Who Does Not Respond to the Agricultural Resource Management Survey and Does It Matter?

Jeremy G. Weber and Dawn Marie Clay

The Agricultural Resource Management Survey is the primary annual source of information on U.S. farms, but in a typical year one-third of sampled farms do not respond. We use Census of Agriculture data to study nonresponse to the survey and how it affects estimates in two econometric models. Despite larger farms responding less, the coefficients estimated from the respondent subsample always fall inside confidence intervals based on draws from the full sample of respondents and nonrespondents. Although nonresponse bias can vary by application, the findings suggest that bias is unlikely to undermine conclusions based on econometrics using respondent data.

Key words: Agricultural Resource Management Survey, nonresponse bias, unit nonresponse.

JEL codes: C81, C80, Q10

The Agricultural Resource Management Survey (ARMS) is the primary annual source of information on U.S. farm businesses and the households who operate them. The survey targets roughly 30,000 farms each year and collects information on practices, production, and finances. Since the survey's inception in 1996, researchers in academia and government have used the responses to study U.S. agriculture and farm and conservation policy, as evidenced by hundreds of reports and journal articles. Despite its central role in research, roughly one-third of sampled farm operators ignore the survey entirely – an occurrence known as unit nonresponse. The unit response rate is well below the 80% level at which the Office of Management and Budget requires that the agency administering the survey conduct a nonresponse bias analysis (U.S. Office of Management and Budget, 2006a, p. 8). Although low unit response rates do not imply nonresponse bias, the lower the rate, the greater the effect that differences between respondents and nonrespondents will have on estimates based on respondent-only data (Groves 2006).

The National Agricultural Statistics Service, which administers ARMS, has researched how unit nonresponse affects estimates of unconditional means by using data from the Census of Agriculture to compare means of Census of Agriculture variables calculated from a full ARMS sample (respondents and nonrespondents) with means calculated from a sample of only respondents (Earp et al. 2008a, 2008b, 2010). The Census of Agriculture provides an ideal opportunity for assessing unit nonresponse bias because it provides information on ARMS respondents and nonrespondents that comes from the same questionnaire collected at the same time for both groups. Relying solely on Census of Agriculture data, we expand the initial work by the National Agricultural Statistics Service by exploring the motivations and characteristics associated with unit nonresponse (hereafter referred to as "nonresponse"), and how point estimates in two econometric models differ when estimated on a respondent subsample and a subsample randomly drawn from respondents and nonrespondents.

Figure 1 shows the response rates for ARMS Phase III for 2003–2006 and 2008–2010 for different deciles of farms ordered by their

Jeremy G. Weber is an economist at the Economic Research Service, U.S. Department of Agriculture. Dawn Marie Clay is a graduate student at North Carolina State University who worked on this project while interning at the Economic Research Service. The views expressed here are the authors' and should not be attributed to the Economic Research Service or the USDA. We thank Tom Hertz, Jason Brown, two anonymous reviewers, and the managing editor for providing very helpful comments. No external funding supported this research.

Amer. J. Agr. Econ. 95(3): 755–771; doi: 10.1093/ajae/aas171
Published online February 4, 2013

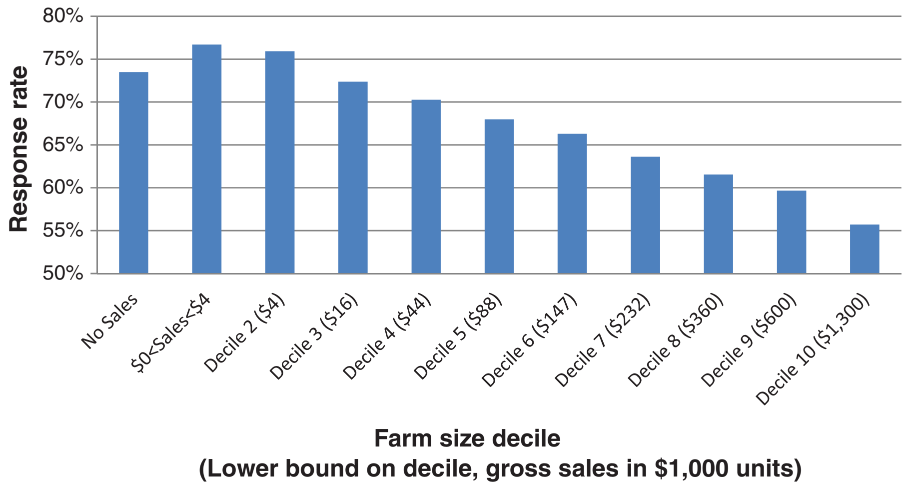

Figure 1. Response rates decrease with farm size (Gross Sales, in $1,000 units)

gross sales. Except for the no sales category, response rates decrease monotonically with farm size. Further study confirms that even after controlling for other farm and household characteristics, farm operators who do not respond have substantially greater sales than respondent operators, part of which reflects that completing the survey takes longer for operators of larger farms.

Despite the observed differences between respondents and nonrespondents, we find minimal nonresponse bias in the two econometric models estimated. The coefficients estimated from the respondent sample always fall inside the confidence intervals generated by repeatedly drawing from the full sample of respondents and nonrespondents. A complementary contribution is the treatment of standard errors where the traditional delete-a-group jackknife variance estimation is either unfeasible or undesirable. Clustering by stratum provides errors much larger than when ignoring the sample design. Clustering by groups created by interacting farm size and location, two attributes used to define ARMS strata, gives errors similar in size to those when clustering by stratum.

Confronting Survey Nonresponse

When an operator only refuses some questions (item nonresponse), information from answered questions can be used to impute values for refused questions, which is standard practice for managers of surveys like the Current Population Survey and the Agricultural Resource Management Survey (Lillard, Smith, and Welch 1986; Ahearn et al. 2011). But when the entire survey is refused, the lack of information on nonrespondents precludes such imputation.

Missing data can be described as missing completely at random (MCAR), where nonresponse is unrelated to a household's unobserved characteristics; missing at random (MAR), where data are missing at random conditional on observed characteristics; or missing not at random (MNAR), where missingness is related to unobserved characteristics even after conditioning on observed ones (Hedeker and Gibbons 2006). If nonrespondent data are MCAR, information from respondents can be used to consistently estimate population means and econometric models. After conditioning on observable variables, the same is true in the MAR case. But unless the sample was drawn from a list frame with ample information on all potential observations, researchers will lack the information on nonrespondents needed to assess whether observations are missing at random.

If households are MNAR, nonresponse can have varying effects on inferences derived from information provided by respondents alone. Abraham, Maitland, and Bianchi (2006) study nonresponse in the American Time Use Survey, and find that incorporating nonresponse propensities into sample weights had modest effects on estimates of time devoted to different activities and often had no effect at all. In contrast, Korinek, Mistiaen, and Ravallion (2006) find that response rates to the Current Population Survey decrease with income. Correcting for the non-random nature of response for one year increased a measure of income concentration, the Gini coefficient, by almost six percentage points (standard error of 1.25) above the uncorrected coefficient.

Nonresponse may also affect estimates of relationships between variables that have economic interpretations. Such nonresponse bias can be understood in the framework of sample selection models commonly used in labor economics. A wage function estimated using only people who work for a wage may poorly approximate the population wage function (Heckman 1979). The decision to work is likely correlated with an unobserved variable such as ability, which would imply that on average, people who work have greater ability than those who do not. Ignoring differences in ability could bias estimates of parameters like the returns to education.

Just as people choose to work, they also choose to respond to surveys. Using the Current Population Survey, Lillard, Smith, and Welch (1986) estimate a two-part model for labor market participants where the first part is an equation for the propensity to report earnings and the second part is an earnings equation, where the earnings for people who did not report them are imputed by the Census Bureau. The results from the reporting equation reveal that people with higher earnings are less likely to report them. Furthermore, joint estimation of the two equations shows that the error terms in the two equations are highly correlated, implying that omitted variables affecting reporting are correlated with omitted variables affecting earnings.

Nonresponse in the Agricultural Resource Management Survey and its predecessor

Low response rates have plagued U.S. agricultural surveys for some time. Before developing the Agricultural Resource Management Survey, the USDA collected farm-level data through the Farm Cost and Returns Survey. Starting in 1984, the survey ran for 11 years and had a typical response rate of 70% (Johnson, Baum, and Prescott 1985; O'Connor 1992). Even more so than its predecessor, ARMS asks many sensitive questions like household assets and debt, and is long. It has two main versions, the Core and the Cost and Returns Report. In 2010 the Core had 16 pages, and over the last three years it has taken the average principal operator one hour and seven minutes to complete. The more exhaustive Cost and Returns Report had 32 pages in 2010 and takes an average of one hour and 36 minutes to complete.

The National Agricultural Statistics Service (NASS) has worked to address nonresponse for more than a decade. As early as 1991, NASS researched reasons for nonresponse in the Farm Cost and Returns Survey (O'Connor 1992). More recently, NASS has identified nonresponse bias as the largest source of non-sampling error (National Research Council 2008, p. 107). To increase response rates, NASS has experimented with providing US$20 gift cards and sending farm operators a prescreening letter that includes an economic brief of results from past ARMS. Gift cards appear to be the most effective: in one experiment the cards increased the response rate by more than 13% (Beckler, Ott, and Horvath 2005).

NASS has used Census of Agriculture data on ARMS respondents and nonrespondents to assess bias in unconditional means of variables of interest. Considering the 2005 and 2006 ARMS, Earp et al. (2008a, 2008b) compare means of Census of Agriculture variables across three groups: matched records (ARMS respondents and nonrespondents who responded to the 2002 Census of Agriculture); matched respondents; and matched respondents with weights calibrated to match population values taken from the Census of Agriculture. The research found that the matched sample means for sales, production expenses, and acres operated all exceeded those from the respondent-only sample, implying that nonrespondents have larger farms than respondents.

The research also found that calibrating weights to known targets like the total acres of corn – the method used to adjust the final weights for ARMS – and applying them to ARMS respondent data from the Census of Agriculture produced means close to those from the full matched sample (ARMS respondents and nonrespondents). To be clear, the ARMS nonresponse bias research by NASS relies solely on data from the Census of Agriculture, but assuming that the bias implied by Census of Agriculture data is similar to the bias in ARMS data, the findings suggest that in many cases calibration can yield unbiased estimates of unconditional means using data from only respondents. However, more recently a similar exercise using the 2008 ARMS and the 2007 Census of Agriculture performed less well, with several variables having substantial bias even after calibration (Earp et al. 2010). While NASS's current calibration techniques applied to Census of Agriculture data appear to mitigate much nonresponse bias in unconditional means for select farm production and expense variables, we focus on the motivations for and patterns in nonresponse and the possible bias that it introduces into econometric models.

Reasons for Nonresponse

Potential reasons for not responding to ARMS abound. Attitudes towards the government and satisfaction with the farm or regulatory policy governing an operator's commodities can influence her propensity to respond. The most common reason for refusing the Farm Cost and Returns Survey in 1991 – accounting for one-quarter of all refusals – was "Would not take the time / too busy," (O'Connor 1992). A similar categorization of reasons for nonresponse in the 2006 ARMS in the state of Washington had the same finding (Gerling, Tran, and Earp 2008). The "too busy" reason, however, can mask several underlying motivations for refusing the survey. Responding does take time away from other activities, but farm operators may also find responding to be tedious. Refusing to take an hour to respond may be less about the shadow value of the hour and more about the disutility from answering.

Other top reasons for nonresponse in the 2006 ARMS in Washington were "Will do other surveys, but not financial surveys" and "Information too personal / none of your business," (Gerling, Tran, and Earp 2008). The explanations possibly reflect the unwillingness of farm operators to provide private information that may become less valuable if given and used to calculate aggregate statistics. Similarly, some operators may be concerned that federal agencies like the Internal Revenue Service or the Environmental Protection Agency might access the information and use it against them, despite the survey's disclaimer that responses are confidential and used for statistical purposes only.

The time and disutility of responding will tend to increase with the size and complexity of the farm. Looking at ARMS response times over the last three years reveals that farm operators in the largest sales decile took about 55% longer (36 minutes) to complete the survey than operators of farms with no sales. Furthermore, it is plausible that larger operations have more private information that they would like to keep private, especially since the operation's influence on aggregate statistics will increase with its size.

Trends in farm structure likely also affect patterns in response propensities. Production has increasingly moved to larger operations, which often encompass multiple enterprises whose legal and contractual complexity make collecting reliable data more difficult (Complex Agricultural Establishments 2011). A farm with multiple owners or managers can mean that several people must consent to responding. Consequently, a sole proprietor may find responding easier than an operator of a farm with multiple equity holders or managers. Furthermore, as an organization increases in complexity it is more likely that some survey questions may be difficult to answer or inappropriate and thereby frustrate operators (Green 2011). For example, a hired manager may make most day-to-day management decisions on a farm owned by an entrepreneur with several businesses. When the survey asks about the principal operator's household and allocation of time, there may be confusion over whose information to use – that of the manager or the business owner.

The Agricultural Resource Management Survey in Detail

ARMS has a complicated design, employing multiple strata (e.g. farm type), modes (phone, mail, and face-to-face enumeration), and phases (I, II, and III). Phase I is a telephone screening to verify that an operation belongs in the sample. Phase II occurs at the field level and targets particular crops, which vary from year to year, and collects information on production practices and chemical use. Phase III gathers information on the farm enterprise and the household associated with its principal operator.

Phase III has three to five versions, depending on the year. The first version is sent to farms of varying commodities, is enumerated in person, and uses the detailed Cost and Returns Report as the instrument. Versions II and III (and sometimes IV) apply a commodity-specific version of the Cost and Returns Report to operations producing the targeted commodity. Crop farms surveyed in Phase II would receive the version that corresponds to their commodity. Version V always corresponds to a mail survey and is a simplified version of the Cost and Returns Report known as the Core.

Having drawn and screened the Phase I sample, NASS contemporaneously draws the Phase II and Phase III sample. The Phase III target sample includes all operators selected for Phase II, and additional operators selected for Phase III only. With 100% compliance in both phases, the total Phase III sample would equal the Phase II sample, plus the Phase III operators not in Phase II. Consequently, increasing the Phase II sample by one will increase the Phase III sample by one, and if an operator refuses Phase II, he effectively refuses Phase III since NASS will not pursue him for Phase III. Because a Phase II nonresponse translates into a Phase III nonresponse, we treat nonresponse in Phase II as an implicit Phase III nonresponse and use the term 'Phase III nonresponse' to refer to the combined group. While some refusals occur in the Phase I screening, we focus on Phase III nonresponse.

The Data

The National Agricultural Statistics Service attempts to collect data on all farms and their operators every five years through the Census of Agriculture.[1] In those years, NASS embeds the ARMS in the Census of Agriculture, so if an operator refused ARMS in 2002 or 2007, which were Census of Agriculture years, no information on the operation exists. Though we can analyze nonresponse in six of the eight most recent ARMS (2003–2006, 2008–2010), relying solely on Census of Agriculture data implies that information on ARMS respondents and nonrespondents was collected a year or more prior to when operators received the ARMS. This is unlikely to render the Census of Agriculture data uninformative since most of the employed variables change slowly over time (e.g. age of the operator, persons in the household, commodity specialization). Furthermore, changes over time would have to affect respondents and nonrespondents differently to potentially lead to qualitatively different conclusions about nonresponse. Consequently, it is common for studies of nonresponse to rely on data collected prior to when the nonresponse occurred (Kennickell and McManus 1993; Lin and Schaeffer 1995; Zabel 1998; Earp et al. 2008a, 2008b, 2010).

Table 1. ARMS Sample – Census of Agriculture Match Rates by Year, 2003–2006 and 2008–2010

Census	Year	Total N	Matched N	Match Rate (%)
2002	2003	30,490	25,046	82.2
	2004	31,138	26,256	84.3
	2005	33,567	27,220	81.1
	2006	33,052	26,407	79.9
2007	2008	33,309	31,670	95.1
	2009	31,863	25,848	81.1
	2010	33,896	27,027	79.7
Total		227,315	189,474	84.0

From NASS data we obtained the principal operator identifier of all farm operators targeted for Phase III of ARMS for the years 2003–2006 and 2008–2010. Summing across years yields 227,315 observations. We then use the principal operator identifier to match the 2003–2006 ARMS samples to the 2002 Census of Agriculture, and the 2008–2010 samples to the 2007 Census of Agriculture. The overall match rate was 84.0%, leaving a total of 189,474 matched observations, of which 67% are ARMS respondents, 28% are refusals, and 5% are inaccessible (a nonresponse is not necessarily a refusal; nonresponse also occurs because the enumerator could not locate the operator, a situation that NASS labels "inaccessible"). Table 1 presents match rates by year. Match rates less than 100% reflect census nonresponse or operators who entered the population, and therefore the list frame for ARMS, after the census occurred. Alternatively, re-organizing agricultural enterprises could cause the identifier to change, which would also preclude making a match.

Dropping unmatched cases could bias the inferences we make using only the matched sample. We would be particularly concerned if the unmatched records were mostly respondents or mostly nonrespondents. Though ARMS respondents had higher match rates than nonrespondents, the difference is small; match rates for ARMS respondents,

[1] The Census of Agriculture attempts to reach all agricultural operations that produce, or would normally produce and sell, $1,000 or more of agricultural products per year. Data are primarily collected through the mail, with supplemental reporting conducted on the internet, and non-response follow-ups by telephone and personal enumeration. The final response rate was 85.2% for the 2007 Census of Agriculture, and 88.0% for the 2002 Census of Agriculture. It could be argued that we should also control for nonresponse in the Census of Agriculture, which we ignore since we have no information on operators who did not respond to ARMS and the Census.

refusals, and inaccessible were 85%, 81%, and 74%.

As with ARMS, NASS ensures that there are no missing values for some Census of Agriculture variables while allowing missing values for others. Except for corn yields and the value of production contracts, which are conditional on engaging in that activity, we focus on key variables with few or no missing values.

We group the variables into three categories: household characteristics, farm characteristics, and farm specialization (table 2). The farm specialization category contains binary variables that indicate a farm's specialty as reflected by its sales, including the value of production contracts. If a farm has no sales, it is classified as such by the binary variable *No Sales*.[2] If more than 50% of a farm's sales are from one category, the farm is classified as being specialized in that category. If not, and the farm had some sales, it is categorized as diversified. Because the Census of Agriculture collects the total payment received for providing commodities under production contracts and not the value of individual commodities, we treat production contracts as a specialization category.

Patterns in Nonresponse

We start exploring patterns in nonresponse by testing for differences in means by respondent status. If nonresponse occurs completely at random, respondent and nonrespondent operators should share similar characteristics and, at an appropriate confidence level, be statistically indistinguishable from each other. In addition to the 26 variables defined in table 2, we include in our comparisons nine binary variables indicating a farm's resource region; the regions are defined by the Economic Research Service according to crop reporting districts and farm characteristics like crop mix.[3]

Of the nonrespondent records matched to Census of Agriculture records, 16% are labeled as inaccessibles. Although an operator may implicitly refuse the survey by purposefully making himself hard to find, it is also possible that the operator was away on vacation or spends little time at the address associated

Table 2. Variable Descriptions

Household Characteristics	Variable Description
Age	The age of the farm's principal operator
Persons in household	The number of persons in the principal operator's household
Worked off farm	0/1 indicating if the principal operator works off the farm
Primary occupation	0/1 if the principal operator's primary occupation is farming or ranching
Percentage income from operation	The percentage of household income that comes from the farm
Farm Characteristics	
Sole proprietorship	0/1 indicating that the farm is an individual or family operation, excluding partnerships
Hired manager	0/1 indicating that the farm has a hired manager
Corn yield	Bushels of corn produced per acre, conditional on growing corn
Uses production contracts	0/1 indicating that the farm delivered commodities under production contracts
Production contract sales	The value of commodities delivered under production contracts, conditional on using production contracts
Total sales	The gross value of agricultural products sold from the farm, including production contracts
Land owned	Acres of land owned by the farm
Cropland harvested	Acres of land harvested by the farm
Farm Specialization	
Grains	0/1 indicating that sales of grains, oilseeds, dry beans, and dry peas compose more than 50 percent of sales
Tobacco	" " sales of tobacco " "
Cotton	" " sales of cotton and cottonseed " "
Vegetable and fruits	" " sales of fruits and vegetables " "
Swine	" " sales of hogs and pigs " "
Dairy	" " sales of milk and other dairy products " "
Cattle and calves	" " sales of beef and dairy cattle " "
Sheep, goats, horses	" " sales of sheep, goats, horses, and their products " "
Poultry	" " sales of poultry and eggs " "
Other	" " sales of horticulture, aquaculture, other animals and their products, etc. " "
Production contracts	" " sales involving production contracts " "
Diversified	" " sales of no category " "
No sales	0/1 indicating that the farm had no sales

[2] To be considered a farm, an operation only needs to produce, or normally produce and sell, $1,000 or more of agricultural products in a year; it does not need to have $1,000 in actual sales, or indeed any sales at all (O'Donoghue et al. 2009)

[3] See www.ers.usda.gov/publications/aib760/aib-760.pdf for a map of the region and details on its construction.

with the farm, in which case the data are more likely to be MCAR or MAR.[4] We therefore break the nonrespondent group into refusals and inaccessibles for descriptive comparisons.

The ARMS survey weights are unavailable for nonrespondents, but we know each farm's survey stratum, which is based on the farm's state, gross sales, and commodity. To capture the non-random sampling design of ARMS, we calculate standard errors for the difference in group means (and in the econometrics to follow) using robust standard errors clustered by stratum. Clustering by stratum assumes dependence between farms of the same stratum, and independence between farms in different stratum. The robust clustered estimator of the covariance matrix is $V = (X'X)^{-1} \sum_{j=1}^{nc} u_j' u_j (X'X)^{-1}$, where $u_j = \sum e_i x_i$ for all i in cluster j. In the simplest case of regression on a constant (estimating the mean), the estimate of the variance would be $n^{-2} \sum_{j=1}^{nc} \left(\sum_J e_i \right)^2$. Comparing this with the unclustered variance estimate $(n^{-2} \sum_{i=1}^{n} e_i^2)$ shows how clustering calculates the total variance by summing the variances of J clusters, while the unclustered estimate treats each observation as its own cluster.[5] For mean comparisons, the clustering is implemented by regressing the variable being compared (e.g. *Age*) on a constant and a binary variable indicating the observation's group (e.g. respondent or nonrespondent). The standard error of the coefficient on the binary group variable permits us to assess the statistical significance of the difference in group means.

Although we do not have the ARMS weights, we calculate our statistics and tests using the Census of Agriculture weights provided by NASS. To account for farms not on its mailing list or that do not respond, NASS assigns each farm a probability weight that reflects how many farms each census respondent farm represents. The weights help to ensure that different subpopulations are properly represented. In the case of unit nonresponse, an algorithm classifies farms based on observed characteristics and response rates. The weights of the responding farms in each group are then calculated to account for nonrespondent farms in the group (Cecere 2009).

Respondent and refusal group means are statistically different for 22 of the 35 variables (table 3). Although we use a high confidence level (99%) for our mean comparisons, which is appropriate given the large number of observations, a statistical difference in means does not imply that the differences are economically meaningful. With many observations – and we do have many – very small differences in means may still be statistically different from zero. However, many of the differences are significant in both senses of the word. The discussion of reasons for nonresponse suggested that operators of larger farms are likely to have lower response rates, and the descriptive statistics bear this out: refusal farms have more sales (\$902,327 compared to \$518,934) and own more land (906 acres compared to 627 acres). They are also more likely to be located in the Heartland (29.5% compared to 23.2%) or the Northern Great Plains (7.2% to 4.7%) – two regions with many large farms. Refusal households also derive more of their income from the farm (57.6% compared to 49.5% for respondent households).

Similarly, the means for respondents and inaccessibles are statistically different from zero for 18 of the 35 variables (results not shown). Operators who were inaccessible have farms with more sales (\$763,651) and own more land (946 acres) than respondent farms. Inaccessible operators are also more likely than respondents to be located in the Fruitful Rim (23.9% compared to 18.0%), which primarily covers Florida, the southern border of Texas, and most of the Pacific coast. Farms in areas with temperate climates may be more likely to serve as partial-year residences, making it more likely that the operator cannot be located.

Because the largest farms account for a disproportionate share of the sector's crop and livestock production, variables related to production have skewed distributions. We therefore test for difference in medians for *Land owned*, *Cropland harvested*, and *Total sales* between respondents and refusals, and between respondents and inaccessibles. In both comparisons, the medians were statistically different for the four variables. More interesting was the size of the differences: the median refusal operator harvested more than twice as many acres as the median respondent operator (307 compared to 132), and the difference in sales was almost as large (\$229,130 compared to \$120,454). Inaccessible operators also had

---

[4] It is also possible that some farms that went out of business between the Census of Agriculture and ARMS were labeled as inaccessibles. It is not clear if these 'out-of-business' inaccessibles would be small or large farms; mid-sized farms (70 to 999 acres) had the highest exit rates in the last decade (Hoppe and Banker 2010).

[5] To adjust for finite sample properties, the clustered variance estimate is multiplied by $\frac{N}{N-K} * \frac{J}{J-1}$, where J is the number of clusters and K the number of regressors.

Table 3. Mean Comparisons, Respondents, and Refusals

Variables	Mean – Respondents	Mean – Refusals	P Value	Percentage Difference
Household Characteristics				
Age	55.0	53.8	0.000	2.3
Persons in household	2.85	2.916	0.003	2.3
Worked off-farm	0.437	0.379	0.000	13.3
Primary occupation	0.756	0.814	0.001	7.7
Percentage income from operation	49.5	57.6	0.000	16.3
Farm Characteristics				
Sole proprietorship	0.755	0.717	0.003	5.0
Hired manager	0.056	0.067	0.001	19.3
Corn yield	123	128	0.000	4.8
Uses production contracts	0.102	0.078	0.025	23.7
Production contract sales	203,714	415,852	0.115	104.1
Total sales	518,934	902,327	0.016	73.9
Land owned	627	906	0.000	44.5
Cropland harvested	501	780	0.000	55.8
Farm Specialization				
Grains	0.244	0.341	0.000	39.5
Tobacco	0.015	0.009	0.000	40.2
Cotton	0.018	0.015	0.369	14.1
Vegetable and fruits	0.015	0.021	0.000	39.6
Swine	0.083	0.083	0.987	0.1
Dairy	0.103	0.113	0.235	9.1
Cattle and calves	0.205	0.164	0.000	19.7
Sheep, goats, horses	0.016	0.010	0.044	36.8
Poultry	0.007	0.007	0.653	5.1
Other	0.099	0.086	0.000	13.4
Production contracts	0.074	0.046	0.004	38.5
Diversified	0.067	0.071	0.191	6.1
No sales	0.053	0.034	0.064	35.8

Note: The percentage difference is the absolute value of the difference in means divided by the respondent group mean. All statistics are calculated using Census of Agriculture weights. The p values come from a regression of the variable being compared (e.g. *Age*) on a constant and a binary variable indicating the observation's group (e.g. respondent or refusal) and reflect the coefficient and standard error (clustered by stratum) of the group dummy variable. Monetary amounts are in 2007 dollars. Except in cases of missing values, the full sample used in the descriptive comparison includes 188,387 observations; 1,087 of the initial matched observations did not have a stratum and were excluded from the comparisons.

more land and sales, though not by as much ($16,410 more in sales and 18 acres more in harvested land).

The propensity to respond

To explore nonresponse patterns in a multivariate setting, we estimate a Probit model where the outcome is whether or not the operator responded to the survey. We include a subset of the variables used in the comparisons in the vector of right-hand side variables. For household characteristics, we include the age of the primary operator, the number of persons in the household, and the percentage of household income that comes from the farm. Attitudes towards government surveys and the value of time may vary with age and household size. Households less dependent on farming for their income likely have smaller, simpler farms that make responding easier. Because state agencies administer the ARMS, we also include state dummy variables to capture any differences across states that affect response propensities.

To capture a farm's scale and type, we include binary variables indicating the decile in the farm sales distribution that the farm belongs to (with the tenth decile excluded): whether it had no sales; whether it uses production contracts; and its commodity specialization.

To estimate the effect of response burden across the versions of ARMS, we group the versions by response burden: the Core (the least burdensome), the Version I Cost and Returns Report (modest burden), and the Version II–IV Cost and Returns Report (most burdensome). We include binary variables indicating if the farm received the Core or the Version I Cost and Returns Report.

We calculate robust standard errors clustered by stratum as was done for the group mean comparisons. To assess the effect of clustering errors by stratum we also calculate robust, unclustered errors, which amounts to ignoring the survey design by treating all observations as being independently drawn. We then cluster the errors based on two observable farm characteristics used to define strata – farm size as measured by gross sales and farm location. Location is given by the farm's region (in this case, USDA/ERS farm resource region) or state if the state is one of the 15 heavily-sampled states.[6] We create one grouping of farms by location and another by interacting location with sales class (less than \$10,000 in gross sales; \$10,000–\$249,999; and \$250,000 or more). In the full sample there are 47 distinct strata. When clustering by location there are 23 clusters, and when interacting location with sales class there are 69 clusters.

The multivariate analysis confirms most of the patterns revealed by the descriptive comparisons. Even after conditioning on farm and operator characteristics, response propensities decrease in a perfectly monotonic fashion when going further out in the distribution of farm size. Relative to operators in the largest sales decile, operators of farms in the fifth sales decile had a 0.118 higher probability of responding, while those with no sales had a 0.200 higher probability. Also consistent with the descriptive comparisons, farms specializing in grains (the excluded commodity group) had the lowest response propensities. In contrast, tobacco farms and dairy farms were most likely to respond.

Contrary to the intuition that farms with a more complex organizational scheme have lower response rates, sole proprietorships were less likely to respond than partnership or incorporated farms. The link between size and complexity, however, suggests that the sales class decile variables likely capture complexity. Indeed, re-estimating the model without the sales class variables reverses the result, with sole proprietorships being more likely to respond (marginal effect of 0.027, with a standard error clustered by stratum of 0.006).

As expected, operators who received the least burdensome version of the survey, the Core, were most likely to respond: they had a 0.037 higher propensity to respond than operators who received the Version I Cost and Returns Report. However, there was no clear difference in response propensities between the Version I Cost and Returns Report and the commodity-specific Cost and Returns Report.

Comparing standard errors reveals the importance of incorporating the sampling design into covariance estimates. For each variable in table 4, we calculate the ratio of the standard error to the standard error clustered by stratum. The typical robust but unclustered standard error is almost 40% smaller than the error clustered by stratum. Table 4 also shows that clustering errors based on observable characteristics related to the sampling design, either location or size-location groups, tend to provide standard errors only slightly smaller than those from clustering by stratum. For the variables listed in the table, the median ratio of the error when clustered by location is 0.94. When clustering by size-location groups it is 0.98.

To what extent can survey burden explain why larger farms respond less?

Conditional on many other variables, receiving the shorter Core version increases the propensity to respond by 0.035 relative to receiving the Version 1 Cost and Returns Report, which over the last three ARMS took an average of 30 minutes longer to complete than the Core. Assuming that the difference in response propensities stems from the length of the survey, the estimates imply that decreasing the time required to complete the survey by one hour would increase the propensity to respond by 7% (or 3.5% per half hour).

For the last three ARMS, farms in the largest sales category required on average 0.6 hours longer to respond than farms with no sales. The greater response burden therefore implies that the largest farms should have a 0.042 lower response propensity (0.6 hours $\times$ 7%) than farms in the smallest sales decile. The Probit model result showed that the largest farms had a 0.200 lower response propensity than farms with no sales. Thus, response burden captures 21% of the different response propensities between the smallest and largest farms. This likely underestimates the role of burden since the response times for the different groups are based on respondent operators, and the nonrespondent group likely includes many operators of larger farms who thought that responding would take a long time. Had

[6] These states are Arkansas, California, Florida, Georgia, Illinois, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Carolina, Texas, Washington, and Wisconsin. Only these 15 states receive the Core version of the survey.

Table 4. Marginal Effects from Probit Model of Response Status (1 = Respondent)

		Standard Error Clustered By …			
Variable	M.E.	Stratum	Nothing	Location	Location & Size
Core version	0.0379***	0.0086	0.0034	0.0123	0.0086
CRR version	0.0022	0.0076	0.0034	0.0057	0.0051
Age	0.0012***	0.0002	0.0001	0.0002	0.0002
Persons in household	0.0051***	0.0009	0.0009	0.0013	0.0012
Percentage income from operation	−0.0001**	0.0000	0.0000	0.0000	0.0000
Sole proprietorship	−0.0158***	0.0042	0.0029	0.0041	0.0040
Uses production contracts	0.0357***	0.0057	0.0067	0.0064	0.0066
Tobacco	0.0567***	0.0128	0.0119	0.0154	0.0234
Cotton	0.0202	0.0246	0.0094	0.0153	0.0127
Vegetable and fruits	0.0199**	0.0087	0.0088	0.0079	0.0088
Swine	0.0135	0.0082	0.0053	0.0081	0.0076
Dairy	0.0663***	0.0072	0.0046	0.0100	0.0092
Cattle and calves	0.0303***	0.0056	0.0039	0.0068	0.0062
Sheep, goats, horses	0.0182*	0.0108	0.0107	0.0131	0.0129
Poultry	0.0258**	0.0120	0.0137	0.0168	0.0184
Other	0.0336***	0.0054	0.0048	0.0079	0.0078
Production contracts	0.0501***	0.0065	0.0084	0.0114	0.0110
Diversified	0.0305***	0.0047	0.0050	0.0063	0.0062
No sales	0.2005***	0.0180	0.0074	0.0140	0.0134
Sales class 1	0.1911***	0.0114	0.0074	0.0143	0.0125
Sales class 2	0.1870***	0.0099	0.0060	0.0090	0.0092
Sales class 3	0.1565***	0.0089	0.0057	0.0107	0.0094
Sales class 4	0.1398***	0.0077	0.0054	0.0107	0.0091
Sales class 5	0.1188***	0.0094	0.0053	0.0094	0.0092
Sales class 6	0.1013***	0.0098	0.0052	0.0081	0.0084
Sales class 7	0.0828***	0.0093	0.0050	0.0077	0.0077
Sales class 8	0.0670***	0.0090	0.0049	0.0078	0.0077
Sales class 9	0.0468***	0.0084	0.0046	0.0053	0.0050
Mean ratio of SE to SE Clustered by Stratum		1.00	0.66	1.02	1.00
Median ratio of SE to SE Clustered by Stratum		1.00	0.61	0.94	0.98
Number of clusters		47	–	23	69
Controls for state	yes				
Controls for year	yes				
Observations	185,880				

*** p value < 0.01, ** p-value < 0.05, * p-value < 0.10. Standard errors are calculated by applying the delta method to the estimate of the variance-covariance matrix. The marginal effects are the effect of changing x by one unit on the response propensity, calculated at the sample mean of x. For discrete variables the marginal effect is for a change from zero to one.

they responded, and had their response time been recorded, it would have increased the estimate of the response burden for the largest farms and therefore the share of the difference in response propensities accounted for by response burden. Still, the modest role of time in explaining response propensities supports the earlier assertion that time is only one of many potential reasons for nonresponse.

Nonresponse Bias and Econometric Estimates

As labor economists have shown, coefficients estimated using a self-selected sample may poorly approximate the analogous population coefficients. We look at nonresponse bias in two econometric models, one a model of labor market participation of the principal farm operator, and the other a model of the farm diversification discount based on a published article that used ARMS data.

Testing for nonresponse bias

We take an approach to testing for nonresponse bias in econometric analysis similar to that of Casari, Ham, and Kagel (2007). In their study of self-selection bias in common value auctions where bankrupt bidders do not return to participate in the auction, the authors test for equality of coefficients from a subsample with attrition (and therefore potentially biased), and with coefficients from a subsample without attrition (no bias). In our case, the subsample with potential bias is the respondent subsample. To create a comparable subsample without bias, we randomly draw from the full sample of respondents and nonrespondents. We draw so as to ensure that the new subsample has the same number of observations as the respondent subsample, as well as the same proportion of respondents and nonrespondents as the full sample.

We also take the approach in Casari, Ham, and Kagel (2007) a step further by creating bootstrapped confidence intervals and coefficient averages from subsamples randomly drawn from the full sample. Specifically, we draw 1,000 subsamples from the full respondent and nonrespondent sample, estimating the model each time. Using the coefficient estimates from each iteration and the average coefficient over all iterations, we calculate a standard error and 95% confidence interval centered around the average coefficient estimate.[7] We can then see if specific coefficients estimated from the respondent subsample fall within the unbiased confidence intervals. We perform the bootstrapping such that each randomly drawn subsample is identical in size to the respondent subsample and has the same proportion of respondents and nonrespondents as the full sample.

Using Census of Agriculture data to estimate two econometric models

Normally data on the full sample (respondents and nonrespondents) are not available and would preclude testing for nonresponse bias as described above. The Census of Agriculture, however, provides data for ARMS respondents and nonrespondents, which we use to estimate two econometric models. Because we only use data from the Census of Agriculture, we are restricted to variables that it provides.

We ignore the ARMS data available for ARMS respondents. Using Census of Agriculture data to impute ARMS responses for nonrespondents would be problematic since errors in imputation and differences in the timing of data collection would make it difficult to separate measurement error from nonresponse bias. By relying solely on Census of Agriculture data we avoid correlation in measurement error and response status, since data for both respondents and nonrespondents come from the same source collected at the same time.

In a world with 100% response rates, the coefficients estimated using the full ARMS sample would likely differ from the coefficients using the same sample but with Census of Agriculture data. For a variable like farm assets, for example, ARMS has a detailed enumeration of asset types, which allows for more complete accounting, whereas the Census of Agriculture has fewer, broader asset categories. Our nonresponse analysis, however, is similar to that of NASS, which focuses on qualitative differences between respondents and nonrespondents revealed by Census of Agriculture data, not on using Census responses as a substitute for ARMS responses. The work by NASS implicitly assumes that differences between ARMS respondents and nonrespondents implied by Census of Agriculture data provide insights into the magnitude and direction of bias in mean estimates based on ARMS respondent data. Similarly, we find it reasonable to believe that the bias of regression coefficients estimated using Census of Agriculture data is an informative proxy for the bias in models estimated using ARMS respondent data.

Model specification and estimation

The first model estimated is a Probit model of the principal farm operator's labor market participation. Although several variables are often included in such models, there is no standard specification in the literature, and the diversity of variables included in past research reflects the variables available to the researcher and the focus of the research. To guide our model's specification, we draw from four articles on farm operator participation in labor markets: Sumner (1982); Huffman and Lang (1989); Kimhi (1994); Ahearn, El-Osta, and Dewbre (2006).

We include a linear and a quadratic term for the operator's age (Sumner 1982; Huffman and Lang 1989; Kimhi 1994; Ahearn, El-Osta, and Dewbre 2006), operator experience on the farm (Sumner 1982), household size (Sumner 1982; Huffman and Lang 1989; Kimhi 1994;

[7] The percentile-t method offers an asymptotic refinement over the common approach of calculating a bootstrapped standard error based on deviations around the coefficient estimate averaged over all iterations, which we use (Cameron and Trivedi 2005). However, the method treats the original sample as the population and uses coefficient and standard error estimates from it to calculate confidence intervals. In the present application, the original sample includes all respondents and nonrespondents. Due to more observations, standard errors from the original sample would give too narrow a confidence interval with which to assess the potential bias of coefficient estimates from the respondent subsample.

Ahearn, El-Osta, and Dewbre 2006), commodity specialization (Sumner 1982; Kimhi 1994; Ahearn, El-Osta, and Dewbre 2006), and region (Sumner 1982; Huffman and Lang 1989; Ahearn, El-Osta, and Dewbre 2006). All of the variables have theoretical justification, which is often provided in the cited articles, though they are clearly not an exhaustive list of variables that could be included, nor do they match the exact covariate set used in any one article. On that matter, none of the four papers shares the same specification. Our goal, however, is not to match the specification of any particular article, but to have a reasonably specified model for an oft-studied outcome that can be estimated using Census of Agriculture data.

The second model estimated is from Katchova (2005), who tests if markets discount diversified farms similar to what has been found for corporate firms; she estimates an excess value model where a farm's excess value compares its actual value to its imputed value if it were specialized (this is taken to be the median farm value for single-enterprise farms). The excess value is calculated as the natural logarithm of the ratio of the farm's actual value to its imputed value. We calculate excess value in the same manner detailed in the published article. The key explanatory variable is an indicator for whether the farm is diversified across crops and livestock, or specialized in one of the two. For leveraged farms, Katchova finds that diversified farms have a discount of 8.4% (table 4, column 2 in the published article).

Except for the operator's education and the household's total off-farm income, neither of which are in the Census of Agriculture, we match the model specification in the published article. Although the Census of Agriculture does not ask for farm debt, it collects information on interest payments, which we use to infer total debt by assuming a common interest rate. This allows us to calculate the debt to asset ratio and examine only leveraged farms, which Katchova defines as having a debt to asset ratio above 0.01.

We estimate the labor market participation model using the same data and years used in the descriptive analysis. To better match the diversification model, which was estimated using only three years of ARMS (1999–2001), we estimate it using the three most recent ARMS, those from 2008–2010. Both models are estimated using only Census of Agriculture data. To be clear, the "full sample" consists of all respondent and nonrespondent observations matched with Census of Agriculture records and, for the diversification discount model, meets the sample criteria specified in Katchova (2005). We calculate robust standard errors clustered by survey stratum.

Respondent and randomly-drawn subsamples: do they give similar results?

The coefficient estimates from the labor participation model using the respondent subsample are as expected (table 5). Age increases the propensity to work off-farm, but at a decreasing rate. The propensity to work decreases with the size of the farm, as evidenced by the coefficients on the sales class binary variables, and there is substantial variation in labor market participation based on the farm's commodity specialization, with operators of dairy farms having the lowest propensity to work off-farm.

The estimates for a randomly-drawn subsample are quite close to those from the respondent subsample (column 2 of table 5). The test for equality of coefficients across the models rejects the hypothesis of equal coefficients at the 1% level, but given the number of variables in the model and the precision that comes with such a large number of observations, the test is arguably a low bar for detecting nonresponse bias. As previously discussed, an alternative method is to draw repeatedly from the full sample and create confidence intervals for each coefficient. In so doing we find that none of the coefficients estimated from the respondent subsample fall strictly outside of the unbiased 95% confidence intervals.

Turning to the farm diversification discount model, there is no reason to expect the same coefficients as Katchova (2005), given the differences in specification (we exclude operator education and off-farm income because they are not available in the Census of Agriculture), differences in how ARMS and Census collect asset values, and the different study periods (1999–2001 by Katchova, and 2008–2010 in this article). Nonetheless, the signs of the coefficients on all variables are the same as the published results from Katchova (2005), and the coefficients themselves are quite similar for several variables. The coefficient on *Operator age (/100)*, for example, is 0.9 compared to 1.4 in the original article (table 4, column 2 of the published article). The coefficient on the variable of interest, the binary variable indicating a diversified crop/livestock farm (*Diversified*), shows a smaller discount than the published estimate (−0.033 compared to −0.084) and is not statistically distinguishable from zero.

Table 5. Farm Operator Participation in the Labor Market

			Bootstrapped Coefficient Estimates and 95% Confidence Interval			
	Respondent Subsample	Randomly Drawn Subsample	Average Value	Lower Bound	Upper Bound	Semi-Parametric Sample Selection
Age	0.045***	0.044***	0.043	0.038	0.047	0.039***
	(0.008)	(0.009)				(0.007)
Age squared (/100)	−0.060***	−0.059***	−0.056	−0.060	−0.051	−0.058***
	(0.011)	(0.012)				(0.010)
Persons in household	0.028***	0.031***	0.025	0.019	0.031	0.017***
	(0.005)	(0.006)				(0.005)
Experience	−0.010***	−0.010***	−0.011	−0.012	−0.010	−0.009***
	(0.001)	(0.001)				(0.001)
No sales	1.175***	1.212***	1.154	1.107	1.202	0.859***
	(0.053)	(0.051)				(0.042)
Sales 1-49,999	1.111***	1.112***	1.103	1.080	1.125	0.768***
	(0.063)	(0.061)				(0.052)
Sales 50,000-250,000	0.553***	0.547***	0.550	0.530	0.570	0.349***
	(0.035)	(0.033)				(0.024)
Crops	0.134***	0.159***	0.136	0.106	0.166	0.198***
	(0.023)	(0.028)				(0.023)
Other livestock	0.147***	0.184***	0.147	0.115	0.180	0.158***
	(0.025)	(0.030)				(0.024)
Dairy	−0.416***	−0.394***	−0.394	−0.434	−0.355	−0.501***
	(0.071)	(0.069)				(0.078)
Uses production contracts	−0.035	0.031	−0.025	−0.066	0.016	−0.143**
	(0.048)	(0.046)				(0.067)
Constant	−1.331***	−1.354***				−1.640***
	(0.180)	(0.215)				(0.224)
F statistic (p value) for test of equality of coefficients	8.5 (0.000)					
Region and year dummies	yes	yes	yes			yes
N	126,364	126,364	126,364			126,364

Note: *** p value < 0.01, ** p-value < 0.05, * p-value < 0.10. Robust standard errors clustered by ARMS survey stratum are in parentheses. The test for equality of coefficients is implemented by pooling the two samples and interacting the variables with an indicator for being drawn from the respondent subsample. Only the coefficients presented in the table were included in the test for equality of coefficients. The 95% confidence intervals are based on standard errors calculated from 1,000 bootstrap iterations and are centered on the average coefficient estimate over all iterations. Bootstrapping is done so that each randomly drawn subsample is identical in size to the respondent subsample, and has the same proportion of respondents and nonrespondents as the full sample.

Comparing these results with those from a randomly drawn subsample again leads to rejecting the hypothesis of equality of coefficients across subsamples, though only at the 5% level. Again, the coefficients are quite similar, with the only exception being the coefficient on *Diversified*, which is one-third larger in the randomly drawn subsample than in the respondent subsample. Similar to the labor participation model, none of the coefficients from the respondent subsample fall outside the bounds of the 95% interval created by drawing from the full sample (table 6). This holds true for the respondent subsample estimate of the diversification discount, which clearly falls within the interval of −0.070 to −0.028.

Altogether, the findings suggest that nonresponse bias in coefficient estimates is small in magnitude and low in frequency.

Controlling for Nonresponse using a Sample Selection Approach?

Despite the evidence of only modest nonresponse bias in the two models studied, an important question for applied researchers is whether using a sample selection model further reduces any bias in a respondent subsample. Because of concerns about distributional assumptions, many studies have proposed

Table 6. The Farm Diversification Discount (Katchova)

	Respondent Subsample	Randomly Drawn Subsample	Bootstrapped Coefficient Estimates and 95% Confidence Interval: Average Value	Lower Bound	Upper Bound	Semi-Parametric Sample Selection
Diversified	−0.033	−0.044*	−0.049	−0.070	−0.028	−0.038*
	(0.021)	(0.024)				(0.020)
Farm assets (/10ˆ6)	0.039*	0.036*	0.033	0.026	0.041	0.099**
	(0.020)	(0.019)				(0.043)
Government payments (/10ˆ6)	−3.744***	−3.989***	−3.568	−3.990	−3.145	−4.211***
	(0.782)	(0.837)				(0.847)
Age (/100)	0.953***	1.039***	0.936	0.831	1.042	0.607***
	(0.101)	(0.097)				(0.133)
Persons in household	−0.025***	−0.021***	−0.027	−0.035	−0.019	−0.035***
	(0.004)	(0.003)				(0.006)
Constant	−0.446***	−0.494***				−0.916**
	(0.085)	(0.072)				(0.419)
F statistic (p value) for test of equality of coef.	2.4 (0.038)					
N	17,915	17,916	17,916			17,915

Note: *** p value < 0.01, ** p-value < 0.05, * p-value < 0.10. Robust standard errors clustered by ARMS survey stratum are in parentheses. The test for equality of coefficients is implemented by pooling the two samples and interacting the variables with an indicator for being drawn from the respondent subsample. The 95% confidence intervals are based on standard errors calculated from 1,000 bootstrap iterations and are centered on the average coefficient estimate over all iterations. Bootstrapping is done so that each randomly drawn subsample is identical in size to the respondent subsample and has the same proportion of respondents and nonrespondents as the full sample.

alternative estimators to the parametric selection model of Heckman (1979), where the errors in the selection and outcome equations are assumed to be jointly normally distributed (for a survey, see Vella 1998). In practice, common applications consist of estimating a response propensity (through either a Probit or a Linear Probability Model) and including some function of the response propensity in the outcome equation (Vella 1998; Duflo and Saez 2002; Hussinger 2008).

We draw on the treatment of two-step sample selection models in Newey (2009) for guidance in specifying a flexible sample selection correction. Through semiparametric estimation we relax the assumption of normality (assumed in a first stage Probit model) in estimating the selection index $Z_i'\gamma$ (from the selection equation $Respond_i = I[Z_i'\gamma + \mu_i > 0]$). Gallant and Nychka (1987) show that a range of density functions with arbitrary skewness and kurtosis can be approximated by a Hermite polynomial expansion. This approximation of the density function can then be used in place of the normal density function (as in the Probit model) in the likelihood function and the parameter vector γ estimated by maximum likelihood (De Luca 2008).[8] We then enter the index $\hat{Z}_i'\gamma$ in the outcome equation as a spline function, which is less sensitive to outliers than a series of polynomials (Newey 2009). For estimation we specify a linear spline with ten knots corresponding to deciles of $\hat{Z}_i'\gamma$.

Relaxing the distributional assumptions of the Heckman model reduces the possibility that misspecification undermines the selection correction. Even so, selection models in general work best with valid exclusion restrictions – variables that affect selection but not the outcome (Puhani, 2000). Valid exclusion restrictions are hard to find in practice, since it is rare for variables that influence participation in economic activities to have no influence on the level of activity. In the present application, a potentially valid exclusion restriction exists. Conditional on meeting basic criteria – for example, having the commodity targeted in the commodity-specific Cost and Returns Report – NASS randomly selects the version of the survey that an operator receives. Conditional on the criteria, survey assignment should

[8] This semiparametric estimation of a binary outcome model is implemented through the snp command in Stata.

be unrelated to outcomes like labor market participation or farm values. And as shown by the Probit model results, due to variation in response burdens associated with the survey versions, the survey assignment affects whether an operator responds to the ARMS. We therefore use variables indicating survey assignment as our exclusion restrictions. In both models the selection equation includes all the variables in the outcome equation and the indicator variables for whether the operator received the Core or the Version I Cost and Returns Report.

In general, applying a sample selection model does not improve upon the base model that ignores selection (see the last column in tables 5 and 6). For the labor participation model, several coefficients move considerably further away from the ones estimated in the randomly drawn subsample. The same holds for all coefficients of the diversification discount model except for that of *Diversified*, which moved closer to the estimate from the randomly drawn subsample.

Using a spline function with only five knots or a series of polynomials (to the fifth degree) and re-estimating the diversification discount model gives qualitatively similar results (results in the supplementary appendix). Because $\hat{Z}_i'\gamma$ enters non-linearly into the second stage, the model is identified even without exclusion restrictions (though the linear term in the polynomial series must be dropped). Estimating the selection model without the exclusion restrictions yields estimates that are no worse than those incorporating the restrictions (columns three and four of the appendix table). The finding suggests that future attempts to account for nonresponse should potentially look to exclusion restrictions other than survey assignment.

Conclusion

Pooling eight of the last ten years of ARMS samples and matching them with Census of Agriculture records reveals that ARMS nonrespondents operate larger farms than respondents. Most striking is that even after controlling for region, commodity specialization, and other farm and household characteristics, the propensity to respond consistently decreases when going from the first to the last decile in farm size distribution.

Despite the clear differences between respondents and nonrespondents, our findings suggest that nonresponse bias is unlikely to undermine conclusions based on econometrics using ARMS. In no case does a coefficient estimated from the respondent subsample fall outside of a 95% confidence interval generated by bootstrapping over the full sample of respondents and nonrespondents. Application of a flexible sample selection model in general does not improve upon the estimates that ignore selection.

The results, however, are conditional on the two models estimated. The U.S. farm population and uses of ARMS data are diverse. Nonresponse bias can vary from model to model and subsample to subsample, depending on correlations between unobserved variables, respondent status, and the outcomes under study. Researchers working with specific types of farms (e.g. those with production contracts) or topics (e.g. government payments) could use supplementary data uniquely available for their target subpopulations for a more focused study of nonresponse – for example, using administrative data on ARMS respondents and nonrespondents who participate in commodity programs.

References

Abraham, K. G., Maitland, A., and S. M. Bianchi. 2006. Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter? *The Public Opinion Quarterly* 70(5): 676–703.

Ahearn, M.C., El-Osta, H., and J. Dewbre. 2006. The Impact of Coupled and Decoupled Government Subsidies on Off-Farm Labor Participation of U.S. Farm Operators. *American Journal of Agricultural Economics* 88(2):393–408.

Ahearn, M., Banker, D., Clay, D., and D. Milkove. 2011. Comparative Survey Imputation Methods for Farm Household Income. *American Journal of Agricultural Economics* 93(2):613–618.

Beckler, D.G., Ott, K., and P. Horvath. 2005. Indirect Monetary Incentives for the 2004 ARMS Phase III Core. Washington DC: U.S. Department of Agriculture, National Agricultural Statistics Service.

Cameron, A.C. and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge, UK: Cambridge University Press.

Casari, M., Ham, J.C., and J.H. Kagel. 2007. Selection Bias, Demographic Effects, and Ability Effects in Common Value Auction Experiments. *The American Economic Review* 97(4):1278–304.

Cecere, W. 2009. 2007 Census of Agriculture Non-Response Methodology. Proceedings of the Survey Research Methods Section, American Statistical Association. Available at: http://www.amstat.org/sections/srms/proceedings/y2009/Files/304163.pdf. Accesed on: 18 January 2013.

Complex Agricultural Establishments. 2011. Paper prepared by the Planning Committee for an International Workshop on Enhancing Data for Complex Agricultural Establishments. Niagara-on-the-lake, Ontario, Canada, June 26–28, 2011. Available at: http://www.farmfoundation.org/news/articlefiles/1749-Concept%20paper.pdf. Accessed on: 18 January 2013.

De Luca, G. 2008. SNP and SML Estimation of Univariate and Bivariate Binary-Choice Models. *The Stata Journal* 8(2):1980–220.

Duflo, E. and E. Saez. 2002. Participation and Investment Decisions in a Retirement Plan:The Influence of Colleagues' Choices. *Journal of Public Economics* 85(1): 121–48.

Earp, M.S., McCarthy, J.S., Porter, E., and P.S. Kott. 2010. Assessing the Effect of Calibration on Nonresponse Bias in the 2008 ARMS Phase III Sample Using Census 2007 Data. National Agricultural Statistics Service. Available at: http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/conferences/JSM-2010/earp-2010_jsm_paper_arms_calibration.pdf. Accesed on: 18 January 2013.

Earp, M.S., McCarthy, J.S., Schauer, N.D., and P. S. Kott. 2008a. Assessing the Effect of Calibration on Nonresponse Bias in the 2005 ARMS Phase III Sample Using 2002 Census of Agriculture Data. Washington, DC: U.S. Department of Agriculture, National Agricultural Statistics Service.

———. 2008b. Assessing the Effect of Calibration on Nonresponse Bias in the 2006 ARMS Phase III Sample Using 2002 Census of Agriculture Data. Washington, DC: U.S. Department of Agriculture, National Agricultural Statistics Service.

Gallant, A. R., and D. W. Nychka. 1987. Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica* 55(2):363–390.

Gerling, M.W., Tran, H.N., and M.S. Earp. 2008. Categorizing Nonresponse in Phase III of the 2006 Agricultural Resource Management Survey in Washington State. NASS Research Report Number RDD-08-08. Washington, DC: U.S. Department of Agriculture, National Agricultural Statistics Service.

Green, K. 2011. Complex Farming Insight. Presentation at the conference "Enhancing Data for Complex Agricultural Establishments", June 26-28, Ontario, Canada. Available at: http://www.farmfoundation.org/news/articlefiles/1749-Kevin%20Green%20PPT.pdf. Accessed on: 18 January 2013.

Groves, R. 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70(5):646–675.

Heckman, J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47(1):153–161.

Hedeker, D. and R.D. Gibbons. 2006. *Longitudinal Data Analysis*. Hoboken, New Jersey: John Wiley & Sons.

Hoppe, R.A. and Banker, D.E. 2010. Structure and Finances of U.S. Farms, Family Farm Report. Economic Information Bulletin Number 66. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Huffman, W.E. and M.D. Lange. 1989. Off-Farm Work Decisions of Husbands and Wives: Joint Decision Making. *The Review of Economics and Statistics* 71(3): 471–480.

Hussinger, K. 2008. R&D and Subsidies at the Firm Level: An Application of Parametric and Semiparametric Two-Step Selection Models. *Journal of Applied Econometrics* 23(6):729–747.

Johnson, J., Baum, K., and R. Prescott. 1985. Errors and Limitations in Economic Indicators and Agricultural Policy Analysis. Available at: http://www.amstat.org/sections/srms/proceedings/papers/1985_011.pdf. Accessed on: 18 January 2013.

Katchova, A. 2005. The Farm Diversification Discount. *American Journal of Agricultural Economics* 87(4):984–994.

Kennickell, A.B. and McManus, D.A. 1993. Sampling for Household Financial Characteristics Using Frame Information on Past Income. Paper presented at the 1993 Joint Statistical Meetings, Atlanta, GA. Available at: http://www.federalre serve.gov/Pubs/oss/oss2/papers/asa93.pdf. Accessed on: 18 January 2013.

Kimhi, A. 1994. Quasi Maximum Likelihood Estimation of Multivariate Probit Models: Farm Couples' Labor Participation. *American Journal of Agricultural Economics* 76(4):828–835.

Korinek, A., Mistiaen, J., and M. Ravallion. 2006. Survey Nonresponse and the Distribution of Income. *Journal of Economic Inequality* 4(1):33–55.

Lillard, L., Smith, J.P., and F. Welch. 1986. What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation. *The Journal of Political Economy* 94(3):489–506.

Lin, I.F. and N. C. Schaeffer. 1995. Using Survey Participants to Estimate the Impact of Nonparticipation. *The Public Opinion Quarterly* 59(2):236–258.

National Research Council. 2008. Understanding American Agriculture: Challenges for the Agricultural Resource Management Survey. Panel to Review USDA's Agricultural Resource Management Survey. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Newey, W.K. 2009. Two-Step Series Estimation of Sample Selection Models. *Econometrics Journal* 12: S217–S229.

O'Connor, T.P. 1992. Identifying and Classifying Reasons for Nonresponse on the 1991 Farm Costs and Returns Survey. Research Report No. SRB-92-10. Washington, DC: National Agricultural Statistics Service, U.S. Department of Agriculture.

O'Donoghue, E.J., Hoppe, R.A., Banker, D.E., and P. Korb. 2009. Exploring Alternative Farm Definitions: Implications for Agricultural Statistics and Program Eligibility. Economic Information Bulletin Number 49. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Puhani, P.A. 2000. The Heckman Correction for Sample Selection and its Critique. *Journal of Economic Surveys* 14(1):53–68.

Sumner, D.A. 1982. The Off-Farm Labor Supply of Farmers. *American Journal of Agricultural Economics* 64(3):499–509.

U.S. Office of Management and Budget. 2006. Standards and guidelines for statistical surveys – September. Available at: http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf. Accessed on: 18 January 2013.

Vella, F. 1998. Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources* 33(1):127–169.

Zabel, J. E. 1998. An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *The Journal of Human Resources* 33(2):479–506.