Using Administrative Data in the 2007 Census of Agriculture Edit

Virginia Harris
USDA National Agricultural Statistics Service
virgina_Harris@nass.usda.gov

Dan Beckler
USDA National Agricultural Statistics Service
dan_beckler@nass.usda.gov

Jeff Beranek
USDA National Agricultural Statistics Service
jeff_beranek@nass.usda.gov

Introduction

The Census of Agriculture has collected data on Federal farm program payments made to farm operations since the 1987 Census of Agriculture. Other USDA agencies annually report the amount of farm program payments paid out. The Census of Agriculture collects farm program payments for two reasons: one, to help determine if an operation qualifies as a farm, and two, to show in published tables the amount of payments received by farms, and the distribution of the payments to operations in terms of amounts of payments, as well as the types of farms which receive such payments.

Historically, published totals of payments received by farm operators are less than the reported amount of payments made by the USDA. In the 2002 Census of Agriculture, payments fell even shorter than the typical historical ratio, and efforts were made to change the imputation strategy for farms not reporting payments.

The 2007 Census of Agriculture was the first to use administrative data for editing government payment questions on the census report form. The goal of project was to use administrative data from the Farm Service Agency (FSA) to improve 2007 Census of Agriculture published totals for government program payments received by agricultural operations

Nature of Farm Program Payments

Payments made are subject to limitations on both the amount of payments that can be received, and the characteristics of the recipient. For the Farm Security and Rural Investment Act of 2002 (more commonly called the 2002 Farm Bill), farmers were ineligible for payments if their adjusted gross income (AGI) was more than $2.5 million and if the farming share of that income was less than 75 percent. Payment limits were by category of program. CRP payments were limited to $50,000 per year, direct payments to $40,000, counter-cyclical payments to $65,000, price support payments to $75,000, and NAP (Noninsured Crop Disaster Assistance Program) payments to $100,000. At a common base yield for corn (of 110 bushels per acre), direct payments would be around $25 per acre. This would imply payments could be received only for the first 1600 acres farmed, if all acres were sown to corn. Producers would have incentives to modify farming arrangements if they believed their current operating arrangements would cause them to come up against payment limits.

Payments may be received by a variety of entities, not all of whom have to be engaged in hands on farming. Basically, an entity is entitled to receive payments if they are subject to the risk of profit or loss in a farming operation. For example, a landowner who rents out his or her land on a share rent basis usually receives half of the crop and pays half the expense. Because this involves a risk to the landlord in the amount received because of growing conditions during the year, this would entitle the landlord to a share of the direct payments received.

The Census of Agriculture target population is not identical to the target population of recipients of farm program payments. The eligible group is farm operators, which is defined as the person making day to day decisions for the farm operation. If a person is renting all their land to another person, the tenant would be the target person for the census.

Table 1: Ratio published census totals to USDA payments made (in billions of dollars)

	Census total	FSA total	Ratio
1992	5.054	9.168	0.55
1997	5.294	7.495	0.71
2002	6.546	12.415	0.53
2007	7.984	11.903	0.67

Source: published census tables, ERS farm income data files

The largest category of payments made is for the direct payment program. Direct payments are limited to farms that have (or had) production of program crops, which are limited to wheat, corn, grain sorghum, barley, oats, upland cotton, rice, soybeans, oilseeds, peanuts, and dry peas, lentils, and chickpeas. Both farm owners and operators are eligible, but all recipients must share in the risk of production. This means that farm owners or landlords must be renting out land under a crop share arrangement to receive direct payments. The amount of payments received is tied to the base acres and program yields on the farm. Payments amounts vary by crop. Base acres and program yields are specific to an individual farm, and unknown to the census edit. Base acres are set based on past production history of the land, and are often based on the amount of production in 1987. Another program based on base acres and program yields is the counter-cyclical program. However, payments under this program are made only with the average marketing year price falls below a crop specific target price.

Within the USDA, most federal farm program payments are administered by the Farm Service Agency (FSA). FSA maintains different databases which contain information on payments made. The database used for this project was one that showed essentially the checks for program payments. The identification number associated with the payments was either a Social Security number (SSN) or an employer identification number (EIN). The payment data also included a name and address of the person or entity receiving the payment, as well as administrative data for the program and program year for which the payment was made. The database included negative payments, which implied a recipient had to repay previous payments.

Data on recipients from 280 different programs were received from FSA. These 280 codes had to be mapped to only 4 census item codes. In addition, not all programs should have been reported on census forms, since not all were related to farming activities. An example of such a program would include those related to warehouse storage.

FSA administers a wide variety of agricultural commodity, credit, conservation, disaster, and loan programs, as specified by the Farm Bill. Persons wishing to participate in these programs are required to complete applications that collect information used to determine program eligibility, conformity, and to calculate (qualifying) governmental payments. These application and payment data formed the administrative dataset NASS received to assist with the governmental payments section of the 2007 Census of Agriculture.

NASS and FSA have a long-standing cooperative agreement – documented in a memorandum of understanding – that serves to fulfill each agency's mission. FSA uses many of the official USDA estimates NASS generates to evaluate and administer the aforementioned programs. In turn, FSA provides NASS with access to much of the data it collects to administer these programs. NASS uses these data to assist in list frame maintenance, as an evaluation tool for its estimation programs, as ground-truth data for the creation of the Cropland Data Layer, and for imputation procedures.

A basic difference exists between NASS and FSA farm definitions. An FSA farm is comprised of land that has the same owner(s) and operator(s) with all of the following elements in common: agricultural labor, agricultural equipment, an accounting system, and management. Associated with each FSA farm are one or more persons who are responsible for the agricultural activity on the Farm or own the land. Persons may be associated with one or more FSA farms, and may be designated as owners for some and operators for others.

A NASS farm is any place from which $1,000 or more of agricultural products were, or normally would be, produced and sold during a calendar year. Associated with each NASS farm are one or more persons who are responsible for the day-to-day decisions of the operation. These persons are collectively referred to as operators. Landowners who do not participate in the day-to-day decisions would not be considered operators.

The relationship between FSA Farms and NASS Farms can be very complicated. This is unfortunate since this can prevent FSA data from being directly associated with NASS Farms. Often the FSA to NASS Farm relationship is fairly simple, resulting in one or more entire FSA Farms mapping to a single NASS Farm operation. However, the FSA to NASS Farm mapping becomes much more complicated when a single FSA Farm is parsed between multiple NASS Farms, or vice versa. For these situations, NASS relies on a probabilistic record linkage process that involves multiple name and address identifiers. However, the record linkage process is not error free and can involve substantial staff and time resources.

It is also worth noting that the coverage of FSA data is not uniform across the country or commodities, since the propensity for farm operators to participate in certain FSA programs is a function of the commodities they produce (and the commodities they produces is, in part, a function of the part of the country in which they live). FSA traditionally has less coverage of producers of livestock and certain specialty crops (e.g., fruits, berries, vegetables) since Federal agricultural programs have targeted these commodities less frequently than other commodities.

NASS obtained FSA governmental payment data in late December 2007; these data represented payments issued to persons associated with FSA Farms throughout calendar year 2007. The late December 2007 delivery date was necessary to allow time for the data to be formatted and loaded to a NASS database for Census processing. Unfortunately, the timing also meant that the data excluded a limited amount of payments that were pending or otherwise were not processed by late December 2007.

The actual dataset FSA provided was a comma separated value (CSV) file containing 12.7 million records - far above the roughly 2.2 million farms eventually estimated by the 2007 Census of Agriculture. The large difference in counts is a result of the payment data containing separate records for each specific payment (or partial payment) FSA issued for every agricultural program and for every person associated with the FSA Farms enrolled in the programs. For example, a person might be receiving a check for direct payments for each of several program crops, and for each of several FSA operations with which he or she is involved. For NASS to make complete use of these data it would be necessary to be able to discern which payments could be mapped to which NASS operations. This difficulty of associating FSA payment data to specific NASS farms allowed NASS to include data for only those programs and persons that best conformed to the requirements of the Census of Agriculture. After collapsing this data into the payment groups defined on the census, and performing record linkage with the NASS CML, FSA data were able to be linked to slightly fewer than 650,000 NASS operations. Because of the uncertainty involved, already described, with the FSA-to-NASS farm identification, NASS statisticians decided to use only the presence or absence of payments in editing the census data, rather than the actual level of payments.

The CSV file was initially read with SAS software. Although no formal editing (i.e., cleansing) was done to the data, basic aggregates were generated at the U.S., state, and agricultural program level. These aggregates were reviewed for reasonableness and to ensure the data were read properly; no anomalies were found. Next, SAS software generated a flat file that conformed to Mass's standard extract, transform, and load (ETL) process for its Redbrick relational database. Lastly, the ETL process loaded the FSA

governmental payment data to the Redbrick database where they were available for the Census processing system.

Questions on 2007 Census Form

In December 2007, NASS mailed census forms to just over 3 million operators on the census mail list (CML). Federal farm program data were requested in the following categories:

1) Amount received in 2007 for participation in the Conservation Reserve Program, (CRP), Wetland Reserve Program (WRP), Farmable Wetlands Program (FWP), or the Conservation Reserve Enhancement Program (CREP);
2) Amount received from direct payments as defined under the 2002 Farm Bill;
3) Amount received from counter-cyclical payments, loan deficiency payments (LDPs), marketing loan gains, and net value of commodity certificates. Include payments received through cooperatives;
4) Amount received from other Federal agricultural program payments, including disaster payments and market loss payments, national dairy market loss payments, non-insured assistance program, EQIP (Environmental Quality Incentives Program), CSP (Conservation Security Program)), livestock programs, and any other Federal programs. Exclude any type of insurance payments received.

In 2002 the questions were for the amount received for CRP, etc., and the amount received participation in other Federal programs (in effect the sum of items 2-4 above). In previous years, the total, including CRP was asked, and the CRP payments asked separately, so that the total 'non-CRP' payments could be calculated. In part the decision to ask for more specific payments data was driven by the perception that government payments tended to be underreported on the Census of Agriculture; by listing some of the specific types of payments separately, it was hoped that item nonresponse might decrease, if the reasons for it were at all due to respondents not being clear on what constituted Federal farm program payments.

Use of Data in Edit

The edit sub-system itself consists of Decision Logic Tables (DLTs), which are 'if-then-else' logic expressions custom-written to check the census data. The DLTs are organized by module, where a module roughly corresponds to a section of the questionnaire. In general, missing data, or those judged inconsistent or incompatible with other data, are altered either by deterministic logic (e.g., replacing a reported sum by the sum of the reported parts), by replacement with a (possibly adjusted) value from a previously reported survey, or by imputation from a similar clean record in a donor pool. Questionable relationships in the data are noted using three indicators of increasing severity, and the records are either reedited via the Data Review interactive edit, or marked clean and posted to the database.

Use of FSA data in editing census reports was, initially and in principle, fairly straightforward. For example, an operator who reported no direct payments (either by writing in a 0, or checking 'None', or by leaving the cell blank), but for whom FSA direct payment data were available (or rather, a variable which indicated that FSA payment data had been mapped to this record. - as explained above, NASS decided against using the actual amount of payments), then direct payments were imputed from a 'similar' clean record in the donor pool. Similar logic was used to impute for counter-cyclical payments and LDPs, and 'other' Federal agricultural program payments. However, in practice some difficulties were encountered. For example, it quickly became apparent that many reporters who were unable to separate payments they received as direct payments, from those they received as counter cyclical or loan deficiency payments. A pattern of reporting (say) direct payments, but no LDPs, while having FSA data for LDPs, but none for direct payments, became apparent as more data were edited, and the edit had to be adjusted to accommodate this misreporting, as otherwise LDPs would be imputed for this record (correctly) but the erroneous direct payments would be retained as well. Another problem, which had been foreseen, was that of 'base acres'. Direct payments are determined by an operation's 'base acres' of certain program crops; these are based on amounts of these crops that the operation had been growing at a specific time in the past. Hence, an operation's base acres are not necessarily the same, or even the same crops, as those currently being grown on the operation. Indeed, the farm may be receiving direct payments, yet not currently growing ANY program crops. Ideally, the best or 'closest' donor for a record to impute direct payments

would be one with similar base acres of the same program crops. For example, to impute direct payments on a record with 100 acres of base acres all in corn, a donor with a similar amount of base acres all in corn would be the 'closest' such donor. Records with a different amount of base acres all in corn could also be used; the payments would simply be prorated to the amount of base acres on the recipients. Records with a mix of base acres in different program crops could also probably be accommodated. Unfortunately, an operation's base acres were not available to the census edit; hence, identification of a suitable donor record, and determining an appropriate 'ratio' variable for the imputation, were problematic. A calculation based on current acres was finally used both to identify 'close' donors, and to prorate the payments from the donor eventually selected, but this was obviously less than ideal.

Results of Project

Administrative data appeared to help compensate for underreporting. Some issues which arose during editing and analysis included a form design problem, which lead respondents to tend to report their totals for direct payments and counter-cyclical payments/LDP payments in one cell, leaving the other blank. As mentioned above, the edit was changed to account for this. Another difficulty was how to impute for missing data when many unknowns exist, such as unknown quantity of base acres (direct payments are made based on program crop acres * 85% * yield in 1987). After editing, macro level data analysis suggested that the level of payments imputed was high compared to reported payment amounts for similar types of farms and size of farms. These problems were state specific (certain states were worse than others), but the result was that NASS concluded that payments in categories asked on census form could not be published. These categories were collapsed to improve published totals.

Areas for Further Research

- Results from the 2007 census showed that breaking the total payments question into components increased the frequency of respondents "bracketing" totals for Change items collected on census to reduce details of categories
- Continue attempts to better match FSA farms to NASS farms. Among other benefits, it might then be possible to use actual payment amounts, instead of merely an indicator variable, if only for a limited number of farms.
- Explore possibility of base acres being made available to the census edit.
- Revisit edit and imputation. Among other things, analysis suggests that, for similar type and size of farms, the level of imputed payments was high compared to reported data.