

Using Administrative Data in Producing U.S. Agricultural Estimates

Virginia Harris

National Agricultural Statistics Service

Enhancing Data for Complex Agricultural Establishments

Workshop 6/28/11



2007 CENSUS OF
AGRICULTURE

Administrative Data Sources

- Lists of establishments reporting farm income (from IRS, Tax Agency)
 - Used for enhancing and updating the list frame
- Death lists (from Social Security Administration)
 - Used for possible updates of the list frame



Administrative Data Sources

- Field, farm, and county level aggregate production data from other USDA agencies
 - Used for estimation program
- Person level data from other USDA agencies
 - Used for list frame building
 - Used as indicator data for census editing
- Price data from agencies within USDA
- Slaughter data from other agencies within USDA



Administrative Data Sources

- Even from sources within the USDA, multiple databases, file formats, and data platforms exist to make sharing data difficult
- Perhaps more challenging is the variety of different definitions for an agricultural operation or farm operator



Administrative Data and Complex Organizations

- Administrative data usefulness varies depends on the type of data required by the program
 - Aggregate administrative data is most helpful for aggregate estimates (example: corn acres planted in a certain state and/or county)
 - Most are compiled by summing field level data so are not impacted by the complexity of the organization structure

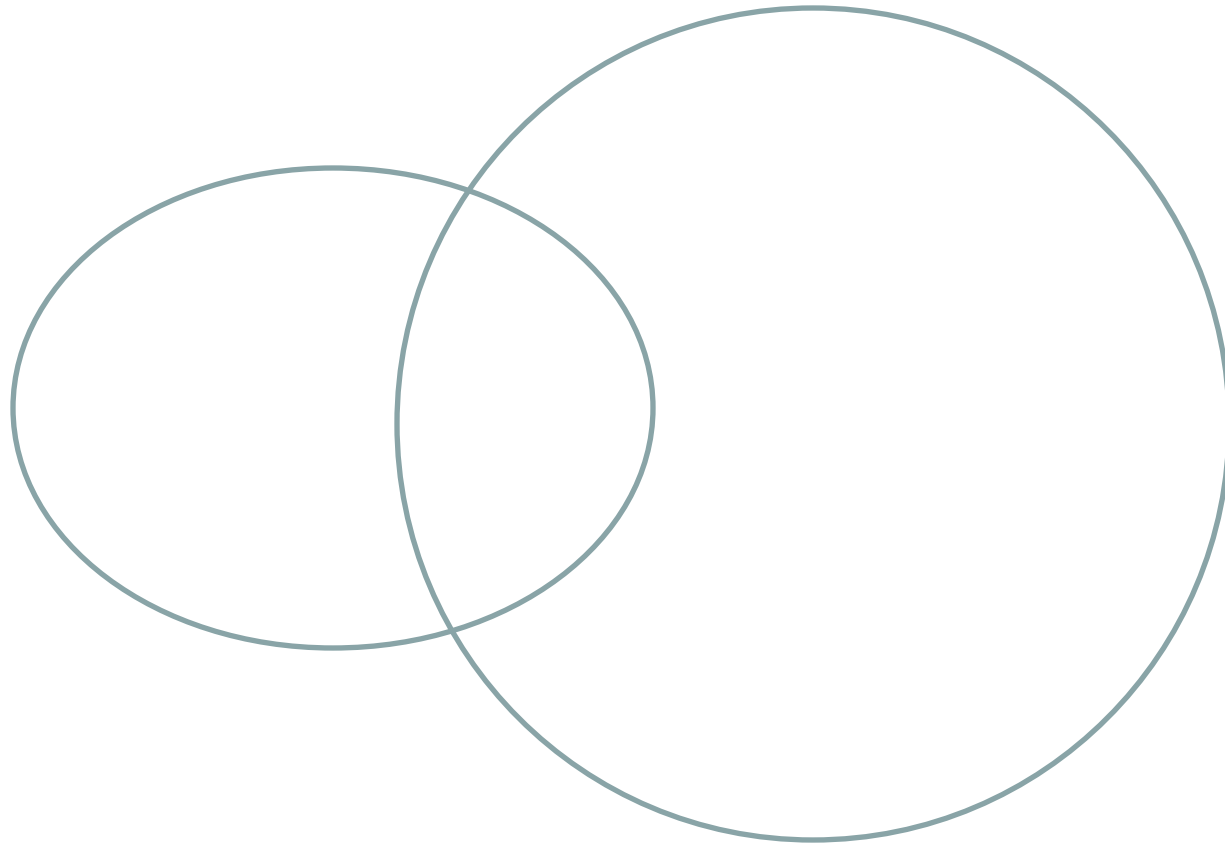


Difficulties in Use of Administrative Data

- Problems arise in use when either the definition or the boundaries (structure/extent) of the entities do not match between the operational survey program and the administrative data source
- For surveys that rely on sample weighting, the boundary of the operation and the boundary of the admin data must match, or data can be either over or under expanded



Illustration



Administrative Data –Record Matching

- Matching records is a difficult and time consuming project, especially in cases where there are millions of records in the survey (or census)



Examples of Uses (Micro)

- Field level data are available from Farm Service Agency on crop production
- Data are used for ground truthing of remote sensing satellite data
- Unit of match is identical- an area of land (one pixel = X acres)
 - Accuracy of match dependent on the accuracy of geographic boundaries obtained from administrative agency



Example of Use (Micro)

- NASS attempted a new use of administrative data in the 2007 Census
- Had received requests to obtain government subsidy payments made to farm operators for use in the census edit
 - Address underreporting by respondents
 - In the future, use instead of respondent reporting



Government payment project

- Obtained 12.7 million data records from the Farm Service Agency, the administrator of subsidy payments in the U.S.
 - Individual payments, representing 1.3 million unique tax ids
- Matched these records to the 3.2 million records on the census mail list
 - Tax ids mapped to 650,000 NASS farms



Government payment project

- Issues
 - Definition of the units represented on the two lists were different
 - FSA records were for persons or tax entities, and were for dollars received by subsidy program
 - Not all recipients of payments will be inscope for the census (landlords, for example)
 - Census records were farm operations, where each farm may have multiple operators, each receiving a separate payment



Accurate matching

- For a census item like government subsidy payments, accurate matching is necessary at the record level
- For data users, item is of interest not in the published totals, but in the distribution of payment dollars to various groups, for example by type of farm, dollar size of farm, or operator characteristics



Administrative Data Timing

- Data from administrative sources is not always as timely as survey sources
 - For example, some agencies collect data for crop subsidy programs at the beginning of the crop season, and then other data not until the season is over (or not until the next crop season)



Future Uses?

- Will future budget pressures necessitate increasing use of admin data?
 - Will nature of estimates published change to match availability of data?
- Data required to analyze distributions requires accurate record level data
 - If units are different on the list frame and admin data source, record level matching is extremely complex



Future Uses?

- Department initiative to standardize data collection, so farmers only need to report once to USDA for program administration
 - Unclear at this point if this includes NASS as a statistical agency, or just the program administration agencies
 - Statistical needs may be different than administrative needs\
- NASS will still need to conduct surveys/censuses to collect the broad range of information needed, but may be able to greater utilize these program data.



Future Issues

- How can data quality measures be adapted to administrative data?
- How can a statistical agency work with an administrative agency to enhance data quality (suitability)?

