AN INVESTIGATION OF INCOMPLETENESS OF LIST FRAMES IN US AGRICULTURAL SURVEYS

Floyd M. Spears[1] and Raj S. Chhikara[1], University of Houston-Clear Lake
Charles R. Perry[2], USDA-NASS
Floyd M. Spears, University of Houston-Clear Lake, 2700 Bay Area Blvd., Houston, TX 77058

KEY WORDS: List Frame Coverage, Ratio Estimator

Abstract

An evaluation is made of the multiple frame USDA/NASS survey estimators for various agricultural items. The list frames for these surveys are incomplete, requiring that additional information be collected using area frames. The level of incompleteness of the list frame can vary substantially at the state and regional level for some of the agricultural items. An investigation is made of the list sample incompleteness and the performance of different estimators as a function of list frame coverage.

1 Introduction

The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) has been investigating estimation approaches that would minimize (or completely eliminate) the use of area-frame samples in adjusting for incompleteness of the list frame. These investigations were initiated because of the relatively high survey cost and respondent burden associated with the area frame samples and also because of the poor precision of the resulting estimates for the area that is non-overlapping with the list frame (NOL).

Several alternative estimators were considered and compared to the currently employed direct expansion (DE) estimator for various labor items as detailed in Perry, et al (1997) and Spears, et al (1996, 1997). Overall, none of the alternative estimators consistently matched the performance of the DE. However, in some cases, alternative estimators performed comparably to the DE. The ratio estimator performed well in some cases and showed the most potential among those estimators that rely less on the area frame than the DE.

The level of coverage of the list frame can vary substantially at the state and regional level for some of the agricultural items. The performance of the ratio estimator is examined as a function of the list frame coverage. Several different estimators currently employed by USDA/NASS for estimating crop, livestock and labor items are included in this study, and ratio estimates are constructed from the estimators. The performance of the estimators is evaluated in Section 3. Section 2 describes the agricultural labor survey data as well as the agricultural items and estimators considered in this study.

2 Background

2.1 Agricultural Surveys

Agricultural surveys are conducted to estimate various crop, livestock and labor items.

Quarterly labor surveys are conducted in July, October, January and April to estimate various characteristics for the three types of farm labor: hired workers, self-employed workers, and unpaid workers. Estimates of the total number of hired, self-employed and unpaid workers are used in our evaluations.

Quarterly crop and stock surveys are conducted to estimate characteristics of various commodities. The survey period varies with commodity. Estimates of the planted acreage, harvested acreage and production for several crop items (alfalfa, corn hay, sorghum, soybean, durum wheat, winter wheat) are used in our evaluations. We also evaluate estimates of corn and soybean stocks.

Quarterly hog surveys are conducted in March, June, September and December to estimate numerous characteristics of hogs. Estimates of the total number of hogs and pigs and the number of hogs marketed in different size categories are included in our evaluations.

[1] These authors' work was supported in part under a cooperative research program of the United States Department of Agriculture at the University of Houston-Clear Lake.

[2] This author is with the National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, VA 22030

2.2 Estimators

USDA/NASS employs a multiple frame estimation methodology that combines separate, independently computed, estimates of the list and NOL components into an estimate of the total for each of the items included in our investigation.

A ratio estimator can be obtained from a multiple frame estimator. Both frames must be estimated during a base survey period. For non-base survey periods, the list frame estimate is multiplied by the the ratio of the multiple frame estimate (list + NOL) to the list frame estimate for the base survey period to obtain the ratio estimate.

Labor Items

The agricultural labor items are estimated using a direct expansion estimator for both the list and area frames. A detailed description of both components of the multiple frame DE estimator is given in Kott (1991). The list frame DE component is based on a sample which is large enough to ensure needed efficiency. The ratio estimator for the labor items is obtained from the DE estimator using July as the base survey period.

Crop and Stock Items

The crop and stock items considered in this study are estimated using a multiple frame estimator (denoted as IMMW) comprised of an imputed estimator (IMP) for the list frame and a modified weight estimator (MW) for the area frame. The base survey period varies by crop and item estimated. For most crops in this study, the base survey period for planted and harvested acreage is March or June. An exception is winter wheat with a base survey period in September. The base survey period is December for corn and soybean stocks, and it is September for wheat stocks and durum wheat stocks. A ratio estimator is constructed from the IMMW estimator using the appropriate base survey period.

Hog Items

The hog items considered in this study are estimated with two different multiple frame estimators. The first estimator (denoted AWMW) combines a revised adjusted list frame estimator with a modified weight estimator for the area frame. A second estimator (denoted RWMW) is comprised of a reweighted estimator of the list frame and a modified weight estimator for the area frame. Ratio estimators are obtained from the ADMW and RWMW estimator using March as the base survey period.

3 Numerical Results

The performance of the estimators described in the previous section is evaluated using estimates from the quarterly agricultural surveys from 1992-1997 for states in each of the 17 agricultural regions in the United States. The state and regional level estimates are compared to the official statistics (OS) and examined as a function of list frame coverage, the ratio of the list frame estimate to the multiple frame estimate for the base survey period. The performance criteria is the relative root mean squared deviation (R-RMSD). The relative root mean squared deviation from the OS, given by,

$$\text{R-RMSD}(\hat{Y}) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - \hat{Y}_{OS,i})^2}}{\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_{OS,i}}, \quad (1)$$

measures the relative variability of an estimator from the corresponding OS relative to the OS using the n quarterly survey estimates. The R-RMSD from the OS was computed at the state and regional levels for each labor item.

In the graphical presentation of the performance results, the R-RMSD values are rescaled relative to the average R-RMSD obtained from the multiple frame estimator in the base survey period.

3.1 Labor Items

In Figure 1, the relative R-RMSD for estimating **number of hired workers** at the regional level is plotted against the list coverage for (1) the DE using all survey periods, (2) the DE using base survey periods, (3) the DE using non-base survey periods and (4) the ratio estimator using non-base survey periods. Note that the relative R-RMSD in each plot is obtained by dividing each R-RMSD by the average of the R-RMSD for the DE using base survey periods.

The list coverage is good, ranging from approximately 0.75 to 0.93 coverage. Even so, there is a noticeable drop in R-RMSD as list coverage increases. The relative R-RMSD is comparable in all four plots indicating that (1) the performance in the non-base period is similar to that in the base period for the DE, and (2) the performance of the ratio estimator is similar to that of the DE.

For the **self-employed** workers at the regional level, the R-RMSD has a significant decreasing trend as a function of list coverage, which ranges from 0.35 to 0.65, with the exception of one region at 0.76. The relative performance of the DE for the base and non-base periods is similar. However, the ratio estimator

has higher relative R-RMSD than the DE, indicating that the ratio estimator is less efficient than the DE.

For the **unpaid** workers at the regional level, the list coverage ranges from 0.38 to 0.84, and the R-RMSD decreases significantly as the list coverage increases. The performance of the DE is similar in the base and non-base periods. The ratio estimator has much higher relative R-RMSD than the DE, particularly in the regions with low levels of list coverage.

The list frame coverage is substantially higher in the case of hired workers than in the cases of self-employed and unpaid workers. Overall, the R-RMSD is substantially higher for the self-employed and unpaid workers than for the hired workers. However, the performance among the three types of workers is similar in the case where the list frame coverage exceeds 0.75.

Estimating at the state level, the range of list coverage was 0.63 to 0.98 for hired workers, 0.31 to 0.85 for self-employed workers and 0.3 to 0.95 for unpaid workers. In all three cases, the relative R-RMSD of the ratio estimator is significantly higher than that of the DE. However, similar to the regional results, the relative R-RMSD decreases significantly as the list coverage increase.

3.2 Crop and Stock Items

The list frame coverage varies across different crops. It ranges from 0.8 to 0.92 for corn and soybean planted acreage, from 0.7 to 1 for durum planted and harvested acreage, from 0.7 to 0.85 for hay harvested acreage, from 0.77 to 0.98 for sorghum planted acreage and from 0.8 to 0.95 for winter wheat planted and harvested acreage. The R-RMSD decreases as list frame coverage increases for corn, soybean and durum, but not for the others. In all cases, the relative R-RMSD for the ratio estimator is comparable to that of the IMMW estimator (imputed list + modified weight NOL). Figure 2 displays the results obtained for the planted corn acreage.

3.3 Hog Items

For the hog items, the list coverage ranges from 0.75 to 0.99, except for a few regions having marginal number of hogs and pigs. Many of the large regions have list coverage of at least 0.9. The R-RMSD varies considerably across regions. For estimating total number of hogs and pigs, there is some decrease in relative R-RMSD as list coverage increases as depicted in Figure 3. The performance of the ADMW estimator (revised adjusted list + modified weight area) is similar for the base and non-base periods. The performance of the ratio estimator is comparable to that of the ADMW. The performance of the RWMW estimator (reweighted list + modified NOL) is similar.

The performance of the ratio estimator is comparable to that of the ADMW and RWMW estimators for all hog items estimated. The R-RMSD decreases as list coverage increases for estimating number of sows furrowed, but not for estimating hogs under 60 lbs, hogs from 60-119 lbs, hogs from 120-179 lbs and hogs over 180 lbs.

References

[1] Chhikara, Raj S., Perry, Charles R., Deng, Lih-Yuan, Iwig, William C., Spears, Floyd M. and Cowles, Susan, "Post-Stratification and Efficient Estimation in U.S. Agricultural Labor Surveys," ASA Proceedings of Survey Research Methods, 1995.

[2] Kott, P.S. "Mathematical Formulae for the 1989 Survey Processing System (SPS) Summary". USDA-NASS SRB Research Report No. SRB-90. May, 1990.

[3] Perry, Charles, Chhikara, Raj, Deng, Lih-Yuan, Iwig, William and Rumburg, Scot, " Generalized Post-stratification Estimators in the Agricultural Labor Survey," USDA-NASS SRB Research Report No. SRB-93-04, Washington, D.C., July 1993.

[4] Perry, Charles, Chhikara, Raj, Spears, Floyd M. and Cowles, Susan, "An Evaluation of List-Only, Reweighted, and Other Estimators for U.S. Agricultural Labor Surveys," USDA-NASS RD Research Report No. RD-97-06, October 1997.

[5] Spears, Floyd M., Chhikara, Raj S., Perry, Charles R., Iwig, William C. and Cowles, Susan, "Agricultural Labor Estimation Using Only List-Frame Sampling," ASA Proceedings of Survey Research Methods, 1996.

[6] Spears, Floyd M., Chhikara, Raj S., Perry, Charles R., and Cowles, Susan. "An Evaluation of Alternative USDA Agricultural Labor Survey Estimators". ASA Proceedings of Survey Research Methods, 1997.

Figure 1. Relative R−RMSD verus List Coverage
Number of Hired Workers

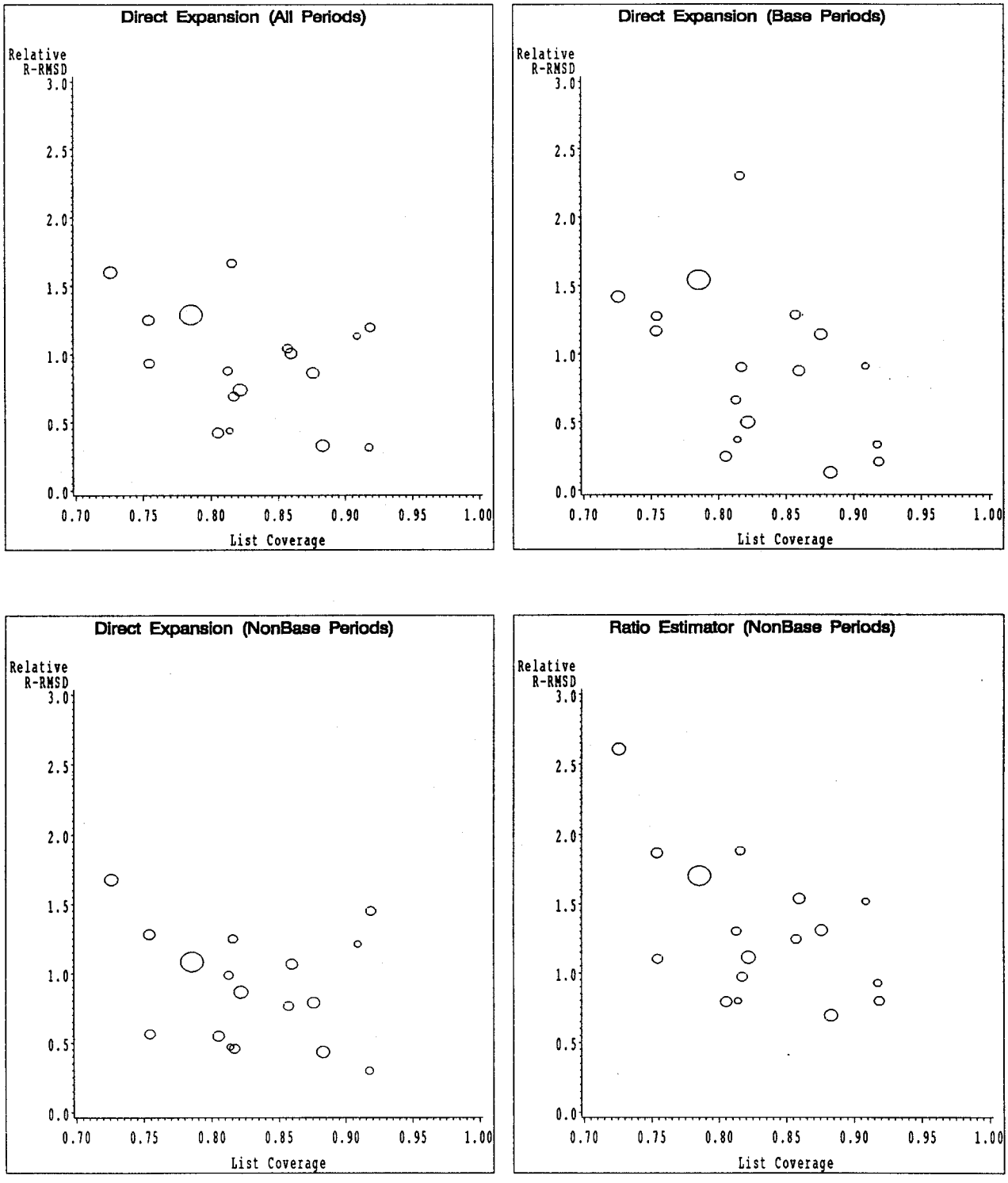

Figure 2. Relative R-RMSD verus List Coverage

Corn – All Planted Acreage

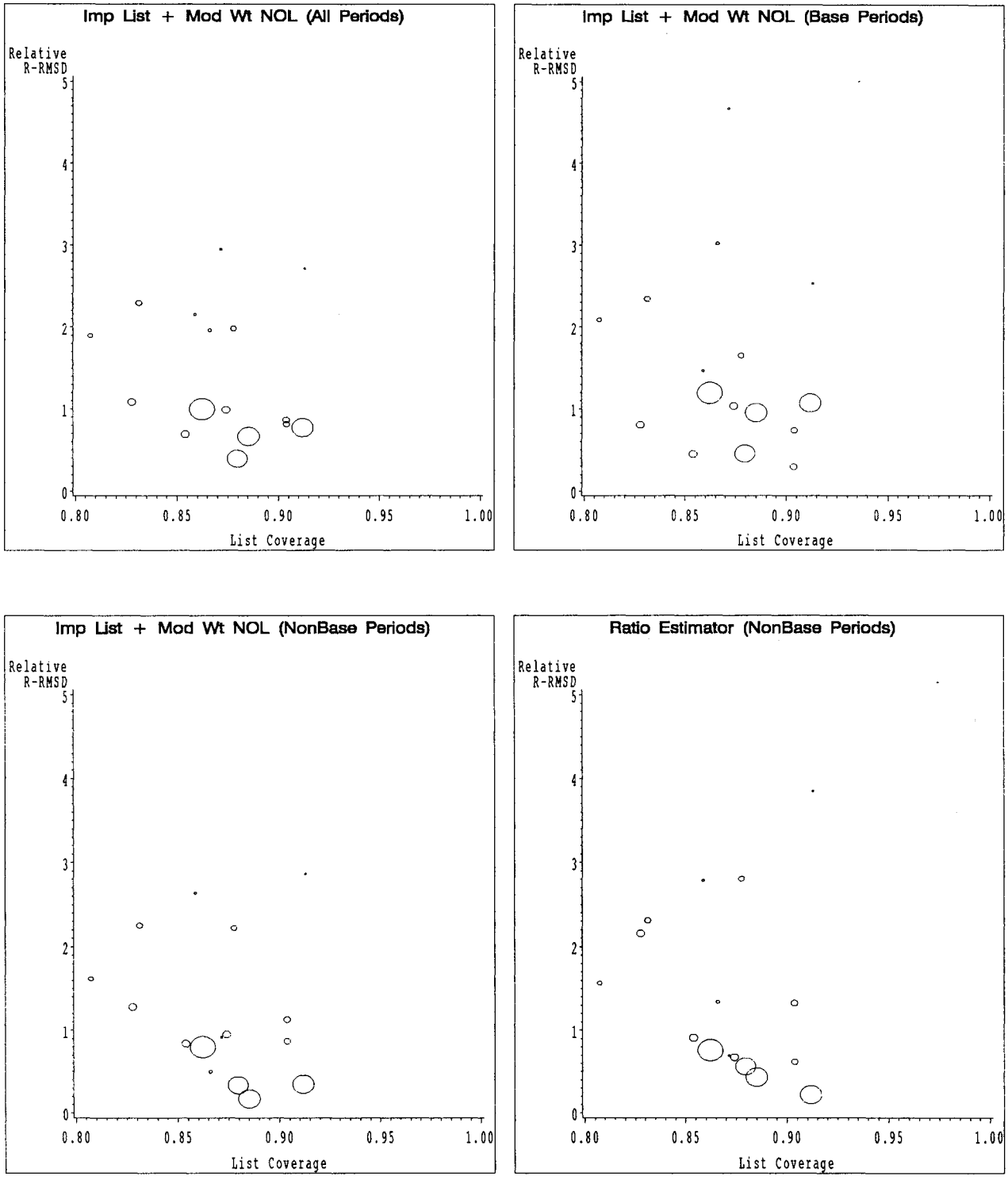

Figure 3. Relative R−RMSD verus List Coverage

Hogs & Pigs