Multivariate Farm Debt Imputation in the Agricultural Resource Management Survey (ARMS)

Mitch Morehart*, Economic Research Service, USDA. (Morehart@ers.usda.gov)
Dan Milkove*, Economic Research Service, USDA. (dmilkove@ers.usda.gov)
Yang Xu, George Mason University


*The views expressed here are those of the author(s), and may not be attributed to the Economic Research Service or the U.S. Department of Agriculture.
Introduction

USDA, through its Agricultural Resource Management Survey (ARMS) collects detailed information from farm operators on specific loan characteristics such as interest rate, loan term, origination date, type of loan, loan purpose, and type of financing. This information is used to construct portions of the farm sector balance sheet in addition to supporting research on credit use, farm solvency, and debt repayment capacity (Kuethe and Morehart 2012 and Harris et. al, 2009). Information collection for sensitive items, such as debt, is subject to item non-response. Item non-response occurs when not all questions are answered. It represents a special challenge to economic surveys (see for example Barcelo, 2008; Drechsler, 2011; Heeringa, et.al, 200; Kennickell, 1998; and Schenker, et.al, 2006). The reason for a “do not know” response is not necessarily the unwillingness to answer. In many cases, it will be lack of knowledge or apprehension.

Ignoring item non-response completely, by setting all missing values to zero, or by taking into account only the existing answers; will result in a bias. Under certain conditions, a bias due to item non-response can be mitigated or even avoided by imputation, depending on how well item non-response can be explained by observed variables (Rubin, 1996 and Schafer, 2010). Imputation is the practice of filling in missing data with plausible values. Historically, the ARMS has used a generalized cell mean imputation approach for general categories of debt and made no systematic effort to impute the detailed components asked in the debt reporting table.

The shortcomings of the current procedures are twofold. First it does not provide for full imputation of all potential debt responses when detailed questions are asked. Secondly, it suffers from the reported drawbacks of univariate imputation approaches as discussed by Rubin and others (Rubin 2009 and Reiter and Raghunathan, 2007). Moreover, recent external review of the ARMS program has highlighted the need to explore alternative imputation methods and has identified unexplained differences between ARMS estimates of debt by lender and administrative data (Briggeman, et.al, 2012 and National Research Council, 2007). Research on ARMS survey methods has also highlighted the need to consider alternative imputation methods (Robbins et.al, 2013 and Ahearn et. al, 2011).
## SECTION K  FARM DEBT

1. Was debt used in funding the operation of this farm/ranch in 2013, including any loans obtained in earlier years?  
   (Include seasonal production and other loans taken and repaid during 2013.)
   
   [ ] Yes – Continue  [ ] No – Go to Section L

2. What was the total amount of all farm business loans taken out and fully repaid in 2013?  
   (Include seasonal production and other loans.)

   Dollars

3. To estimate the financial position of farms correctly and their ability to service debt and to categorize debt by types, we need to list loans this operation had on December 31, 2013, including any line of credit. (Include farm/ranch loans, debt on the operator’s house if owned by the operation, and multi-purpose loans used for both farm and non-farm purposes. Exclude CCC commodity loans and any loans used exclusively for non-farm purposes.)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who is the lender?</td>
<td>What was the balance owed on Dec 31, 2013 including outstanding principal plus unpaid interest?</td>
<td>What was the interest rate on Dec 31, 2013? (Report in hundredths of a percent. Example: 8% = 0.80)</td>
<td>What is the type of loan? (From Loan Type Codes Below)</td>
<td>What year was it obtained?</td>
<td>What is the original term of the loan? (Number of Years)</td>
<td>What percent is for operating expenses, capital expenditures, or other expenses of the farm operation? (From Loan Purpose Codes Below)</td>
<td>Is this loan a:</td>
<td></td>
</tr>
<tr>
<td>(Code)</td>
<td>(Dollars)</td>
<td>(Percent)</td>
<td>(Code)</td>
<td>(Year)</td>
<td>(Percent)</td>
<td>(Code)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>1002</td>
<td>1003</td>
<td>1004</td>
<td>1005</td>
<td>1006</td>
<td>1007</td>
<td>1008</td>
<td>1009</td>
</tr>
<tr>
<td>1010</td>
<td>1011</td>
<td>1012</td>
<td>1013</td>
<td>1014</td>
<td>1015</td>
<td>1016</td>
<td>1017</td>
<td>1018</td>
</tr>
<tr>
<td>1019</td>
<td>1020</td>
<td>1021</td>
<td>1022</td>
<td>1023</td>
<td>1024</td>
<td>1025</td>
<td>1026</td>
<td>1027</td>
</tr>
<tr>
<td>1028</td>
<td>1029</td>
<td>1030</td>
<td>1031</td>
<td>1032</td>
<td>1033</td>
<td>1034</td>
<td>1035</td>
<td>1036</td>
</tr>
<tr>
<td>1037</td>
<td>1038</td>
<td>1039</td>
<td>1040</td>
<td>1041</td>
<td>1042</td>
<td>1043</td>
<td>1044</td>
<td>1045</td>
</tr>
</tbody>
</table>

If more space is needed, please use a separate sheet of paper.

### Lender Codes (Column 1)
- Farm credit system
- USDA Farm Service Agency (FSA)
- Small Business Administration (SBA)
- State & county government lending agencies
- Savings and loan associations, residential mortgage lenders
- Commercial banks
- Life insurance companies
- Implement dealers and financing corporations
- Credit unions
- Other lenders
- Co-ops and other merchants
- Contractors
- Individuals from whom any land in this operation was bought under a mortgage or deed of trust
- Individuals from whom any land in this operation was bought under a land purchase contract

### Lender Codes (Column 1) (continued)
- Any other individuals
- Credit cards
- Farmer Mac
- Credit union
- Any other banks
- Other loans
- Any other individuals
- Credit cards
- Farmer Mac
- Credit union
- Any other banks
- Other loans

### Loan Purpose Codes (Column 2)
- Purchased real estate (land & its attachments)
- Building construction
- Construction of livestock and poultry facilities
- Development and rehabilitation
- Purchase feeder livestock
- Purchase other livestock
- Other current operating expenses
- Core crops production
- Core and feeding livestock including poultry
- Labor, feed, seed, fertilizer, graze, rent, repair and maintenance
- Farm machinery and equipment
- Debt consolidation
- Other

### Loan Purpose Codes (Column 3)
- One year or less production or other loans
- Non-real estate loan more than one year
- Real estate loan more than one year for operator’s dwelling
- Other real estate loans more than one year

4. If you had farm loans in addition to the five reported in Item 3, what is the total amount of debt from these loans owed on December 31, 2013? (Include farm/ranch loans and debt on the operator’s house if it is owned by the operation. Include any loans exclusively for non-farm purposes that are secured by assets of the farm/ranch.)

   Dollars

5. How much of the total debt owed on December 31, 2013 (reported in items 3 and 4 above), was for the operator’s dwelling? (If the operator’s dwelling is owned by the operation debt should be included here and above. Exclude operator’s dwelling if not owned by the operation.)

   Dollars

Office Use Only

Dollars

0999
Imputation Process

There were a number of variables that were significantly associated with the incidence of debt and whether or not this information was missing. These associations formed the basis of multiple imputation models, which include covariates of interest to this analysis, as well as other variables previously identified in the literature as being related to use of credit (see for example: Briggeman, 2009; Harris et.al, 2009; and Katchova, 2005).

Imputation for the ARMS debt table represents a unique challenge since the data contains a mixture of categorical variables and continuous variables with skewed distributions and a variety of often hierarchical skip patterns and logical constraints. As a result, we apply the fully conditional specification approach, iteratively imputing one variable at a time, conditioning on the other variables available in the dataset as a multi-stage process (Templ, et.al, 2011).

Imputation and statistical analysis were performed with SAS 9.2 (SAS Institute, Cary NC) and the SAS-callable implementation of IVEware (Raghunathan, 2002). For multivariate imputation we rely on four different model specifications. Linear models are used for continuous variables, the logit model is used for binary variables and the multinomial logit for variables with more than two categories, and a two stage (logistic then linear) model is used for mixed variables that are both categorical and continuous.

<table>
<thead>
<tr>
<th>Categorical Imputation Model (R1001, R1004, R1007, R1009)</th>
<th>Continuous Imputation Model (R1002, R1003, R1005, R1008, R1006)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRATAY</td>
<td>STRATAY</td>
<td>Collapsed sample strata (1-32)</td>
</tr>
<tr>
<td>STREG</td>
<td>STREG</td>
<td>Core state (1-15) or region based on census divisions (1-5)</td>
</tr>
<tr>
<td>AGECLS</td>
<td>AGECLS</td>
<td>Operator age class (1-5)</td>
</tr>
<tr>
<td>RENTLAND</td>
<td>RENTLAND</td>
<td>IF P44 &gt; 0 THEN RENTLAND=1</td>
</tr>
<tr>
<td>RENTEQ</td>
<td>RENTEQ</td>
<td>IF P750 &gt; 0 THEN RENTEQ=1</td>
</tr>
<tr>
<td>PCONTRACT</td>
<td>PCONTRACT</td>
<td>IF P400=1 THEN PCONTRACT=1</td>
</tr>
<tr>
<td>CAPLAND</td>
<td>CAPLAND</td>
<td>IF P807+P810+P813 &gt; 0 THEN CAPLAND=1</td>
</tr>
<tr>
<td>CAPEQUIP</td>
<td>CAPEQUIP</td>
<td>IF P821+P822+P823+P824 &gt; 0 THEN CAPEQUIP=1</td>
</tr>
<tr>
<td>GOVPAY</td>
<td>GOVPAY</td>
<td>IF IGOVT &gt; 0 THEN GOVPAY=1</td>
</tr>
<tr>
<td>PTAX</td>
<td>PTAX</td>
<td>IF P744 &gt; 0 THEN PTAX=1</td>
</tr>
<tr>
<td>INSUR</td>
<td>INSUR</td>
<td>IF P729 &gt; 0 THEN INSUR=1</td>
</tr>
<tr>
<td>BREED</td>
<td>BREED</td>
<td>IF P621 &gt; 0 THEN BREED=1</td>
</tr>
<tr>
<td>E_CAPLAND</td>
<td>E_CAPLAND</td>
<td>Capital expenses, land improvement (P807+P810+P813)</td>
</tr>
<tr>
<td>E_CAPEQUIP</td>
<td>E_CAPEQUIP</td>
<td>Capital expenses, equipment (P821+P822+P823+P824)</td>
</tr>
<tr>
<td>VEHICLE</td>
<td>VEHICLE</td>
<td>Capital expenses, cars and trucks (P816+P818)</td>
</tr>
<tr>
<td>IGOVT</td>
<td>IGOVT</td>
<td>Total government payments</td>
</tr>
<tr>
<td>P738</td>
<td>P738</td>
<td>Interest expense, debt secured by real estate</td>
</tr>
<tr>
<td>P741</td>
<td>P741</td>
<td>Interest expense, debt NOT secured by real estate</td>
</tr>
<tr>
<td>P744</td>
<td>P744</td>
<td>Property taxes paid on real estate</td>
</tr>
<tr>
<td>P756</td>
<td>P756</td>
<td>Depreciation expense</td>
</tr>
<tr>
<td>P621</td>
<td>P621</td>
<td>Breeding stock purchase expense</td>
</tr>
<tr>
<td>P820</td>
<td>P820</td>
<td>Capital expenses, tractors</td>
</tr>
<tr>
<td>P803</td>
<td>P803</td>
<td>Capital expense, farm land</td>
</tr>
<tr>
<td>P860</td>
<td>P860</td>
<td>Value of operators dwelling</td>
</tr>
<tr>
<td>Z1001-Z1045</td>
<td>Z1001-Z1045</td>
<td>Prior imputed values</td>
</tr>
</tbody>
</table>
Results

Imputations can be validated using a reasonability standard by examining the differences between observed and missing values, and the distribution of the completed data as a whole (Abayomi 2008). The outcomes for categorical variables indicate relatively small changes in the distribution across response codes. For example, for loan types (1-3), the share of farms with production loans of one year or less declined by 1.6 percent and the share with non-real estate loans of more than one year went up by 1.8 percent. The share with real estate loans was virtually the same. For continuous variables, such as debt, the consistency after imputation is illustrated by the similarity in the overall distribution and the scatter plot versus total production expenses.
Implications

When summarized across all ARMS versions, multiple imputation can be compared with the mean imputation approach used in the past. The results suggest a $27 billion (17 percent) increase in total debt. Imputation differences were more pronounced for current liabilities (23 percent increase) and non-real estate, non-current liabilities (25 percent increase). Real estate liabilities increased when using multiple imputation by 12 percent.

As a result of the increase in debt, estimated equity (the difference between total assets and total debt) declined by less than 2 percent. In addition, debt-to-asset ratio estimates increased across a broad range of producers with the potential of more debt repayment challenges emerging. For farm businesses, the largest increase in debt-to-asset ratios occurred for poultry, hog, and dairy farms where debt was highly concentrated prior to imputation.

For the previously un-imputed ARMS version 1 debt table, results show that applying multivariate imputation procedures would increase the estimate of total farm debt by $55 billion; an increase of approximately 40 percent. Increases for major lenders ranged from 24 percent for life insurance companies to almost 70 percent for individuals and others. Coverage relative to lender administrative data improved across all major lenders with the largest gains for Commercial Bank debt, with more than 75 percent of reported loans outstanding captured by ARMS after imputation.
References


