts, Kirwan, and Hopkins (2003), the coefficient α_d may be biased in a linear regression of CR_i on NFI_i and DP_i when using cross-sectional data for several reasons. Differences between expected and realized net farm income show up in the error term and bias α_d toward zero. And since direct payments tend to be geographically correlated, estimates of α_d may be biased because of unobserved geographic heterogeneity in factors such as yields. With panel data, we could control for unobserved heterogeneity with, for example, farm fixed effects.[6] We have only cross-sectional data.[7] However, our main goal is to assess the effect of various imputation methods on estimates of α_d.

---

6 Some previous studies of the effects of government payments on cash rents and/or land values have used farm-level panel data. Roberts, Kirwan, and Hopkins (2003) used a farm-level panel constructed from the 1992 and 1997 *Census of Agriculture*. Neither the imputation methods used nor the ways in which the imputed data were identified in those censuses have been made available to researchers. Furthermore, while the 2007 *Census of Agriculture* asked a separate question about direct payments, previous versions lumped direct payments together with other types of government payments. Thus it is not yet possible to construct a farm-level panel data set from the agricultural census that includes direct payments as a farm-level variable. Ifft, Kuethe, and Morehart (2013) used a five-year rotating field-level panel constructed from NASS's annual June Area Survey (JAS). The JAS includes information about farm land values but does not include data on program payments. The authors obtained program payment information from Internal Revenue Service Form 1099 data aggregated at a county level. Information on the methods used for imputation in the JAS also has not been made available to researchers. Unfortunately, also, one cannot construct a farm-level panel data set from the ARMS data. To reduce the burden on respondents, the ARMS sampling procedure is designed to minimize the probability that a farm that was sampled in the last ARMS survey is sampled in the current one.

7 In theory, the bias in our estimators of α_d and β_d could be positive or negative depending on the correlation between unobserved geographic heterogeneity and the level of direct payment per acre. However, in practice, the bias seems to be positive. Roberts, Kirwan, and Hopkins (2003) constructed 40 estimates of the effect of government payments on cash rents (analogous to our coefficient α_d) using three estimators, three samples, five specifications with county fixed effects, and five specifications without county fixed effects. In every case, they found that the estimate from

Following the same logic, the value of crop land should increase with both net farm income (excluding direct payments) and direct payments. We operationalize this by regressing the per-acre value of land rented from others (VLR_i) on NFI_i and DP_i:

$$(3) \qquad VLR_i \,/\, AV_i = \beta_0 + \beta_n(NFI_i \,/\, A_i) + \beta_d(DP_i \,/\, A_i) + \mathbf{X}_i\boldsymbol{\beta}_X + v_i$$

where AV_i is another acreage variable (more fully described under sample selection and data for regression analysis) and v_i is an error term. For the vector **X**, we use three categorical variables (farm type, farm sales class, and farm region) that form a framework for the strata of the ARMS data. We exclude interactions to avoid augmentation of the design matrix. As is standard practice, each categorical variable is input into the regression scheme as a sequence of binary variables (each indicating the category of the categorical variable).

Sample Selection and Data for Regression Analysis

Our data set is the 2008 ARMS survey, which is jointly designed and administered annually by NASS and ERS. The survey covers U.S. farming operations and their operators in the 48 contiguous states. In our model, when estimating the quantities given in equations 2 and 3, we limit the sample to farms that had nonzero values for crop land acreage, direct payments, cash rents paid, and acres rented.

The amount of direct payment received is an ARMS survey variable that is to be scaled by an acreage variable, *A*, that identifies the number of acres of the farm associated with the appropriate payment program. It is unclear which ARMS variable should be used as *A*; the two best options are crop land acres and acres operated. For the empirical analysis, we use the total crop acreage of the farm, which will yield the most economically meaningful estimations. In the simulation study (the purpose of which is to gauge the efficacy of various procedures rather than to produce meaningful estimations), on the other hand, the results are based on direct payments being scaled by both crop land acres and acres operated.

We use the ARMS variable *Cash Rent Paid for Land and Buildings* as a measure of rents, and that variable is scaled by *AR*, the number of acres rented. We measure *NFI* as follows. First we subtract direct payments and cash rents paid from the ARMS variable *Net Farm Income*, which ERS constructs from other farm-level revenue and cost variables. That quantity also must be scaled by *A*. The dependent quantity in equation 3 is calculated by dividing *VLR* (represented by the ARMS variable *Market Value of Land Rented from Others*) by *AV*, which is the sum of *Acres Cash-Rented*, *Acres Share-Rented*, and *Acres Rented for Free*.

NASS imputes data for about 150 ARMS variables that may be missing values. Table 1 presents the total number of observations, the number of missing values, and the number of missing values as a percentage of total observations for the key variables in our analysis, including *DP* and *VLR*. Table 1 also lists missingness rates for the variables used in the exhaustive ISR model. The percent missing varies considerably, from only 1 percent for ARMS's *Income from Federal Crop Insurance* to 43 percent for *Government Payments Received by Landlord*. It is also important to note the variables that

the specification with county fixed effects was lower than the estimate without county fixed effects.

Table 1. Number of Observed and Missing Values for Selected ARMS Variables

		Observations		
ARMS Item	Variable Name	Total	Number Missing	Percent Missing
855	Market value of land rented from others	10,495	2,185	17
525	Direct and countercyclical payments received	9,342	329	3
520	Direct payments as percent of item 525	8115	931	10
552	Income from federal crop insurance	1,544	24	1
543	Government payments received by landlord	1,719	1,297	43
852	Value of other farm buildings	15,298	940	5
854	Value of land	17,365	2,556	13
878	Value of inputs owned on January 1	8,523	540	6
883	Value of tractors owned	18,398	1,188	6

Source: ARMS (2008).

are not included in Table 1. Most of the acreage variables (e.g., *Acres Cash-Rented*, *Acres of Crop Land*, and *Acres Share-Rented*) are not imputed by NASS; they are fully observable (have no missing values) in the ARMS data set. *Cash Rent Paid for Land and Buildings* is also not eligible for computer imputation by NASS.

Estimation for Regression Analysis

Both NASS and our study use imputation to create a complete data set, and we now describe how the coefficients in equations 2 and 3 are estimated using that data set. The ARMS's design weights are crucial to analysis of its data. Letting w_i represent the calibrated design weight for unit i, we prefer to estimate the regression coefficients with weighted least squares while using $w_i^* = w_i \times A_i$ as the weights. However, the ARMS data tend to be highly skewed (Robbins, Ghosh, and Habiger 2013), and per-acre versions of the pertinent variables are highly skewed as well. This skewness results in several extreme observations that have a large influence on the values of coefficients found using least squares. To compensate for these influential observations, Roberts, Kirwan, and Hopkins (2003) removed the largest 1 percent of each relevant variable as outliers. We instead employ robust regression (Huber and Ronchetti 2009) in which outliers are iteratively reweighted (as opposed to being discarded) to reduce their influence. We use the package *RLM* in *R* to calculate regression coefficients (the algorithm returns estimates of all regression coefficients and their standard errors), and we input the vector of w_i as prior weights into the algorithm.

We calculate our estimates using MI, which includes an assumption that the imputation method randomly samples from a predictive distribution. Therefore, the imputation process can be repeated to create m imputed data sets for which the imputations are assumed to be independent across the data sets. After creating m data sets, we use the estimation process previously described to determine values for all of the regression coefficients and their respective variances (i.e., standard errors squared) for each imputed data set.

We pool information across data sets using Rubin's combining formulas for MI (Rubin 1987). Let β represent a regression coefficient of interest and $\hat{\beta}^{[k]}$ represent the estimated value of β found using the kth completed data set. The MI point estimate of β thus is $\bar{\beta} = \sum_{k=1}^{m} \hat{\beta}^{[k]} / m$. Let $v(\hat{\beta}^{[k]})$ represent the estimated variance of $\hat{\beta}^{[k]}$. The quantity is called within-imputation variance and represents a point estimate of the variance of $\bar{\beta}$ had there been no missingness. A confidence interval for β is calculated using the total variance of $\bar{\beta}$, which is calculated as $T = \bar{v}(\hat{\beta}^{[k]}) + (1 + [1 / m])B$ where B is the between-imputation variance and is calculated as the sample variance of $\hat{\beta}^{[k]}$. The between-imputation variance provides a measure of the error of the imputations and is included to ensure that imputation error factors into the interval estimate.

The Imputer's Model: Iterative Sequential Regression with Varying Depth of Input

Our preferred method of imputation is ISR (Robbins et al. 2011b), which was developed specifically for use with ARMS and designed with the flexibility to include a wide range of input data. A similar procedure (Robbins and White 2011) was shown to improve econometric analysis of ARMS data relative to older methods, and Robbins, Ghosh, and Habiger (2013) describe the procedure in detail and give several illustrations of the utility of the method. Our focus is to study the effect of the imputer's model on the value of regression coefficients estimated from the models in equations 2 and 3.

ISR involves two primary phases: transformation and imputation. The transformation phase applies robust transformations that were designed specifically for ARMS data and were used in Robbins, Ghosh, and Habiger (2013). Note that the ARMS data predominantly consist of skewed semicontinuous variables (i.e., variables that are positive, continuous, and highly skewed apart from a large mass at zero). Such variables are handled by first creating dummy (0/1) variables that indicate whether the variable is positive and then by treating all observed zeroes in the original semicontinuous variables as missing. Since survey enumerators are usually able to determine whether a respondent should have positive values for each item, all of the values originally coded as missing are treated as being positive—a longstanding characteristic of machine imputation in the ARMS. As a result, all of the dummy variables are fully observed. Next, a density-based transformation (here, as in Robbins, Ghosh, and Habiger (2013), the transformation is of a skew normal density family) is applied to the continuous portions of the variables, which ensures approximate normality (following transformation) of all of the variables with missing values.

The second phase is a form of data augmentation (Tanner and Wong 1987, Little and Rubin 2002) that uses the MCMC algorithm to iteratively draw imputations from a predictive model. A key characteristic of the data augmentation phase is that it jointly models all of the variables that require imputation. Specifically, letting $X_1, \ldots, X_p$ (where the index now denotes the variable) represent the variables (there are p of them) that have missing values following transformation and letting $\mathbf{Z}$ denote a set of fully observed covariates, the ISR constructs a joint model for pertinent ARMS variables using the fact that the joint distribution of the variables that have missing values can be expressed as a product of conditional distributions. That is,

$$P(X_1, \ldots, X_p \mid \mathbf{Z}) = \prod_{j=1}^{p} P(X_j \mid \mathbf{Z}, X_1, \ldots, X_{j-1}) \tag{4}$$

where $P(\cdot)$ denotes general notation for a distribution function. This formula allows the imputer to specify the form of each conditional distribution occurring in the righthand side of the equation. Here, we assume the linear form

$$X_j = \gamma_{j,0} + \boldsymbol{\gamma}_j Z + \gamma_{j,1} X_1 + \ldots + \gamma_{j,j-1} X_{j-1} + \varepsilon_j \tag{5}$$

for $j = 1, \ldots, p$ where $\boldsymbol{\gamma}_j$ represents a vector of regression parameters that correspond to the fully observed variables and ε_j represents a standard normal error. One advantage of using this expression for imputation of high-dimensional economic data is that the imputer has the liberty to remove variables from inclusion in any of the conditional linear models.

Like most data augmentation methods, ISR uses a Bayesian model to place distributional assumptions on parameter values. Within each iteration of the MCMC are two steps. In the first (the *I* step), ISR samples imputations, and in the second (the *P* step), it samples parameter values. In this description, "sampling" is used in a Bayesian (or Monte Carlo) sense in that the imputations and parameter values are sampled by simulating values from theoretical probability distributions. In the regression, θ represents the set of all model parameter values, $\boldsymbol{\chi}_{mis}$ represents the missing portion of the data, $\boldsymbol{\chi}_{obs}$ represents the observed portion of the data, and $\boldsymbol{\chi}_{mis}^{(t)}$ and $\theta^{(t)}$ denote the values of $\boldsymbol{\chi}_{mis}$ and θ, respectively, at the t^{th} iteration. $P(\cdot)$ represents general notation for a distribution function. The *I* step of the $(t + 1)$th iteration samples updated imputations using

$$\boldsymbol{\chi}_{mis}^{(t+1)} \backsim P\left(\boldsymbol{\chi}_{mis}^{(t)} \middle| \boldsymbol{\chi}_{obs}, \theta\right)$$

and the *P* step samples updated parameters values using

$$\theta^{(t+1)} \sim P\left(\theta^{(t)} \middle| \boldsymbol{\chi}_{mis}^{(t)}, \boldsymbol{\chi}_{obs}\right).$$

The form of

$$P\left(\boldsymbol{\chi}_{mis}^{(t)} \mid \boldsymbol{\chi}_{obs}, \theta\right) \text{ and } P\left(\theta^{(t)} \middle| \boldsymbol{\chi}_{mis}^{(t)}, \boldsymbol{\chi}_{obs}\right)$$

can be determined using the models expressed in equations 4 and 5 with the assumption of a noninformative prior for parameters in each conditional model. After a fixed number of iterations (b), the process is stopped and $\{\boldsymbol{\chi}_{mis}^{(b)}, \boldsymbol{\chi}_{obs}\}$ is returned as the imputed data set. To generate multiple imputations, the Markov chain can be extended to sample one more set of imputations after each additional c iteration beyond the bth iteration (see Schafer (1997), among others). Here, we use $b = 500$ and $c = 250$, which have been shown to be sufficient for the ARMS data by prior analysis.

As is evident from its description, ISR is a rather costly process computationally. Each imputed data set requires hundreds to thousands of iterations of the MCMC algorithm plus several imputed data sets must be created to implement MI. In addition, there are 30,000 to 40,000 data units included in this process. Incorporating hundreds of the available ARMS variables into the model used for imputation results in an algorithm that takes weeks (or months) to run even when using sophisticated computing, time that agencies involved do not have.

With such a vast amount of computation time at stake, how does one select the scope of input for use in the imputation model? Is it sufficient to use only variables that are relevant for the analysis? Or are marked gains produced when the scope of the imputation model is broadened? Consequently, we apply the ISR method under three input regimes (i.e., imputer's models):

- *The deficient model* uses only the three categorical variables incorporated into equations 2 and 3 as covariates for imputation.

Table 2. Explanatory Variables in the Exhaustive Iterative Sequential Regression Imputation Models

		Imputation Model	
ARMS Item	Variable Name	Cash Rents	Value of Rented Land
21	Acres cash rented	●	●
23	Acres share rented	●	●
26	Total acres operated	●	●
30	Total crop land acres	●	●
39	Acres covered, federal insurance	●	●
44	Cash rent paid for land and buildings	●	●
47	Farm type	●	●
54	Count of landlords		●
501	Major grains, cash sales	●	●
502	Other grains and oilseeds, cash sales	●	●
503	Tobacco and cotton, cash sales	●	
511	Cattle and calves, cash sales		●
—	Direct payments	●	●
—	Adjusted net farm income	●	●
—	Region	●	●
—	Gross value of sales	●	●
—	Cash crop sales	●	●
—	Fertilizer expenses	●	●
—	Tax expenses	●	●
—	Calibrated design weights	●	●
—	Design weights, crop land acres	●	●
477	CRP / CREP / WRP payments[a]	●	
552	Income from federal crop insurance	●	
543	Government payments received by landlord	●	
852	Value of other farm buildings		●
854	Value of land		●
878	Value of inputs owned on January 1		●
883	Value of tractors owned		●

[a] CRP stands for Conservation Reserve Program, CREP for Conservation Reserve Enhancement Program, and WRP for Wetlands Reserve Program.

Notes: Variables without item numbers are constructed from other variables in the survey questionnaire. Adjusted net farm income is net farm income less direct payments.

- *The parsimonious model* uses only the information that is directly relevant to the analyst's models (i.e., all of the variables in equations 2 and 3 but no others) as input into the imputation procedure.
- *The exhaustive model* incorporates variables for information that is not directly relevant to the analyst's model as input. The variables used in this model are given in Table 2.

We follow Schenker et al. (2006) and use sample weights as predictors for imputation in the exhaustive model. Robbins et al. (2011b) illustrated that incorporating such weights can improve the reliability of weighted estimations. Since product weights (w_i^*) are used in the estimation procedure, those are included in the regression as predictors for both the exhaustive and the parsimonious imputation models. We are interested in determining whether the parsimonious model is sufficient for estimation and whether improvements are gained using the exhaustive model.

Rudimentary Procedures: NASS and Approximate Bayesian Bootstrap

ISR was designed in 2011 to replace the method used by NASS to create the official imputations. The prior method (hereafter referred to as the NASS procedure) used a stratum-based form of conditional mean imputation. Specifically, NASS created a donor pool for a missing value of variable Y by collecting all positive and observed values of Y for farms observed to have the same values for sales class, farm type, and region (the three variables used as fixed effects in the analyst's models) as the farm with the missing value of Y. The imputation was set as the mean of the donor pool, and fall-back groupings were used when the donor pool was not sufficiently large. We refer the reader to Banker (2007) for a more detailed description of this method.

Drawbacks of the NASS procedure are many (see Miller, Robbins, and Habiger (2010) for further details regarding the drawbacks) and range from its limited scope of input to its incorporation of mean imputation. Conditional-mean imputation distorts the marginal distributions of the imputed variables, leading to downward-biased estimates of the variance of the variable (e.g., Little and Rubin 2002, Schafer and Graham 2002). Robbins and White (2011) and Robbins, Ghosh, and Habiger (2013) illustrated that the NASS procedure has little utility compared to more sophisticated procedures such as ISR, which, when used in conjunction with ARMS data, has been shown to preserve all of the characteristics of marginal distributions as well as all relevant aspects of the joint distribution (neither is achieved by NASS's method). As a result, ISR exceeds the ability of the NASS procedure in estimating large numbers and types of parameters, including means (when missingness is not completely random), variance components, and regression parameters.

However, several questions regarding the utility of ISR and the procedures associated with it remain. One relates to the utility of MI with complex survey data that has been subjected to imputation via ISR. Hence, we consider an approximate Bayesian bootstrap (ABB) extension (Rubin and Schenker 1986) of the NASS procedure to gauge the utility of its imputation strategy when implemented in conjunction with MI to compare the efficacy of ISR when used with MI. We use the same donor pools for the ABB method, but each ABB imputation is a random draw from the donor pool.

When estimating equations 2 and 3, we expect that the ISR method using the parsimonious and exhaustive models will provide better imputations than the ISR method using the deficient model, the NASS imputation method, and the ABB imputation method, which use only three explanatory variables: sales class, farm type, and region—the same categorical variables that we include on the righthand side of equations 2 and 3. Any variation in direct payment per acre that is independent of sales class, farm type, and region will not be captured by those methods. Parsimonious and exhaustive ISR imputation models, on the other hand, include all of the variables of interest from equations 2 and 3—measures of cash rents, land values, net farm incomes, and direct payments.

Empirical Results

We produce five imputed data sets, one for each method—NASS, ABB, and the three ISR models. For the NASS imputations, we use the official 2008 ARMS imputations (which were not randomly sampled and thus are not conducive to MI). For the other four, we generate ten imputations for each missing item and method, resulting in ten complete data sets. We then estimate equations 2 and 3 separately for each data set to produce regression coefficients and standard errors for each of the ten data sets for each method. Finally, we apply Rubin's combining formulas to pool the results from each method's ten data sets and calculate interval estimates of the regression coefficients for each method.

Table 3 presents our estimates of pertinent quantities related to estimation of α_d in equation 2 (cash rents per acre) for each imputation method: $\widehat{\alpha}_d$ (the MI point estimate of α_d), $se(\widehat{\alpha}_d)$ (the square root of the within-imputation variance of the MI estimate of $\widehat{\alpha}_d$, which quantifies the standard error of $\widehat{\alpha}_d$ under complete data), B (the between-imputation variance), and L_{MI} and U_{MI} (the upper and lower bounds, respectively, of the MI interval estimate of α_d as found with 95 percent confidence). Recall that B is a quantity that measures the variability induced into the estimate of α_d by the imputations. Therefore, a comparatively smaller value of B indicates that the imputations are more accurate (contain less error).

Table 3. Regressions of Cash Rents per Acre on Direct Payment per Acre and Adjusted Net Farm Income per Acre

	Imputation Method				
			ISR		
Quantity	(1) NASS	(2) ABB	(3) Deficient	(4) Parsimonious	(5) Exhaustive
$\widehat{\alpha}_d$	0.7610	0.6351	0.5392	0.7489	0.7542
$se(\widehat{\alpha}_d)$	0.0357	0.0331	0.0305	0.0354	0.0358
B	—	0.0020	0.0019	0.0010	0.0003
L_{MI}	—	0.5158	0.4245	0.6519	0.6740
U_{MI}	—	0.7543	0.6538	0.8459	0.8346

Note: Estimates of regression coefficients and standard errors were combined across the ten imputed data sets using Rubin's combining formulas. Observations were weighted by the product of total acres and the sampling weight.
Source: ARMS (2008).

Table 3 demonstrates that the imputation method chosen can influence the estimated coefficient for the direct payment. For instance, the NASS procedure and exhaustive ISR model produce similar values of $\hat{\alpha}_d$ while the ABB method yields a noticeably different result. Only 10 percent of the direct payment data (and none of the cash rent data) is imputed, so perhaps the small difference between the NASS estimate and the exhaustive ISR estimate should not be surprising. Habiger, Robbins, and Ghosh (2010) showed that NASS imputations induced a downward bias into estimates of sample variances and covariances whereas ABB imputations only added bias to estimates of covariance. As a result, the NASS method may inadvertently preserve ratios of sample covariances to sample variances (such ratios are used in calculating regression coefficients via least squares). Table 3 also shows that the exhaustive and parsimonious ISR imputations yield similar values of $\hat{\alpha}_d$ and $se(\hat{\alpha}_d)$. The primary distinction between the results is the values of B. The exhaustive model yields a small between-imputation variance (in fact, the confidence interval corresponding to the exhaustive model is contained within the confidence interval corresponding to the parsimonious model), suggesting that the imputations from the more comprehensive model contain less error. Note also that the point estimates of α_d under the ABB method and the deficient ISR model lie outside the interval estimates for that parameter from the parsimonious and exhaustive ISR models. This result illustrates that the choice of imputation method and model can profoundly influence the inferences drawn.

Table 4 provides results for quantities relevant to the estimation of β_d in equation 3 (per-acre value of land rented from others), which mimic those presented in Table 3. The values of $\hat{\beta}_d$ show even greater sensitivity to imputation method than values of $\hat{\alpha}_d$—perhaps because of the increase in the missingness rate (in equation 2, the dependent variable has no missing values while 17 percent of the farms have missing values for the dependent variables in equation 3). Furthermore, there is again evidence that imputations created using the exhaustive ISR model contain the least error.

Both tables illustrate the significant impact that choice of imputation method can have on estimates of the interactions between farm subsidies and cash rents

Table 4. Regressions of the Per-Acre Value of Land Rented from Others on Direct Payment per Acre and Adjusted Net Farm Income per Acre

	Imputation Method				
			ISR		
Quantity	(1) NASS	(2) ABB	(3) Deficient	(4) Parsimonious	(5) Exhaustive
$\hat{\beta}_d$	16.78	16.51	15.08	18.13	19.47
$se(\hat{\beta}_d)$	0.902	0.994	0.929	1.046	1.030
B	—	1.887	1.021	0.905	0.585
L_{MI}	—	12.86	12.19	15.21	16.85
U_{MI}	—	20.17	17.96	21.05	22.07

Note: Estimates of regression coefficients and standard errors were combined across the ten imputed data sets using Rubin's combining formulas. Observations were weighted by the product of total acres and the sampling weight.
Source: ARMS (2008).

and land values, especially when a relatively large percentage of the data is imputed. But which estimates are closer to the truth? While we caution against interpreting any single estimate as the truth, we note that our estimates from the exhaustive ISR model are similar to estimates from other studies that used much larger panel data sets and employed a more complete set of controls. For example, Goodwin, Mishra, and Ortalo-Magné (2011) used farm-level data from the ARMS survey and county-level data from a variety of sources and found that a one-dollar increase in direct payment is associated with a \$0.72 increase in cash rents, which is comparable to our estimate of \$0.75 per acre but is not economically significantly different from the estimate of \$0.76 per acre using the NASS procedure. Using field-level panel data from the JAS and county-level data on federal payments to farms, Ifft, Kuethe, and Morehart (2013) found that an extra dollar of a decoupled payment (which included any direct payment) is associated with an increase in land value of \$17.72 per acre under their preferred specification. That estimate is well within the 95 percent confidence interval (\$16.85– \$22.07) for our preferred-method estimate of \$19.47 per acre based on the exhaustive ISR model. In contrast, our estimate using the NASS imputations (\$16.78 per acre) is outside the 95 percent confidence interval for the estimate from our preferred method.

What then is the value of ISR imputation methodology for applied researchers and policymakers? According to the results in Table 3, the additional computational cost associated with the exhaustive ISR method may not be worthwhile when a relatively small percentage of the data is imputed (e.g., 10 percent or less) and only one or two key variables contain imputed data. On the other hand, as shown in Table 4, a larger percentage of imputed data and the presence of imputed data in both the dependent variable and a key explanatory variable could make the exhaustive ISR method worthwhile. Our estimate of the per-acre value of land under the parsimonious ISR model (\$18.13 per acre) is closer to the estimate by Ifft et al. using their preferred method than to our estimate under the NASS procedure. Furthermore, for the results shown in Table 4, all three ISR models provide tighter 95 percent confidence intervals than the ABB method, which used the same donor pools as the NASS method. Intuitively, then, the estimates based on the ISR imputations have a smaller degree of uncertainty due to imputations for missing data. Estimates based on single imputations using the NASS method give researchers and policymakers a false sense of certainty because they fail to account for uncertainty associated with the imputation process.

Simulations

Tables 3 and 4 indicate that different imputation methods and models can yield substantially different estimates of regression coefficients. However, we do not know which value is closest to the truth. To provide guidance as to which estimates are most trustworthy, we execute a jackknife-type simulation study using the 2008 ARMS data to compare benchmark values of the parameters to the ones estimated using each imputation method.

Since the ARMS data set contains missing values for a number of variables, we cannot calculate benchmark coefficient values using only observed data. One option is to remove units that have a non-response in at least one pertinent variable. However, that approach yields a sample size that is too small because of the large number of variables used in this study. Therefore, we first generate a single set of imputations (using a method described later) that we use to create the completed benchmark data set of regression coefficients. We then randomly "poke holes"—introduce missing values—in 50 percent of the observations of cash rents and of value of land rented in the benchmark data set without regard for which values were missing originally. This missingness mechanism thus is "missing completely at random" (MCAR) (Little and Rubin 2002). For each missing value, we create an imputation using each of the methods in the study. Then, for each of the five completed data sets, we calculate MI points and interval estimates using our previous estimation procedures. That entire process is then repeated 249 times, generating 250 data sets with imputed values standing in for the simulated missing ones. Finally, we calculate values for each of the quantities listed in Tables 3 and 4 for each method.

Earlier exploratory studies similar to the one presented here and conducted using only fully observed ARMS data indicate that ISR is preferable to more rudimentary methods. Furthermore, other exploratory studies that used benchmark data that contained imputations derived via the NASS method also indicated that ISR was preferable. Therefore, to create the benchmark data set, we use ISR with a range of inputs that exceeds the inputs in our exhaustive model. Note, however, that the observed missingness rate of *DP* and *VLR* is much lower than the missingness rate we impose in the simulation study so the imputation method used to create the benchmark data set will not be of great consequence (this belief was verified by exploratory studies).

We are interested in determining how well each imputation method maintains the benchmark values for the regression coefficients and respective standard errors. Letting θ denote a quantity of interest (such as α_d, β_d, or the corresponding standard errors), we calculate

$$\Delta(\theta)^{(j)} = 100(\hat{\theta}^{(j)} - \hat{\theta}) / \hat{\theta},$$

which represents the percent change in θ when $\hat{\theta}^{(j)}$ is the value of θ estimated from the *j*th data set with simulated missingness and $\hat{\theta}$ is the value of θ calculated from the benchmark value.

We present the results as box plots for the 250 values of $\Delta(\theta)^{(j)}$ for each imputation method and relevant quantities of interest. In each plot, the vertical axis represents the percentage difference between estimates from the multiply imputed completed data and the corresponding estimate from the benchmark data. The thick dark line in each plot depicts averages of 100-percentage-point differences for a given imputation, and the upper and lower ends of the boxes show the upper and lower quartiles.

Our results for the cash rent model (equation 2) are presented in Figure 1. The lefthand plots in Figure 1 show the coefficients on direct payments (i.e., $\theta = \alpha_d$) and the righthand plots show the estimated standard errors of $\hat{\alpha}_d$. Corresponding results for the model of value of land rented (equation 3) are presented in Figure 2. As previously mentioned, both figures provide results for direct payments scaled by crop land acres and acres operated.

The results shown in the figures mirror the patterns seen in Tables 3 and 4. The figures thus verify the efficacy of the exhaustive ISR model and specific deficiencies of the other methods. Specifically, Figure 1 shows that the parsimonious ISR results in biased estimates of α_d. This is not particularly surprising. As mentioned earlier, the model in formula 2 incorporates complexities such as sample design weights and per-acre forms of dependent

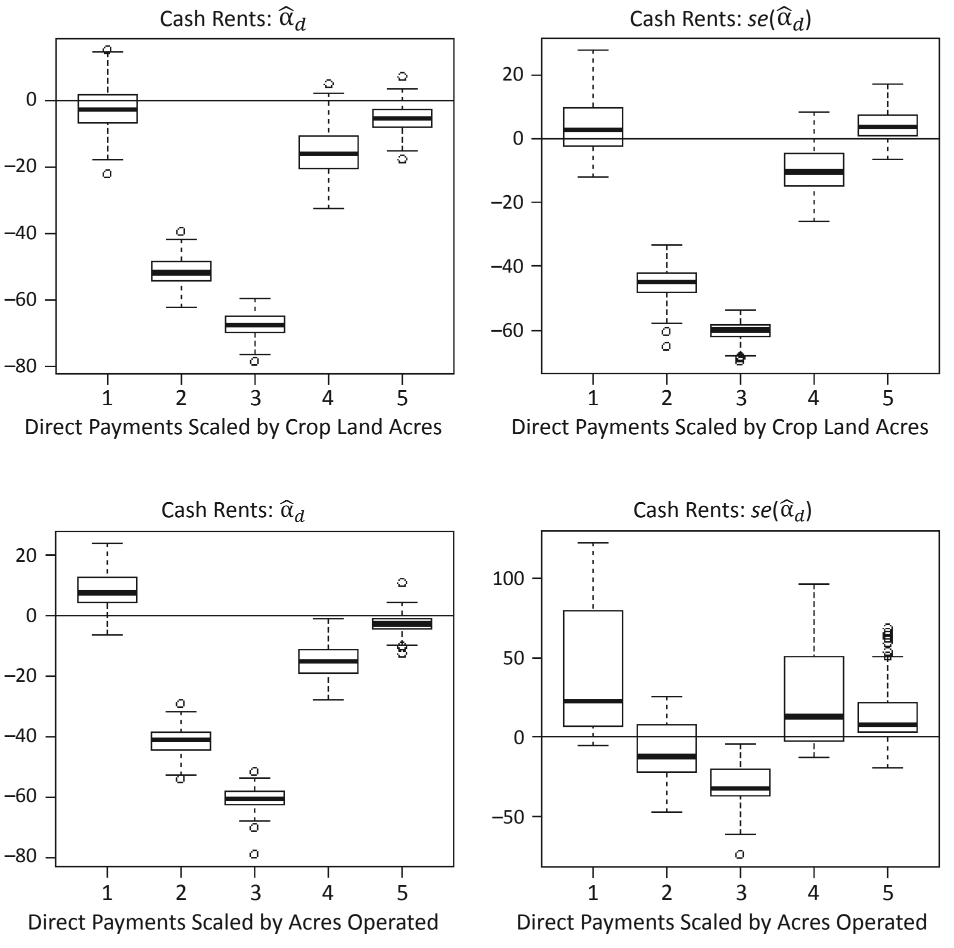

Figure 1. Box Plots of Percentage Differences between Cash Rent Model Estimates from Multiply Imputed Data and Fully Observed Data by Imputation Method

Note: Methods plotted: (1) NASS, (2) ABB, (3) ISR deficient model, (4) ISR parsimonious model, (5) ISR exhaustive model. For the plots in the top (bottom) row, estimators are calculated after scaling direct payments by crop land acres (acres operated).

and response variables that are not directly incorporated into the imputation procedure in the simulation. However, by expanding the depth of inputs and thereby garnering more accurate imputations, we can improve the reliability of estimates of the regression coefficients while using the same general imputation method.

We are also interested in gauging the appropriateness of the confidence intervals derived using MI in the simulation study so we monitor quantities that are specific to MI while running the simulations. Specifically, we track the value of the between-imputation variance, B, and the width of the 95 percent confidence interval estimated using MI for each method for each of the 250 runs of the simulation. Also of interest is the portion of runs in which the benchmark

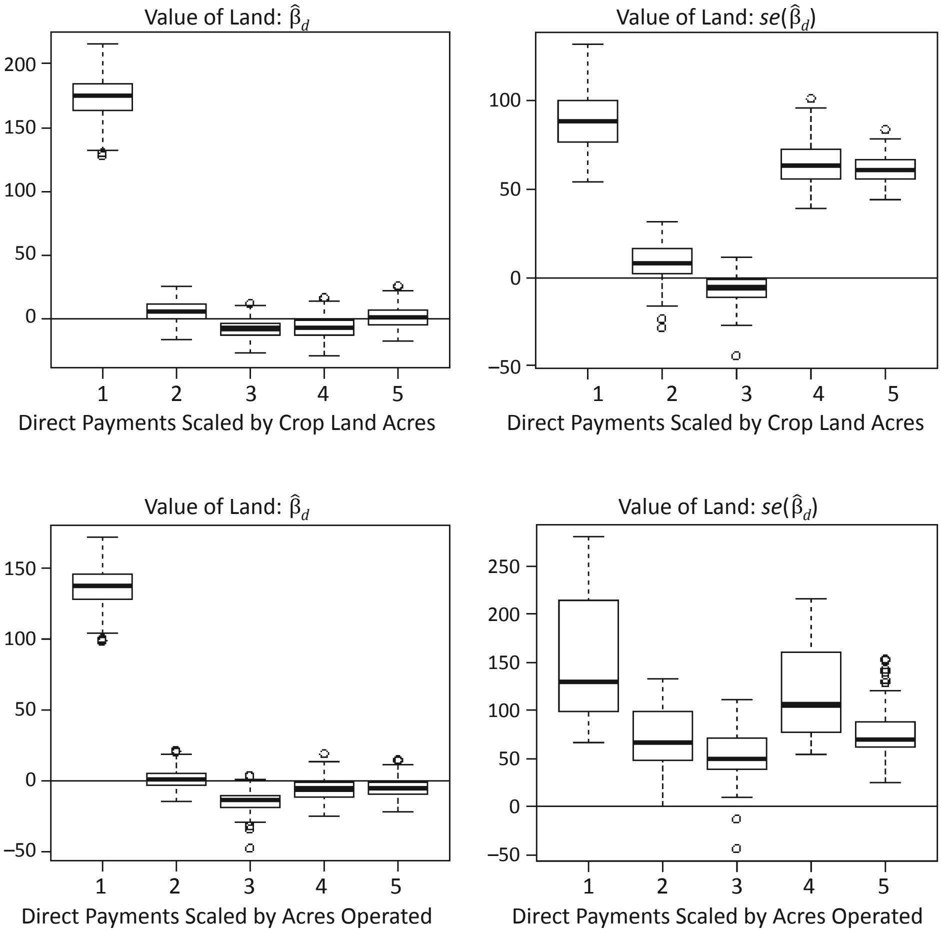

Figure 2. Box Plots of Percentage Differences between Value-of-Rented-Land-Model Estimates from Multiply Imputed Data and Fully Observed Data by Imputation Method

Note: Methods plotted: (1) the NASS method, (2) ABB, (3) ISR deficient model, (4) ISR parsimonious model, (5) ISR exhaustive model. For the plots in the top (bottom) row, estimators are calculated after scaling direct payments by crop land acres (acres operated).

value of the regression coefficient falls within the interval estimated using MI. We refer to that portion of runs as the estimated coverage probability of the interval estimate and denote it as $\hat{p}$. Table 5 provides the results of this analysis for the cash rent model and Table 6 provides the results for the land value model. The tables give average between-imputation variances and confidence interval widths across the 250 runs plus estimated coverage probabilities of the 95 percent confidence interval for the coefficients on direct payment (α_d and β_d). As in Figures 1 and 2, these results are for direct payments scaled by crop land acres and acres operated.

The results of this analysis confirm those presented in Tables 3 and 4: increasing the scope of an imputation model decreases the between-imputation variance (and, thus, the width of resulting interval estimates), which is indicative of more accurate imputations. Note that there is a discrepancy in between-imputation variance values from Tables 3 and 4 and Tables 5 and 6, a result of the fact that the missingness rates imposed in the simulation are much larger than the ones observed empirically. Higher missingness rates tend to yield higher values of between-imputation variance.

Table 5. Results for Multiple Imputation of Simulated Missingness for the Cash-Rents Model

	Imputation Method			
		ISR		
Quantity	(2) ABB	(3) Deficient	(4) Parsimonious	(5) Exhaustive
Acreage Measure: Acres Operated				
B	0.0033	0.0023	0.0038	0.0014
Width	0.3259	0.2655	0.3470	0.2273
$\widehat{p}$	0.0000	0.0000	0.3880	0.9280
Acreage Measure: Crop Land Acres				
B	0.0017	0.0011	0.0026	0.0008
Width	0.2396	0.1908	0.2944	0.1961
$\widehat{p}$	0.0000	0.0000	0.6120	0.9400

Note: Information provided includes average between-imputation variance (B) and interval width as well as estimated coverage probability ($\widehat{p}$) of the 95 percent confidence interval. Results are given using two separate acreage variables to scale direct payments.

Table 6. Results for Multiple Imputation of Simulated Missingness for the Value-of-Land-Rented Model

	Imputation Method			
		ISR		
Quantity	(2) ABB	(3) Deficient	(4) Parsimonious	(5) Exhaustive
Acreage Measure: Acres Operated				
B	7.311	5.764	5.326	4.639
Width	15.22	13.13	13.19	12.29
$\widehat{p}$	0.9880	0.8040	0.9320	0.9560
Acreage Measure: Crop Land Acres				
B	4.326	3.658	3.783	2.952
Width	11.71	10.62	11.25	10.11
$\widehat{p}$	0.9800	0.9720	0.9640	1.0000

Note: Information provided includes average between-imputation variance (B) and interval width as well as estimated coverage probability ($\widehat{p}$) of the 95 percent confidence interval. Results are given using two separate acreage variables to scale direct payments.

The results presented in Tables 5 and 6 also illustrate that confidence intervals calculated using MI with imputations sampled from the exhaustive ISR model are likely to observe the appropriate coverage (i.e., $\hat{p} \approx 0.95$). This observation is of particular interest since the theory that validates the MI interval estimates does not hold in these circumstances because of lack of congeniality between the imputer's and analyst's models.

Conclusions

We investigate the effect of various imputation methods on regression coefficient estimates and standard errors in two models of the effect of direct payments on cash rents and the market value of land using data from the ARMS. We replace NASS's single imputations with multiple imputations created using the ABB method plus the recently developed ISR method applied with three levels of input. We find that regression coefficient estimates and standard errors differ significantly based not only on the imputation method used but also on the depth of the imputation model used in ISR. Our comparison of estimates using the official USDA imputations with estimates using our preferred imputation method and model—exhaustive ISR—points to an additional $52.64 in land value per acre associated with direct payments. At a national level, that could translate to a $13.7 billion increase in land valuations associated with direct payments.

We also simulate missing data in fully observed cases and compare regression coefficient estimates and standard errors based on ABB and ISR multiple imputations to estimates based on the original fully observed cases. We find that the ISR method consistently produces regression coefficient estimates and standard errors with significantly less bias than those based on ABB imputations. Furthermore, we make the surprising observation that parsimonious ISR may produce biased point estimates of regression coefficients while no such bias is seen in the exhaustive imputations. Likewise, the exhaustive imputation model yields interval estimates that have a higher level of precision (i.e., a smaller width) and appropriate coverage probabilities. These observations lead us to conclude that an exhaustive imputation model is indeed worth the computational cost in many cases. Furthermore, use of an exhaustive imputation model may speed up the rate of convergence of the Markov chain, thereby allowing for fewer MCMC iterations, which may help compensate for the increased computational time required by an exhaustive model to some degree.

References

Banker, D. 2007. "ARMS Phase III: Data Processing and Analysis." Working paper, Economic Research Service, USDA.

Barnard, C., R. Nehring, J. Ryan, and R. Collender. 2001. "Higher Cropland Value from Farm Program Payments: Who Gains?" *Agricultural Outlook* 2001(Nov): 26–30.

Floyd, J. 1965. "The Effects of Farm Price Supports on the Returns to Land and Labor in Agriculture." *Journal of Political Economy* 73(2): 148–158.

Gardner, B. 1992. "Changing Economic Perspectives on the Farm Problem." *Journal of Economic Literature* 30(1): 62–101.

Goodwin, B.K., A.K. Mishra, and F. Ortalo-Magné. 2011. "The Buck Stops Where? The Distribution of Agricultural Subsidies." NBER Working Paper 16693, National Bureau of Economic Research, Cambridge, MA.

Habiger, J., M. Robbins, and S. Ghosh. 2010. "An Assessment of Imputation Methods for the USDA's Agricultural Resource Management Survey." *JSM Proceedings: Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
Huber, P.J., and E.M. Ronchetti. 2009. *Robust Statistics* (2nd edition). Hoboken, NJ: John Wiley & Sons.
Ifft, J., T. Kuethe, and M. Morehart. 2013. "The Impact of Decoupled Payments on U.S. Cropland Values." Working Paper, Economic Research Service, USDA, Washington, DC.
Ifft, J., C. Nickerson, T. Kuethe, and C. You. 2012. "Potential Farm-level Effects of Eliminating Direct Payments." Economic Information Brief 103, Economic Research Service, USDA, Washington, DC.
Kirwan, B.E. 2009. "The Incidence of U.S. Agricultural Subsidies on Farmland Rental Rates." *Journal of Political Economy* 117(1): 138–164.
Kott, P.S. 1995. "A Paradox of Multiple Imputation." Working paper, National Agricultural Statistics Service, USDA, Fairfax, VA.
Kuchler, F., and A. Tegene. 1993. "Asset Fixity and the Distribution of Rents from Agricultural Policies." *Land Economics* 69(4): 428–437.
Little, R.J.A., and D.B. Rubin. 2002. *Statistical Analysis with Missing Data* (2nd edition). Hoboken, NJ: John Wiley & Sons.
Miller, D., M. Robbins, and J. Habiger. 2010. "Examining the Challenges of Missing Data Analysis in Phase Three of the Agricultural Resource Management Survey." *JSM Proceedings: Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
Robbins, M.W., S. Ghosh, B. Goodwin, J. Habiger, D. Miller, and T.K. White. 2011a. "Multivariate Imputation Methods for Addressing Missing Data in the Agricultural Resource Management Survey." Working paper, National Institute of Statistical Sciences, Research Triangle Park, NC.
———. 2011b. "Multivariate Imputation Methods for Addressing Missing Data in the Agricultural Resource Management Survey." NISS/NASS collaborative research project, National Agricultural Statistics Service / National Institute of Statistical Sciences, Research Triangle Park, NC.
Robbins, M.W., S.K. Ghosh, and J.D. Habiger. 2013. "Imputation in High-dimensional Economic Data as Applied to the Agricultural Resource Management Survey." *Journal of the American Statistical Association* 108(501): 81–95.
Robbins, M.W., and T.K. White. 2011. "Farm Commodity Payments and Imputation in the Agricultural Resource Management Survey." *American Journal of Agricultural Economics* 93(2): 606–612.
Roberts, M.J., B. Kirwan, and J. Hopkins. 2003. "The Incidence of Government Program Payments on Agricultural Land Rents: The Challenges of Identification." *American Journal of Agricultural Economics* 85(3): 762–769.
Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons.
Rubin, D.B., and N. Schenker. 1986. "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse." *Journal of the American Statistical Association* 81(394): 366–374.
Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*. New York, NY: Chapman and Hall / CRC.
Schafer, J.L., and J.W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7(2): 147–177.
Schenker, N., T.E. Raghunathan, P.L. Chiu, D.M. Makuc, G. Zhang, and A.J. Cohen. 2006. "Multiple Imputation of Missing Income Data in the National Health Interview Survey." *Journal of the American Statistical Association* 101(475): 924–933.
Tanner, M.A., and W.H. Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation (With Discussion)." *Journal of the American Statistical Association* 82(398): 528–550.
U.S. Department of Agriculture. 2008. "Land Values and Cash Rents 2008 Summary." Summary Report, National Agricultural Statistics Service, USDA, Washington, DC.
White, T.K., and R. Hoppe. 2012. "Changing Farm Structure and the Distribution of Farm Payments and Federal Crop Insurance." Economic Information Brief 91, Economic Research Service, USDA, Washington, DC.