United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research Division

SRB Research Report
Number SRB-93-05

July 1993

ANALYSIS OF A GENERALIZED POST-STRATIFICATION APPROACH FOR THE AGRICULTURAL LABOR SURVEY

Scot Rumburg
Charles R. Perry
Raj S. Chhikara
William C. Iwig

ANALYSIS OF A GENERALIZED POST-STRATIFICATION APPROACH FOR THE AGRICULTURAL LABOR SURVEY, by Scot Rumburg, Charles R. Perry and Raj S. Chhikara*, William C. Iwig, Sampling and Estimation Research Section, Survey Research Branch, Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC 20250-2000, July 1993, Report No. SRB-93-05.

ABSTRACT

The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture conducted monthly Agricultural Labor Surveys (ALS) from April 1991 through November 1992 in major agricultural labor states to estimate, among other items, the number of hired workers. The survey is still being conducted; however, it has become a quarterly survey for all states beginning in January, 1993. A multiple frame (MF) consisting of both a list and an area frame was utilized. The list frame does not have complete coverage of all agricultural operations while the area frame does. In a MF survey the area frame is used to account for the lack of coverage by the list. In this study all list respondents and all area respondents found to be non-overlap (NOL) with the list are post-stratified to construct a MF post-stratified estimate. A list-only approach is constructed that accounts for non-coverage of the NOL through post-stratification of list respondents based on the farm value of sales, type of farm and peak number of workers expected during a year. List-only estimates are obtained for each post-stratum and expanded using an estimated population count for each post-stratum derived from the area sample of the June Agricultural Survey (JAS:A). For this procedure to be effective the list respondents must also be representative of NOL respondents for all post-strata. Another approach considered is to obtain post-stratified ratio estimates based on the previous quarter estimates. For this technique to be effective, list respondents need only represent the rate of change for the NOL respondents within a post-stratum. We found that list respondent values do not necessarily represent NOL respondent values within a post-stratum; however the two types of respondent values appear to be similar in their rate of change from one quarter to the next.

KEYWORDS

Multiple frame; Non-overlap modeling; Post-stratified estimators; Combined ratio; Ratio estimate; Taylor series linear approximation of variance.

ACKNOWLEDGEMENTS

The authors would like to thank Fred Vogel for putting forward the "strawman" concept and thus stimulating our thinking about new approaches to estimation. We also thank Gary Keough, Stan Hoge, Tom Kurtz and Dan Ledbury for their time in relating the methodology of the Agricultural Labor and June Area Surveys.

* Raj S. Chhikara is Professor, Division of Computing and Mathematics, University of Houston - Clear Lake, Houston, Texas 77058

TABLE OF CONTENTS

SUMMARY

The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture conducted monthly Agricultural Labor Surveys (ALS) from April 1991 through November 1992 in major agricultural labor states to estimate, among other items, the number of hired workers. The survey is still being conducted; however, it has become a quarterly survey for all states beginning in January, 1993. This multiple frame (MF) survey uses a list and an area frame to select respondents from the target population consisting of agricultural operations with labor. The list frame does not have a complete coverage of the target population, but the area frame does, compensating for the incompleteness of the list frame. Respondents which are selected from the area frame but are not contained on the list frame are labeled as non-overlap (NOL). A MF estimate is obtained by adding the list frame component estimate and the NOL component estimate. Estimates are set at the state level for larger agricultural labor states and regionally for the remaining smaller states. The list frame component accounts for the bulk of the MF estimate for the number of hired workers while only a small portion of the MF estimate is attributable to the NOL. However, the NOL often accounts for a larger portion of the overall variance of a MF estimate. Among other factors, a much smaller sample size for the NOL causes its estimator to be unstable. A post-stratification of all NOL and list respondent values is considered to achieve a MF post-stratified estimator that might be more stable than the current MF estimate. Proposed is another approach of constructing a list-only post-stratified estimator which would allow the list respondents to represent the entire target population. One estimation method studied is as follows:

1) List-only respondents are post-stratified following a list-only labor survey. Post-stratification variables to be used are annual farm value of sales, type of farm, and peak number of workers expected during the year.
2) Estimates are made for each post-stratum for the variable of interest - number of hired workers. Generally this is an average number of workers per agricultural operation and can be calculated using either a weighted or unweighted average.
3) Population sizes for the post-strata are estimated using the area sample from the June Agricultural Survey (JAS:A). This survey has a larger sample size than an agricultural labor survey, and is expected to provide more accurate target population size estimates for the post-strata.
4) Post-stratum estimates are expanded to the state level by aggregating the product of the estimated size of each post-stratum as calculated in Step 3 and its estimated number of workers per operation as calculated in Step 2 for all post-strata.

For this methodology to be effective, list and NOL respondents that are classified into the same post-stratum must be similar with respect to the number of hired workers (that is, have similar conditional distributions).

A second method studied was the use of a post-stratified list-only combined ratio estimator. The combined ratio does not require that list operations be similar in size of hired workers to the

NOL operations within post-strata. It does require, however, that the rate of change for number of hired workers be similar between the list and NOL across surveys used to construct the ratio.

If reliable estimates could be produced using either of the above procedures, several problems with the current NASS survey methodology could be alleviated. For example, elimination of area frame NOL sampling for any survey would result in reduced respondent burden, reduced variability of sampling weights, and reduced need for checking its overlap with the list.

Analysis of list and NOL respondents for the monthly agricultural labor surveys showed that list respondents were distributed differently from NOL respondents. These differences occurred even within post-strata expected to be homogeneous regardless of which frame the respondents originated from. This disparity means that the list-only post-stratified estimators would be biased in estimating the entire target population. Biases were particularly acute for unweighted estimators.

Analysis of list-only post-stratified survey to survey combined ratio estimates provided some promise for achieving unbiased list-only estimates. Rates of change between list and NOL respondents appear to be similar across surveys used to obtain ratios. Ratio estimates for two states showed improved accuracy and precision over the MF direct expansion (MF-DE) estimator.

The list-only post-stratified combined ratio also produced a more accurate estimate than was produced by the list-only survey design combined ratio. The list-only survey design ratio estimate was produced from list-only DE survey totals and then multiplied by a base population indication. However, analyses of the superiority of any post-stratified estimate to a non-post-stratified estimate must also consider the increased complexity involved with post-stratification.

Attempts at a simplified post-stratification scheme whereby post-strata were defined using only the variable of the peak number of workers (the variable best correlated with the number of hired workers) also resulted in a loss of precision for all estimators examined.

The list-only post-stratified combined ratio provided more accurate and more precise estimates than the MF DE. Analyses indicated however that an effective list-only survey design combined ratio estimator could result in nearly all goals currently sought through the post-stratification proposal without the need for post-stratification. Further study of this type of estimator is required to evaluate its effectiveness. Furthermore, its performance can be easily tracked, and it should be evaluated in its ability to estimate across quarters. It should be noted that the bulk of this study is based on a monthly agricultural labor survey series and the inference made must now be extended to a quarterly series.

INTRODUCTION

A multiple frame approach, employing both a list and an area frame, has long been a cornerstone for many of the agricultural surveys which are conducted by the National Agricultural Statistics Service. Area frame responses often account for a majority of the total variance for multiple frame estimates but add little to the total indication. For this reason and others, it was recommended that a study be performed into alternatives to the current multiple-frame approach for administering surveys. A post-stratification approach, labeled "strawman" whereby list respondents could be used to represent the entire target population, was recommended for consideration (Vogel, 1990a, 1990b and 1991). Kott (1990a and 1990b) elaborated on the proposal and outlined the two model-based estimators, their variance and potential bias. Perry, et al. (1993) provide an estimation method for the variance of a generalized post-stratification estimator based on its linear approximation using a Taylor series expansion.

A list-only approach would result in reduced respondent burden, reduced variance of estimates, and simplified survey procedures. List-only estimators do not come without a price, however. Bias is an intrinsic part of any estimate of a target population derived from a sample which ignores a specific subgroup. Bias can be reduced or eliminated provided the sample can accurately model the non-sampled subgroup.

The survey data from the Agricultural Labor Survey series from July 1991 through June 1992 were used to investigate the alternative estimators. This investigation has centered on two states, California and Florida, which, together, make-up a quarter or more of the total U.S. agricultural labor force in any given month. Their large sample sizes allow for effective list-only modeling as well as for verification of the accuracy of any model-based estimates as compared with multiple frame estimates. Because of the varied nature of agricultural commodities it is difficult to draw conclusions beyond the range of the data about any model-based estimates produced.

This interim report characterizes the major results to-date of the ongoing investigation into models employing list-only estimators. It reports findings and provides insight into the potential estimation process in the future. Several different types of estimators were investigated and presumably some of the results obtained would hold for other commodities, as well. The investigation into the effectiveness of list-only estimators in other commodities is the next natural step in this study and currently work is being done towards that goal. Preliminary emphasis in the study was placed on previous research in an effort to understand issues concerning post-stratification.

METHODOLOGY AND JUSTIFICATION

General NASS Survey Methodology.

The National Agricultural Statistics Service (NASS) conducts numerous surveys with regard to agricultural commodities and related subjects. Depending on the commodity, estimates are produced annually, quarterly or monthly. The majority of these surveys employ a multiple-frame (MF) methodology using both a list frame and an area frame. The survey estimates are used in conjunction with other information by the Agricultural Statistics Board (Board), a designated committee of NASS statisticians, to develop official statistics. The official statistics for a commodity, which are set at various levels (usually national, regional and/or state), reflect the expert judgement of the designated Board members based on all available survey and administrative data.

The list frame is stratified based on known data about agricultural operations with regard to the survey item(s) of interest. The list frame is not a complete listing of all agricultural operations. For the 1992 survey year beginning in June, the entire NASS list frame is estimated to contain 56% of all agricultural operations (often referred to simply as farms) and 81% of all land in farms. Emphasis is placed on locating and retaining larger agricultural operations with higher values of annual sales and/or possessing larger acreage.

The area frame is stratified based on the agricultural intensity of a region. Unlike the list frame it has complete coverage of all agricultural operations in the United States. The area frame compensates for the incomplete coverage of the list frame and allows for known probability sampling and unbiased estimates.

All area reporting units (agricultural operations) in June are classified as either overlap (OL) or as non-overlap (NOL) with the list frame. All operations found to be NOL are divided into several sampling pools to be used in follow-on surveys for the year. The list frame takes precedence over all OL operations when a multiple frame (MF) estimate is calculated. A MF estimate is obtained by summing the list frame sample component estimate with the area frame's NOL sample component estimate. In most cases, the list frame provides about 75% of the total MF estimate while the NOL component adds only about 25%. However, the NOL estimate is often a large contributor to the overall variance of the MF estimate, due to both the high variability of sampled units for many commodities and the sizable sample weights associated with small sampling fractions. The post-stratification approach investigated in this paper is an attempt to improve the reliability of the NOL component of MF estimates.

Benefits Derived from NOL Modeling.

The proposed list-only estimator based on modeling of the NOL population represents a departure from the present NASS survey design and estimation methodology. Attempts at modeling and estimations for the NOL have become a notable research objective for three primary reasons: (1) the NOL sample units are highly burdened, (2) the current NOL estimates are often unreliable, and (3) the presence of NOL sample units increases the complexity of a survey.

One benefit of modeling the NOL would be a reduction in respondent burden. Reducing respondent burden has been an issue at NASS for many years and has become a goal with respect to the administration of many surveys. At a recent NASS Program Planning Committee (1992) it was recommended that reducing respondent burden become a high priority. Complete replacement of NOL sampling by modeling would be impossible since trends in commodities require that overall model accuracy would have to be checked at specific intervals during the survey series. However, such concepts as replacing NOL enumeration by NOL modeling for monthly surveys with continued enumeration of the NOL sample quarterly, or modeling the NOL every other quarter for quarterly surveys, would result in significant reductions in respondent burden.

For nearly all surveys, respondents within the NOL portion can and do represent some of the largest expanded response values for the variables of interest. Since a larger portion of the target population is present on the list frame and the list frame is more heavily sampled, the sampling can be better controlled by restricting it to the list frame only. Sampling weights for the list frame are nearly always less than 100. Alternatively, the area frame, from which the NOL estimates are derived, is not sampled as heavily. This is true even in high agricultural land use strata. Sampling weights for area frame respondents are usually above 100 and weights above 1,500 are not uncommon. When positive responses for commodities are found within an area frame sample unit, the expanded values can often be quite large, even for a small measure of a commodity. Modeling the NOL would certainly lead to reductions in the number and magnitude of outliers, conceivably leading to more reliable estimates.

Modeling the NOL would also result in some reduction in overlap determination between the list and area frames. All area respondents are labeled as OL or NOL during the June Agricultural Area Survey (JAS:A). This insures that no operation is counted twice in a MF estimate and enables those respondents designated as NOL to be allocated to follow-on agricultural MF surveys for the coming survey year. This is a time-consuming task which must be undertaken every year in June as well as any time an NOL operation is moved into the overlap portion in a follow-on MF survey. Though checking for the overlap following the JAS:A accounts for most of this work and would not be eliminated with NOL modeling, overlap checks for follow-on surveys based on the list-only methodology would be eliminated.

Though list-only surveys cannot completely replace the current MF approach, their use, however limited, would provide some relief from response burden and improve upon the stability of the estimate. List-only surveys will only be appropriate, however, if list respondents are representative of the entire target population.

Overview of List-Only Post-Stratification Methodology.
Post-stratification was proposed as an approach to modeling NOL operations through list operations. See Appendix A - Part 1 for an overview of standard post-stratification methodology. Post-stratification is currently being used in the January Cattle-on-Feed survey within NASS and is under consideration for other surveys. This procedure is being proposed as a means of maximizing usefulness of the area and list frames while simplifying survey procedures and reducing respondent burden on the NOL. This list-only procedure assumes that list respondents alone are representative of the entire list/NOL population within each, suitably defined post-stratum. NOL modeling can then be accomplished through the use of appropriate population counts within each post-stratum. In the simplest form of the proposed estimator, an average of unexpanded list sample responses would be computed within each post-stratum with each respondent having a weight of one (unweighted). An alternative procedure would be the use of list-only sample responses with each respondent having weight equal to the inverse of its sampling fraction (weighted). For details of a generalized post-stratified estimator, one may refer to Perry, et al. (1993).

Steps involved in the construction of the post-stratified estimator are as follows:

1) The population count for each post-stratum is estimated using the JAS:A. The total estimated population count is fixed until the next JAS:A produces a new estimate.
2) Once population counts are estimated, a survey for the commodity of interest is conducted. Respondents are post-stratified based on classification variables obtained during this survey. If a MF survey is conducted for the commodity of interest, the June estimated population counts from the JAS:A could be used to provide a more precise estimate, assuming the JAS:A provides better information regarding post-stratum population totals than does the commodity survey. For a list-only survey where NOL modeling is required, the June estimated population counts are a necessity, since the list-only respondents would provide an estimate of only the list population.

3) Following proper post-stratification of all surveyed respondents, estimates are obtained within each post-stratum. This is usually an average, though ratio or proportion are other possible quantities.
4) Once post-stratum estimates have been made, they are expanded based on the product of the estimate and the estimated population count for that post-stratum. These post-stratum expanded estimates are then summed to obtain an estimate of the total for the target population.

For post-stratification to be effective in improving upon a commodity estimate, the following conditions must hold:

- Subgroups created within a post-stratum should form similar distributions regardless of which frame they originate from; and these distributions should appear different across post-strata.
- Subgroups must be mutually exclusive for respondents and have complete coverage of the target population.
- Information obtained from a respondent during the survey process to be used to post-stratify that respondent should not have been used in the initial survey design, but it should be well correlated with the variable of interest.
- Population counts must be accurate for each post-stratum.

Post-stratification for the Agricultural Labor Survey (ALS) was based on three classification variables: (1) The peak number of agricultural workers an operation expected to have over the course of a year (Peak), (2) the annual farm value of sales for agricultural goods (FVS), and (3) the type of farm operation (FType). These classification variables were selected based on their ability to describe distinct post-stratum populations and to correlate with the number of hired agricultural workers, which is the variable of interest. Basic strategy to obtain homogeneous post-strata populations involved selecting class boundary values for the two numerical classification variables (Peak and FVS), and creating combinations of the third categorical variable (FType). No more than twelve total post-strata could be created in order to maintain adequate sample counts for all post-strata across all surveys. Depending on cutoff values and FType groups selected, fewer post-strata could be constructed. An attempt was made to maintain a minimum of 20 respondents per post-stratum for all post-strata, though this was not always possible. (For more information on how post-strata were defined, see Appendix B.) The following criteria were used to evaluate post-strata as defined by the cutoff values:

1) Minimize total variance.
2) Define distinct populations within post-strata.
3) Maintain adequate sample counts within all post-strata.

Four measures were used in evaluating each estimator:

- Accuracy (BIAS) of an estimator - a measure of the average deviation of the post-stratified estimate from the actual population value over the survey year. The true population value was assumed to be the monthly total number of hired workers at the state level as published by the Agricultural Statistics Board. (Note: For purposes of this report, accuracy is defined as bias from the target population true value and not as the total of both bias and sampling variability (i.e., mean squared error)).
- Precision (AVE CV) - a measure of average sampling variability over the survey year. It was measured by the average coefficient of variation.
- Mean absolute deviation or mean error (ME) - a measure of both accuracy and precision. It was measured by the average absolute deviation over the survey year.
- Maximum absolute deviation (MAX) - a measure of the largest deviation over the survey year. It was measured as the largest deviation between the survey estimate and the Board specified value over the survey year.

Additionally, another estimate of the total number of hired workers in California was available through the Employment Development Department (EDD), what would be the Department of Labor in many states. EDD conducts a probability labor survey each month which provides a state level agricultural labor estimate (referred to as Administrative Data) and was used as a comparison of the true population value in the final summary results.

Post-Stratified Estimators.

Though post-stratification is often used as a variance reduction tool in a design unbiased survey, it can also compensate for the undercoverage of a target population by a particular selected sample. In the case of a list-only approach, post-stratification must compensate not only for inaccurate post-stratum coverage by a sample, but also for the complete lack of a particular population (i.e. NOL). This would result in biased estimates if the list and NOL do not act similarly. For the approach explored in this paper, the list frame is used for the selected sample for follow-on surveys, and the sample is then post-stratified to obtain post-stratum estimates.

Population counts for each post-stratum are determined once yearly from the JAS:A and then are fixed for all follow-on surveys during the year. As mentioned previously, post-stratum estimates could be either unweighted or weighted, depending upon the type of modeling desired.

In the case of unweighted list responses,the estimator of the characteristic of interest Y is of the form:

$$\hat{Y}^{PS}_{(unex)} = \sum_{\substack{all\,k \\ post\text{-}strata}} \left(\hat{N}_{k\,(June)}\right) \cdot \overline{y_k} = \sum_{\substack{all\,k \\ post\text{-}strata}} \left(\hat{N}_{k\,(June)}\right) \cdot \left(\frac{1}{n_k} \sum_{i \in U_k} y_i \right), \qquad (Eq.\,1)$$

where

$\hat{N}_{k\,(June)}$ = k^{th} post-stratum population size estimate from the June survey (JAS:A),

n_k = k^{th} post-stratum ALS List sample size, and

U_k = the set of all useable ALS List sample reporting units in the k^{th} post-stratum.

Similarly, a weighted estimator of Y is of the form:

$$\hat{Y}^{PS}_{(expd)} = \sum_{\substack{all\,k \\ post\text{-}strata}} \left(\hat{N}_{k\,(June)}\right) \cdot \overline{y}_{k\,(wtd)}$$

$$= \sum_{\substack{all\,k \\ post\text{-}strata}} \left(\hat{N}_{k\,(June)}\right) \cdot \left(\frac{1}{\hat{N}_{k\,(Labor)}} \right) \cdot \sum_{i \in U_k} w_i y_i = \sum_{\substack{all\,k \\ post\text{-}strata}} \left(\hat{N}_{k\,(June)}\right) \cdot \left(\frac{\sum_{i \in U_k} w_i y_i}{\sum_{i \in U_k} w_i} \right), \qquad (Eq.\,2)$$

where

$\hat{N}_{k\,(Labor)}$ = k^{th} post-stratum population estimate from the ALS List sample,

w_i = i^{th} List sample reporting unit weight, and

other variables are defined analagous to Equation 1.

The list-only post-stratified estimators differ from the standard post-stratification estimator given in Appendix A since population counts ($\hat{N}_{k(June)}$) in Equations (1) and (2) are estimated rather than known. This means that the population size estimate adds variance to the overall list-only post-stratified estimate. But any increase may be offset or compensated for by the reduction in variance due to post-stratification. For both the June population count estimates ($\hat{N}_{k(June)}$) and the ALS population estimates ($\hat{N}_{k(Labor)}$) used in Equation (2), the population counts are the aggregate total of all sampling weights associated with all usable respondents. For the June counts this weight is the product of the sampling fraction, non-response adjustment and the percentage of total farm acreage contained in the sampled area unit. (Note: Normally non-response is not recognized for this survey. Respondents that refuse or are inaccessible are manually imputed. For this study, however, only respondents which contained no imputation were used since it was felt that they would better delineate the target population.) For the labor population count derived using only list respondents, the weight associated with the respondent is simply the sampling fraction adjusted for non-response. Note that the June count is an estimate of the total number of farms in the target labor population, whereas the labor list-only count is an estimate of the total number of farms on the list.

Post-Stratified Ratio Expansion Estimators.

A second estimator, based on list-only post-stratification, that was evaluated was a survey to survey (i.e.current to previous quarter) ratio expansion estimate. For more information on ratio and ratio expansion estimators, see Appendix A - Part 2. This methodology does not require that the list and NOL respondents have the same distribution within a post-stratum, but that they exhibit the same rate of change between the two survey periods. This implies that comparability of the means from either frame within a post-stratum is no longer important. It is the comparability of the rate of change for each frame over the time period where ratios are produced that is important. The ratio estimator is based on post-stratified list-only survey totals (a combined ratio estimator). Only useable matched respondents were used (useable matched respondents are defined as reporting units which appear in both surveys, both having valid responses). The product of this rate of change ratio estimate and the base quarterly Board estimate produced a ratio expansion estimate for the number of hired workers. It is of the form:

$$\hat{Y}^{PS}_{RE_1} = Y_B \cdot \hat{R}^{PS} = Y_B \cdot \frac{\hat{Y}^{PS'}}{\hat{X}^{PS'}} = Y_B \cdot \frac{\displaystyle\sum_{\substack{\text{all } k \\ \text{post-strata}}} \left(\frac{\hat{N}_k}{\hat{N}'_k} \sum_{i \in m_k} y_i \right)}{\displaystyle\sum_{\substack{\text{all } k \\ \text{post-strata}}} \left(\frac{\hat{N}_k}{\hat{N}'_k} \sum_{i \in m_k} x_i \right)}, \qquad (Eq.\ 3)$$

where

Y_B = Board Indication for number of hired workers from the previous quarter,

$\hat{N}_k$ = k^{th} post-stratum estimated population size from the June Survey (JAS:A),

$\hat{N}'_k$ = k^{th} post-stratum estimated population size from all matched useable ALS List sample reporting units,

m_k = set of all matched useable List sample reporting units in post-stratum k,

y_i = i^{th} matched useable List sample expanded value from the current survey, and

x_i = i^{th} matched useable List sample expanded value from the previous quarter.

Ratio, and likewise ratio expansion, estimates are most efficient when produced at a level which maintains homogeneity, have a large sample size, and where the variable of interest is well correlated between the two surveys.

Multiple-Frame Post-Stratified Estimators.

Considered were two multiple-frame (MF) post-stratified estimators, an unweighted and a weighted one, similar to those given in equations (1) and (2) respectively, for the list-only estimators. The difference between MF and list-only estimators lies in the computation of the unweighted or weighted average response of a post-stratum. For the MF case, the average y_k was based on sample values obtained from both NOL respondents and list respondents that belonged to the k^{th} post-stratum.

Variance Estimation.
An exact variance formula that would encompass numerous alternative post-stratification schemes was intractable. In order to calculate a variance formula which would be easily computable, a Taylor series linear approximation to the overall variance was obtained by Perry, et al. (1993). Because post-stratification approach involves a ratio estimate, this approximation will, in general, underestimate the true variance slightly unless the sample size is large. Variances for the post-stratified ratio estimates were computed using a similar methodology of linear approximation to the variance. For details, refer to Perry, et al. (1993).

DATA

The area portion of the June Agricultural Survey (JAS:A) was used to estimate population counts (number of farms at the state level) for the post-strata in the area weighted estimator. For an overview of the weighted estimator, see Nealon (1984). The JAS:A has the largest area sample of all NASS surveys and is thought to provide the best area estimate for population counts because of its large sample size. One classification variable which was not on the JAS:A prior to 1991 - the peak number of workers expected over the next year - was placed on the 1991 JAS:A questionnaire. This allowed population units to be classified identical to any post-strata that could potentially be defined by the three classification variables. The June population counts are estimated once yearly; the total size estimate is then fixed. Thus, population counts within post-strata do not change unless post-strata definitions change, and the sum of all post-strata population counts must equal to the fixed population total for the most recent JAS:A estimate.

The JAS:A weighted estimate for total number of farms was somewhat less than desirable in precision and accuracy. Population counts estimated using the June area weighted estimator underestimated the Board number of farms in California by 12% and overestimated the number of farms in Florida by 9%. This is a matter of concern since the post-stratified estimator is sensitive to inaccuracies in population counts. For the ratio estimators, this is not as much of a problem since only the rate of change is what matters, and the inaccuracies in counts, in general, tend to average out in the ratio. It is, however, important that the population estimate be correctly proportioned across post-strata.

The Peak Worker variable currently on the JAS:A was added to five state's JAS list questionnaires (CA, MI, NC, TX, and WA) in 1992 in order to evaluate the benefits of a MF estimate for farm numbers. It is presumed that the MF estimate for farm numbers will be more stable than the area weighted estimate.

The ALS provided the variable of interest, the number of hired workers, as well as a sample estimate of the list or total population count depending on whether a list-only or a MF estimate was obtained. All NOL respondents allocated to the ALS (approximately 40% of the all JAS:A NOL respondents) were sampled in the July ALS and post-stratification for FVS, FType and Peak were determined. Post-stratification of list respondents was made throughout the survey year whenever they were first selected for the ALS.

The ALS series had four states (CA, FL, NM & TX) that were sampled in all 12 months, seven seasonal states (MI, NC, NY, OR, PA, WA & WI) sampled monthly from April through October and 49 states sampled quarterly beginning in July. Alaska is estimated once annually in July. Because of the enormity of the analysis it was decided to concentrate on Florida and California for the present analysis study.

One large outlier was present in the December Florida data - a list respondent with FVS less than $2,500 yet with a Peak equal to 200. In December this operation reported 62 workers which expanded to just under 1,700 workers, and the influence can be seen in several survey estimates for that month.

One problem associated with the ALS is the seasonality involved with much of agricultural labor. Because of the labor variation it would seem appropriate to allow post-strata definitions to vary from month to month. For example, the flexibility of farm types to be regrouped, depending on the season, may define better post-strata populations across the year. However, since thirty-eight of the states are sampled only on a quarterly basis for the entire year, and six of the remaining eleven states are sampled quarterly for a portion of the year, much of the continuity of the ALS series is lost. This fact, coupled with the increased work involved in redefining monthly post-strata and maintaining historical information, would make such a task difficult.

Adding to the problem of seasonality is the transitory nature of much of the agricultural labor force. The ALS series provides a snapshot of agricultural labor - principally a specific week within the month. For many operations a peak labor force is brought in on a short term basis for harvesting, working livestock or other needs for which increased temporary labor is required. These laborers are often employed contingent on numerous conditions including weather, economic factors, and availability. It is conceivable that a peak labor force could be hired and dismissed in a matter of days and never be recorded by the survey. Though it is expected that the randomized sample of "hits and misses" will average out in a large sample and provide an unbiased estimate, the transitory nature will assuredly affect precision by adding additional variability to the estimate.

Another problem with the ALS is the presence of subtracts - additional operations associated with an area respondent after the JAS:A. All subtracts must be combined to the tract level in the ALS in order that farms be defined identically with the JAS:A. Although subtracts in the ALS are rare, when they occur they must be combined to the tract level in order that they represent a true farm unit in the same sense as a JAS:A farm unit. Subtracts can, and often are, of different farm types, value of sales or peak workers. This makes classification of the combined tract level operation into a distinct post-stratum difficult. A priority scheme was instituted which captures all labor data and classifies the combined "farm" into a specific post-stratum. This, however, leads to additional variability in the estimate.

The ALS series also presents several problems in terms of post-stratification. First, the sample sizes are relatively small even for the larger labor states. It is essential that all post-strata have enough respondents to make an accurate estimate for each post-stratum. This requires a continual trade-off between increases in the total number of post-strata to help define distinct populations and decreases in sample counts per post-stratum. For 1991, state level estimates were made only for the monthly labor states and the seasonal states surveyed monthly from April through October. If post-stratification methodology were to go operational, it may be necessary to make regional estimates for combinations of these states in order to maintain samples sizes within post-strata and produce accurate estimates. This means an allocation scheme would have to be devised if state level estimates were still required. This last question was avoided for the time being by limiting our initial investigation to California and Florida. These two states have the first and third largest ALS sample sizes respectively.

Another post-stratification problem created by the ALS data was post-stratification variables which were unassignable. Particularly problematic is the "Don't Know" (DK) response for Peak. In one classification scheme DKs were imputed based on FVS and FType, while in a second scheme DKs were classified as a distinct category.

Estimation of agricultural labor is an arduous task even under the best of conditions. The ALS seeks information from respondents over the span of one week in a given month and often only quarterly. Because of the transitory nature of much of agricultural labor the ALS is heavily dependent on a random sampling of regions, farm types and sizes. A host of reasons combine to create this transientness, including weather, geographical differences in crop progress due to geography or species, different farm types and much more. Reliance on randomization to compensate for all these factors is reinforced by the loss of overall sample size when the NOL is modeled. This can only make the analyst's task of achieving accurate estimates that much more difficult.

RESULTS

Preliminary Research - Simulation Studies.

Simulation studies provided a theoretical perspective into several aspects of the post-stratification methodology. Use of simulated data provides one with a known population target parameter which one wishes to estimate through sampling. This luxury is not afforded in actual surveys where the true number is seldom known. These studies show the effect of altering one or more variables by measuring the distance (usually a combination of bias (accuracy) and sampling error (precision)) between the estimator and the known true value. However, it should be noted that none of the simulation studies employed exact design survey parameters or methodology.

Population Counts and Post-Strata Estimates.

Initially some textbook examples worked out by Flores-Cervantes (1991a, 1991b and 1991c) provided insight into potential error costs associated with using estimated population counts versus known true counts. These examples showed, through the use of data simulated under a simplistic model, that small deviations in the estimated population count from the actual value could result in a significant increase in the mean squared error (MSE). The sampling variance associated with the post-strata means had a lesser affect on the MSE of the post-stratified estimator than did the variance associated with post-strata population counts.

An extensive simulation study was performed by Perry, et al. (1993) to evaluate numerically the performance of several post-stratified estimators. The numerical evaluations showed that the performance of a post-stratified estimator is largely a function of the sample size used to estimate the post-stratum sizes, the sample size used to estimate the post stratum means of the variable of interest, and the ratio of these two sample sizes. The relative efficiency of the post-stratified estimators all increased as the ratio of the two sample sizes increased. Given the sample size for the follow-on survey, the sample size for the base survey should be at least twice as large for gains in efficiency. Moreover for post-stratification to be effective, the entire sample size in the follow-on survey should be at least 50 (preferably much larger) with the sample size in all post-strata at least 10 (preferably 20 or more).

Verification of the Variance Approximation.

Simulated results provided evidence for the validity of the Taylor series linear approximation to the overall actual variance of the post-stratified estimate, see Perry, et al. (1993). Using known population variances and a list-only sampling scheme, repeated sub-sampling of the population showed that the linear approximation underestimated the actual MSE within 10% for large sample cases. This reflects well on the variance estimates made in our evaluations despite the expected bias resulting from the ratio estimate and the use of a linear approximation to estimate the variance of a non-linear function.

Effect of Sample Size Differences Between the JAS:A and ALS.

Post-stratification becomes more efficient as the ratio of the JAS:A sample size to the ALS sample size increases (ie., when the JAS:A sample is much larger than the ALS sample) as demonstrated by the simulation study in Perry, et al. (1993). This is because the information within the JAS:A sample with respect to farm counts provides a more precise and more accurate estimate than can be obtained from the smaller ALS sample. The JAS:A sample size is predominantly a function of the area size of a state and its agricultural intensity while the ALS sample size is a function of the amount of state level agricultural labor (at least for the list frame). It is assumed that the ALS MF sample could better estimate the number of farms in the higher FVS and Peak post-strata since the list represents these populations well. The JAS:A would perhaps better estimate farm counts for the lower FVS and Peak post-strata where it better represents this portion of the target population. For simulation purposes though, these assumptions were not considered.

It was found that ratios greater than two would imply that the JAS:A sample would be effective in better estimating the post-strata population counts. Current sample size ratios of JAS:A to ALS range from 2.5 to 5 and thus the JAS:A sample appears to be large enough for efficient estimation of the population counts. This is especially true considering the sensitivity of the estimator to imprecise post-strata counts as discussed.

Comparison of List and NOL Respondents.

In order for post-stratification to be effective, all sampled units from the ALS that are placed in a post-stratum must come from the same distribution. This is true even though the unit may be coming from the list or the area frame. Note that all sampling units coming from the area frame are NOL units. Thus, it was important to assess whether the list and area frame distributions appeared similar within a post-stratum. For the most part it was found that a list unit does not behave like an NOL unit, even within a particular post-stratum. Within post-stratum, NOL units are more likely than list units to contain no hired labor and a smaller number of laborers on average. The list sampling scheme omits operations whose control data for FVS is less than $20,000. This reduces the number of smaller operations where minimal workers might be found. This does not preclude a list sample respondent from having FVS less than $20,000 and being post-stratified accordingly, as many are, since control data and actual data are not always well correlated. It will however, greatly reduce the number of list respondents found in smaller FVS post-strata (FVS not greater than $50,000). Most of these smaller operations will only be covered by the area frame and will be classified NOL. Thus within the smaller FVS post-strata one would expect the NOL number of hired worker responses to be smaller on average. However, within post-strata defined by larger FVS (FVS greater than $50,000) this reasoning does not explain the lower average values found for NOL responses.

TABLE 1. Counts and Mean Number of Hired Workers Within Post-Strata For the California July 1991 Agriculture Labor Survey

Post-strata Definitions			Survey Counts		Weighted Mean		Unweighted Mean	
FVS	FType	Peak	List	NOL	List	NOL	List	NOL
$1-50K	Crops&Misc	0-4	49	70	0.28	0.12	0.24	0.24
$1-50K	Crops&Misc	5+	1	1	0.00	0.00	0.00	0.00
$1-50K	Veg,Frt&Nut	0-4	70	79	0.21	0.11	0.17	0.13
$1-50K	Veg,Frt&Nut	5+	28	15	2.03	0.27	6.89	0.40
$1-50K	Dairy,Pltry, GH&Nrsry	0-4	4	0	1.36	.	0.75	.
$1-50K	Dairy,Pltry, GH&Nrsry	5+	0	0	.	.	.	.
$50K+	Crops&Misc	0-4	56	35	1.11	0.39	0.93	0.69
$50K+	Crops&Misc	5+	59	17	7.90	15.10	13.80	35.10
$50K+	Veg,Frt&Nut	0-4	57	15	0.70	0.48	0.77	0.67
$50K+	Veg,Frt&Nut	5+	249	38	15.30	5.29	38.80	16.30
$50K+	Dairy,Pltry, GH&Nrsry	0-4	30	0	1.15	.	1.30	.
$50K+	Dairy,Pltry, GH&Nrsry	5+	63	5	22.90	16.30	33.70	19.40

Cell counts and means for the weighted and unweighted response values by frame. *Note that the NOL cell averages tend to be smaller than the list averages and that the weighted cell averages tend to be smaller than the unweighted averages.*

Table 1 shows that the NOL has a lower average estimate within nearly all post-strata for California, whether one compares weighted or unweighted responses. Particularly troubling are the large FVS post-strata with open-ended peak workers (5 or more) and specifically the fruit, nut and vegetable post-stratum. The few NOL respondents which fell into this category had many fewer hired workers than did their list counterparts. This post-stratum represents slightly less than 10% of all California farms and the fruit, nut and vegetable FType category represents 57% of California farms. Another high FVS post-stratum, with Peak 5+ and FType Crop & Misc produced larger NOL average hired workers than did the list in both the weighted and unweighted category. This was due largely to one NOL respondent which reported 391 hired workers. For this post-stratum the *average* Peak reported was 27. The operation's expansion value was small (about 3), especially considering it was an NOL respondent, and the respondent's influence on the weighted average was lessened, allowing for much more comparable weighted average values between the two frames. Again though, this example points out problems one can expect with open ended post-strata.

The preponderance of NOL respondents with few or no workers can be seen in a comparison of NOL and list cumulative distributions in Figures 1a and 1b for the number of unweighted hired workers reported in the July 1991 ALS for California and Florida. For California (Figure 1a), 67% of all NOL respondents had no hired workers compared to only 35% of all list respondents, and 93% percent of the NOL respondents had 10 or fewer workers compared to only 75% of list respondents. In Florida (Figure 1b) the differences are even more acute with 85% of the NOL not having any hired workers versus only 43% for the list. No NOL respondent had more than 8 hired workers while 20% of the list respondents sampled were larger than this value. Although these distributions represent the entire July sample for both states, similar trends are found in nearly all post-strata.

Comparison of Unweighted and Weighted Means Within Post-Strata.
Table 1 also characterizes the difference between weighted and unweighted averages. Unweighted averages are consistently higher than weighted averages for both list and NOL respondents for nearly all post-strata. Since operations with larger numbers of hired workers are sampled at a higher rate, and because operations with larger numbers of workers tend to represent fewer number of farms, the sampling weights are negatively correlated with the number of hired workers, the variable of interest. This situation occurs even within post-strata. For the same California July ALS sample, within post-stratum correlation between the hired worker response and sample weight for respondents who had at least one hired worker ranged from -0.40 to 0.24 with an average value of -0.12. For the unweighted average to be equivalent to the weighted average, weights must be uncorrelated with the variable of interest, or all weights must be equivalent (which would necessarily imply no correlation). The negative correlation of weights and number of hired workers within post-strata suggests that the unweighted average will tend to overestimate the number of hired workers per farm for both frames.

Overall Performance of the Estimators.
Post-Stratified Estimators.
The combinations provided by selecting unweighted or weighted averages and an ability to select for list-only, NOL-only or both respondent types, produced six possible post-stratification estimators to study and evaluate. The NOL-only estimators were used only in conjunction with list-only estimators to provide comparative differences between the two frames on a state level basis. The MF post-stratified estimators were used to evaluate changes in variance due to list-only post-stratification. Numerical evaluations of all estimators considered are given in Appendix C. Only what are considered feasible options will be discussed in this section, and references to the numerical evaluations will be limited. Again a reminder that the post-stratum classification is optimized for California but not necessarily for Florida.

Figure 1a.

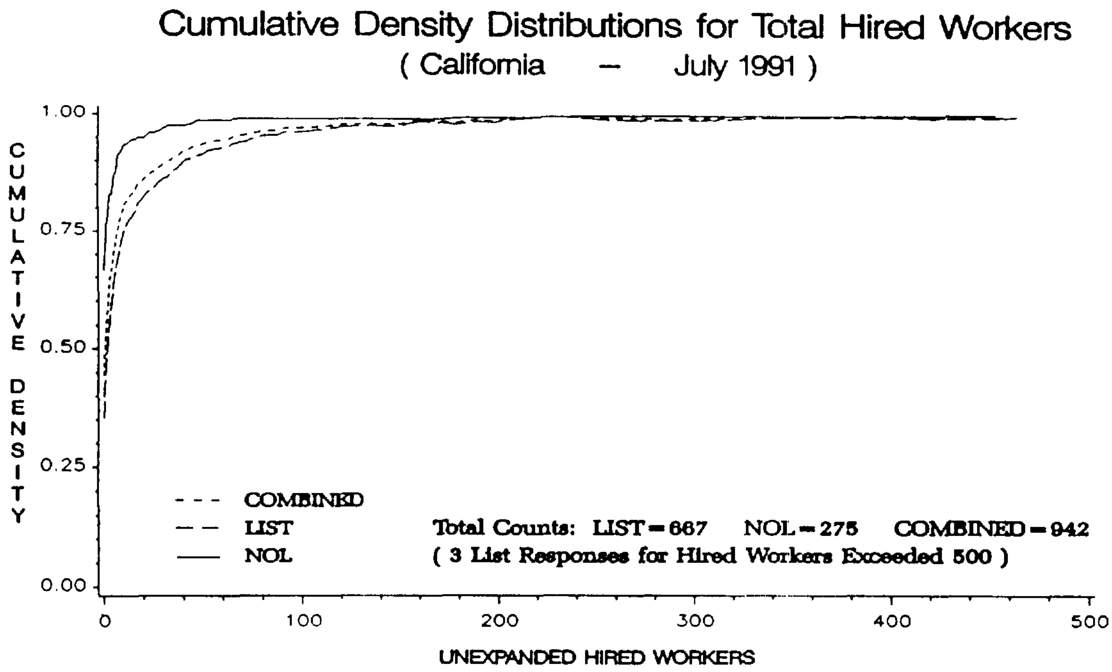
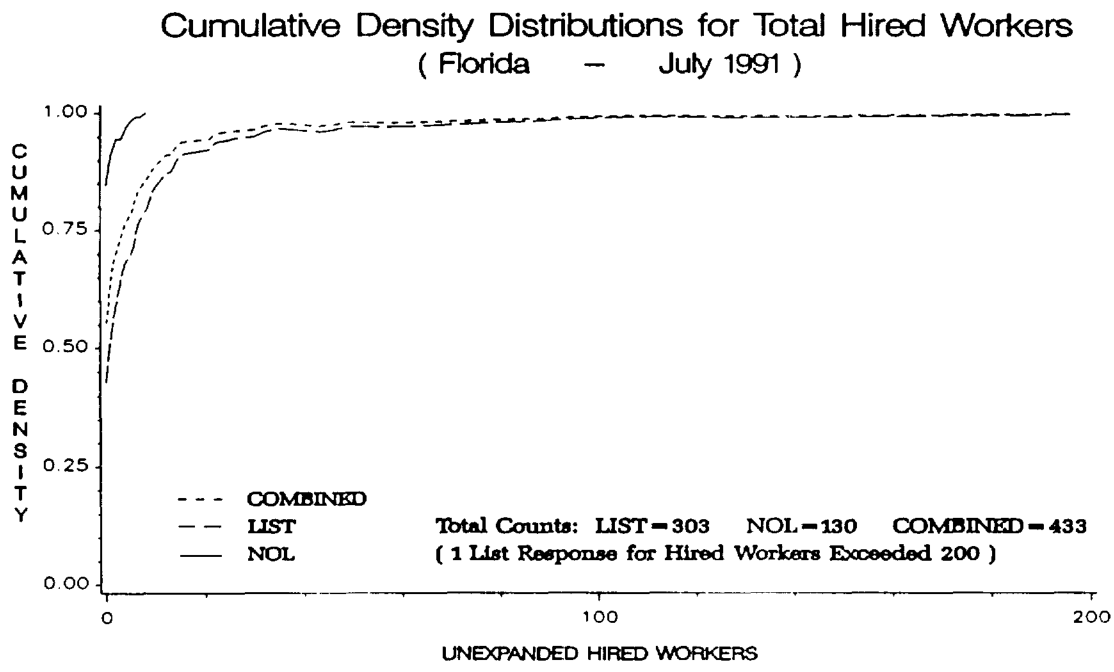


Figure 1b.



Cumulative Density Distributions for List, NOL and Combined. *The NOL sample had a much greater proportion of records with zero workers than did the list. Also positive NOL records tend to be smaller than positive list records. Although these graph are given at the state level, similar differences occurred within post-strata.*

Not surprisingly, it was found that the unweighted estimator consistently overestimated the actual labor force by a large margin (recall Table 1). The estimators using unweighted survey values produced the largest biases of all the estimators. Use of weighted survey values produced adequate, though somewhat more variable, estimates when compared to MF survey design direct expansion (MF DE) estimates. Since much of the post-stratification information is included in the list survey design (FVS and FType) and because the bulk of the ALS estimate comes from the list, it is not surprising the weighted MF post-stratified and the MF DE estimates are comparable.

Figures 2a and 2b for California and Florida respectively, depict the level of bias produced by using a strictly unweighted post-stratified estimator and compare survey estimates across the 1991 ALS series year. For these and all succeeding graphs of this type, the vertical length of each estimate represents one standard error from the survey estimate in either direction. In some extreme cases the length in one or both directions has been truncated.

Also not surprising, given the post-stratum mean differences as shown in Table 1, it was found that the list-only estimator consistently overestimated the actual number of laborers while the NOL-only underestimated the actual labor force number. Figures 3a and 3b for the two states illustrate graphically the problems inherent in the weighted list and NOL-only post-stratified estimators, again comparing survey estimates to the Board as well as to the combined MF estimate across the 1991 ALS year.

Overall, list-only post-stratification CVs for California and Florida were mostly comparable with original MF DE CVs. This occurs for the most part because list-only post-stratified estimates generally are larger than the survey indication and have more variance introduced through the use of estimated June population counts. This leaves the overall percentage error of the total (CV) roughly equal to the MF DE CV. One must remember however, that the computed variance underestimates actual variance by as much as 10% resulting in a CV increase of approximately 5% since the $\sqrt{1.1} = 1.049$. For purposes of this report however, all CVs displayed will be the actual value computed with no compensation for bias. For California, the average CV for the weighted list-only post-stratification estimate for the survey year 1991 averaged 15.6%. This compares to an average MF DE CV for California of 14.2%. For Florida the average CV for the weighted list-only post-stratified estimator was 16.4% compared to 21.1% for survey design over the same period.

Figure 2a.

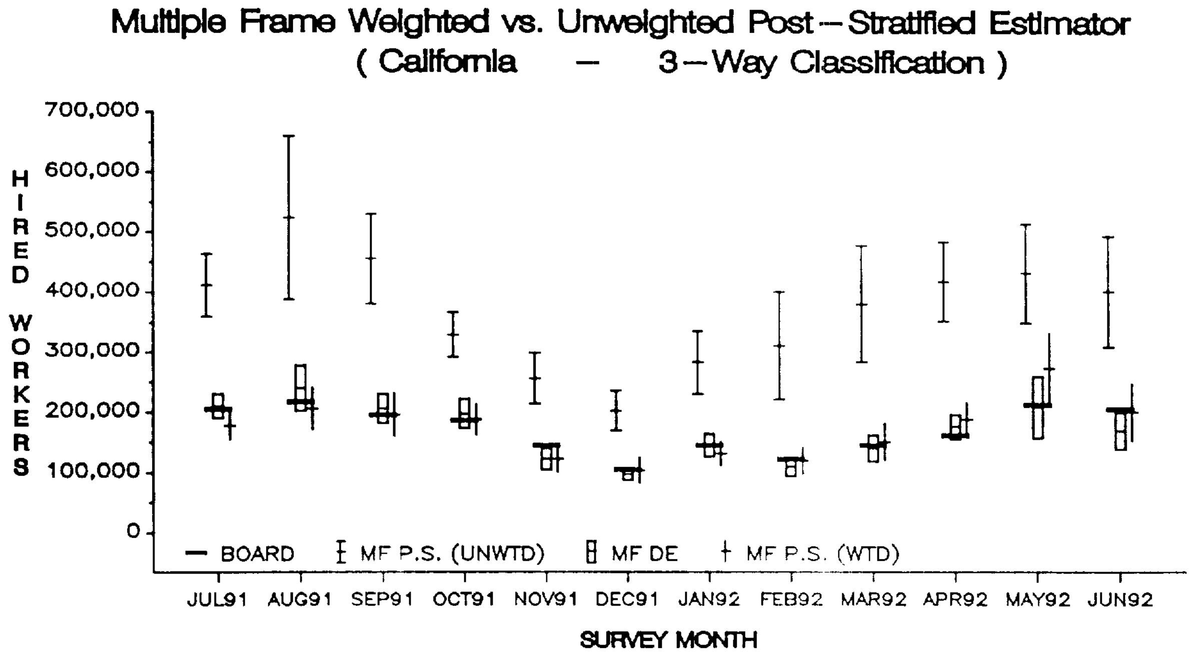
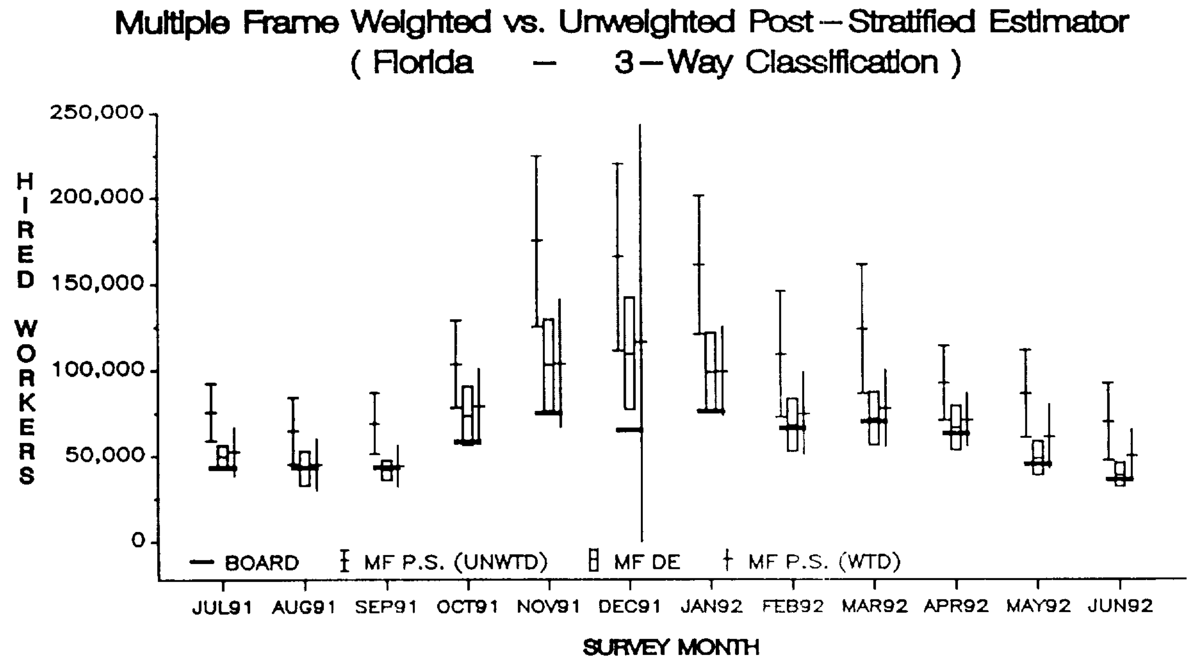


Figure 2b.



(Vertical Symbol Length Represents Two Standard Errors)

Comparison of Weighted versus Unweighted Post-Stratified Estimators. *The unweighted estimator over-estimates total hired workers while the weighted estimator tracks well with the Board estimate for both states.*

Figure 3a.

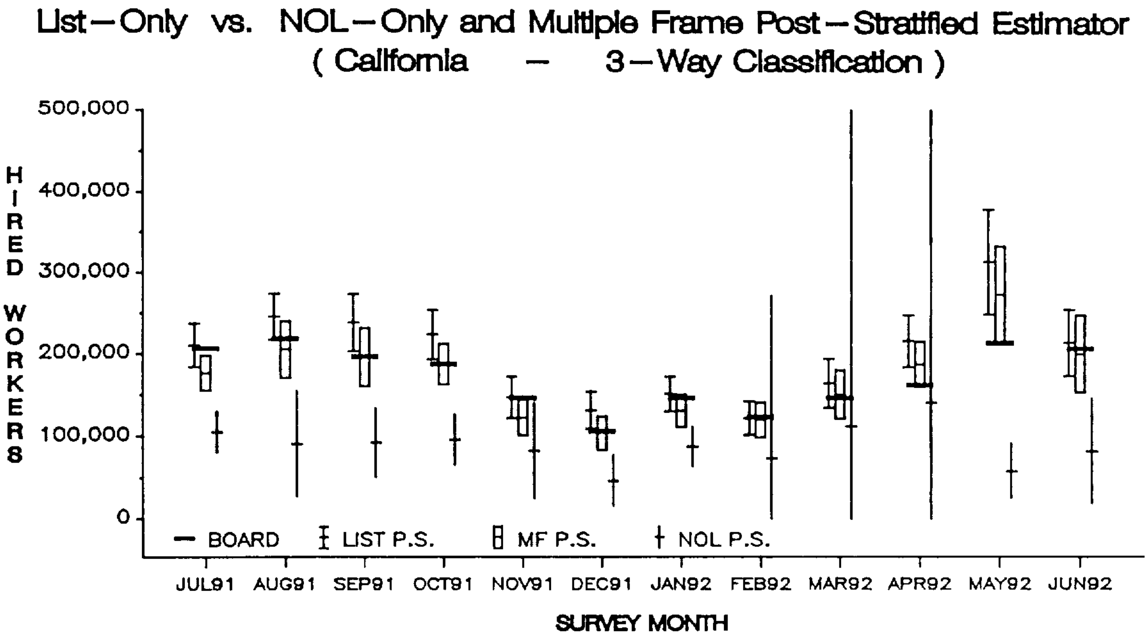
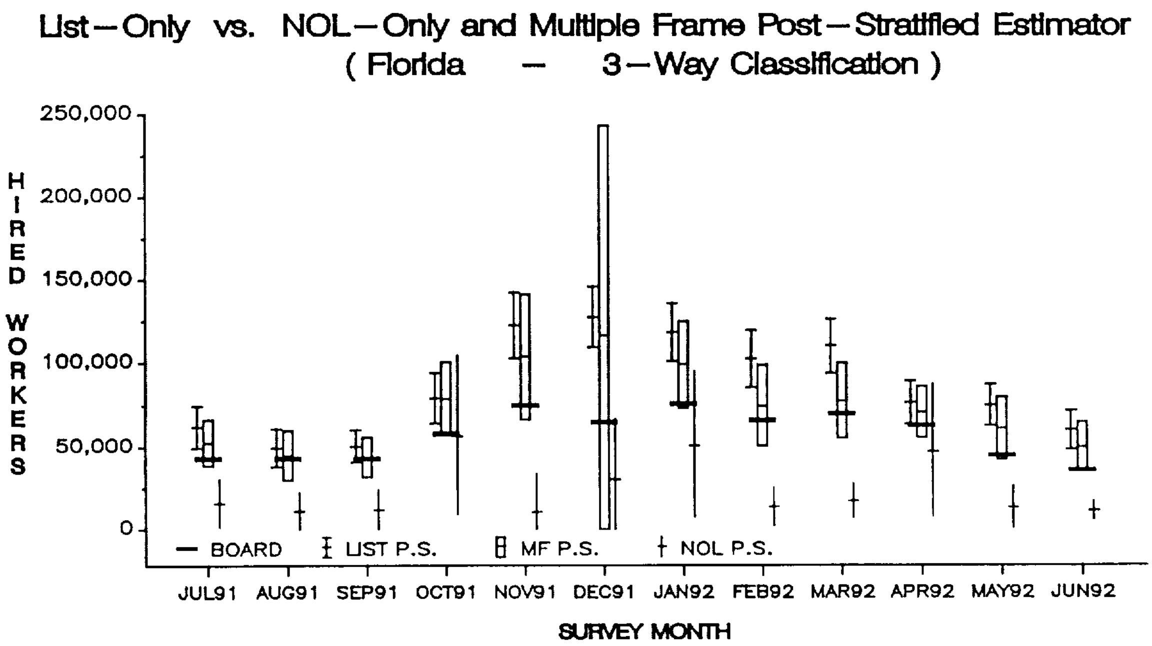


Figure 3b.



(Vertical Symbol Length Represents Two Standard Errors)

Comparison of the List-Only, NOL-Only and Multiple Frame Post-Stratified Estimators. *Use of only list respondents produces an estimate which tends to overestimate the actual number of hired workers; use of only NOL respondents greatly underestimates the number of hired workers; and use of the Multi-frame data tracks the Board fairly well.*

Ratio Estimators

Post-stratified combined ratio expansion estimates were calculated using MF and list-only data. In addition, a combined survey design ratio expansion estimate was computed using list-only data. For the list-only post-stratified ratio estimator a second simplified post-stratification scheme was tried using almost exclusively the Peak post-stratification variable. This was the classification variable which was best correlated with the number of hired workers. Again, the evaluation of these ratio estimators was based on the four numerical measures: BIAS, AVE CV, ME and MAX. A summary of these evaluations can be found in Appendix C. Eleven monthly estimates were produced over the survey year for each estimator since a ratio estimate for July 1991 was not feasible.

Post-Stratified Ratio - Initially, ratio expansion estimates were obtained using Equation (3), a combined post-stratified ratio estimator, where the three-way post-strata classification scheme was utilized. Post-stratified survey total estimates using either list-only or MF respondents were constructed, and the results are shown in Figures 4a and 4b alongside the actual MF DE and the Board number for that month. The weighted list-only post-stratified survey total ratio tracks well with the Board estimate and, in fact, seven of the eleven ratio expansion estimates obtained for California and four of the eleven ratio expansion estimates produced for Florida were closer to the Board estimate than the MF DE indication. The average CV for California was 11.3% for the list-only ratio expansion estimate, which was less than the MF DE average CV of 14.6% over the same eleven surveys. For Florida the average CV for the list-only ratio expansion estimate was 16.4% compared to an eleven survey MF DE average of 21.8% Since these estimates are not biased upwards (at least not to the extent that the non-ratio list estimates were) gains appear to have been made in both accuracy and precision for California, and at least in precision for Florida.

Peak Worker-Only Classification - The three-way classification represents a rather complex scheme for post-stratifying respondents. We investigated whether it could be simplified. An individualized classification scheme for both California and Florida was determined which classified all respondents with zero Peak into one stratum, all respondents with FVS less than \$20,000 into a second post-stratum (FVS less than \$20,000 is the cut-off for list sampling), and the remaining respondents with FVS greater than \$20,000 into three post-strata based on Peak. In reality this is not a Peak-only classification but also uses a two-way classification for FVS. Since control data for the list is not always accurate, some of the list respondents can and do fall into the FVS less than \$20,000 category. In fact, enough list respondents fall in this post-stratum to make NOL modeling possible. Combined post-stratified ratios were produced and results are shown in Figures 5a and 5b along with the combined ratio 3-way post-stratified estimator and the MF DE. It appears that the Peak-only classification scheme tends to have a larger estimate than the Board number and seems to have less precision than that using three-way classification scheme which produces estimates that are comparable with the Board numbers. Thus overall, the three-way classification produces a better estimator, but at the cost of increased complexity for the post-stratification model.

Figure 4a.

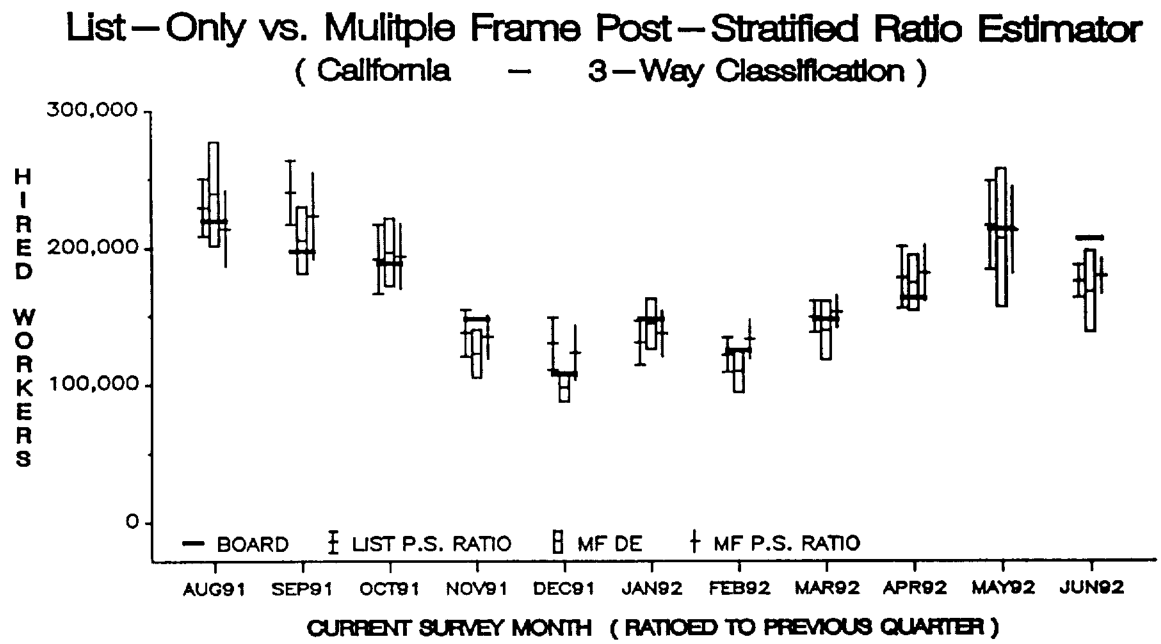
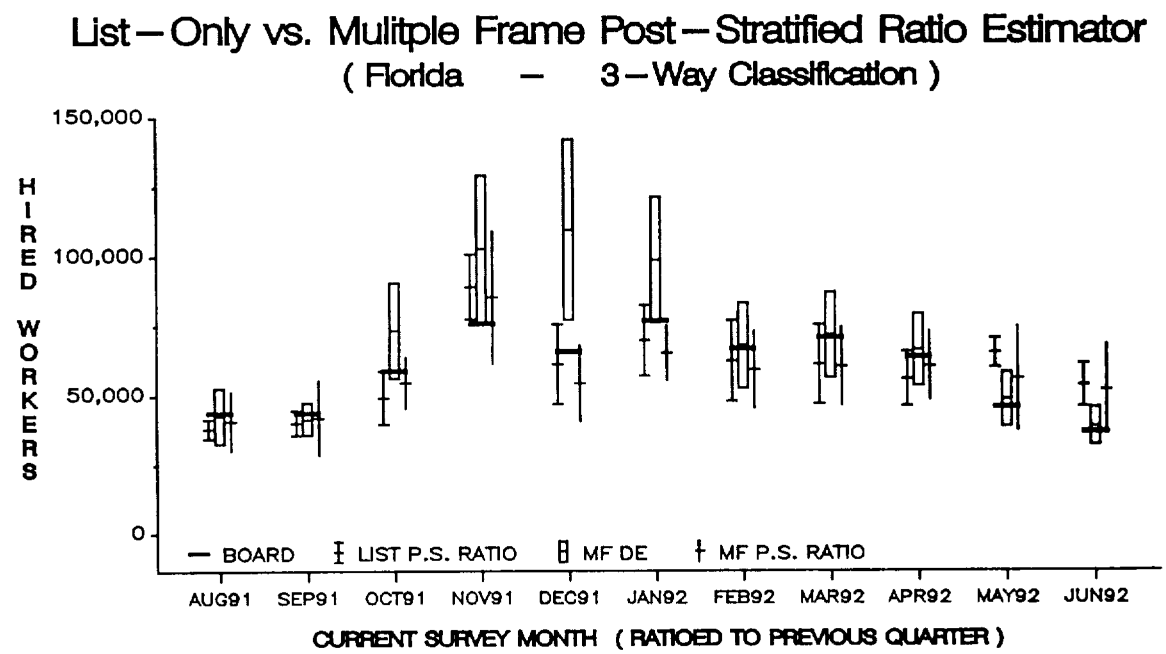


Figure 4b.



(Vertical Symbol Length Represents Two Standard Errors)

Comparison of the List-Only and Multiple Frame Post-Stratified Ratio Estimators. *The list-only ratio expansion estimator tracks well with the Multiple-Frame ratio and the Board number. Variance reductions for the list-only versus the MF DE survey indication can also be seen for several months.*

Figure 5a.

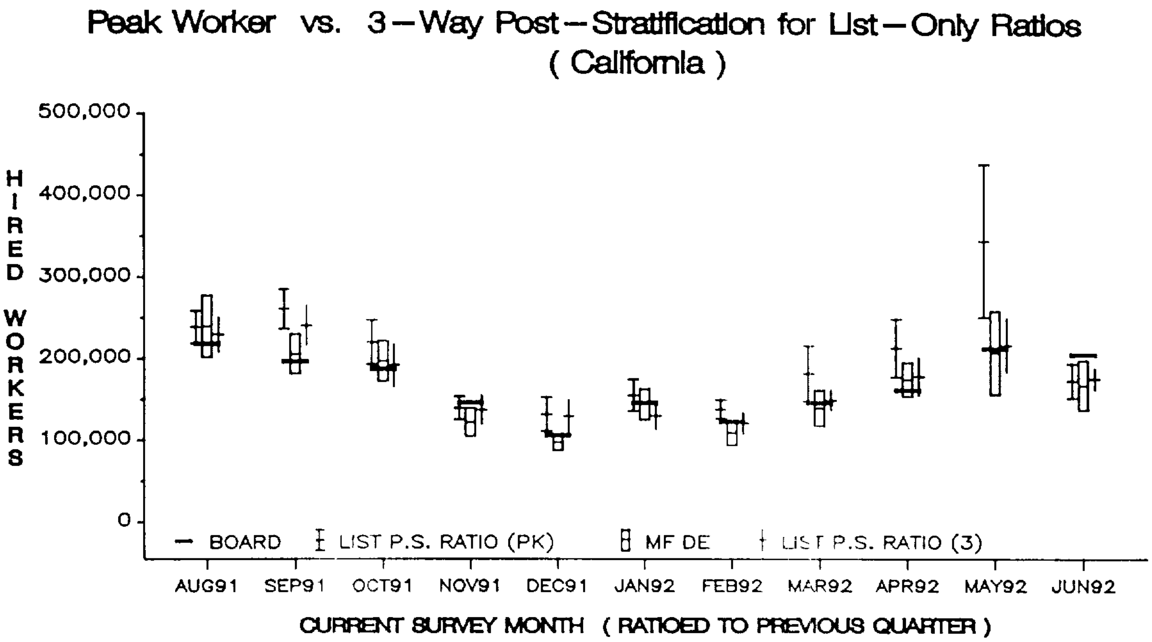
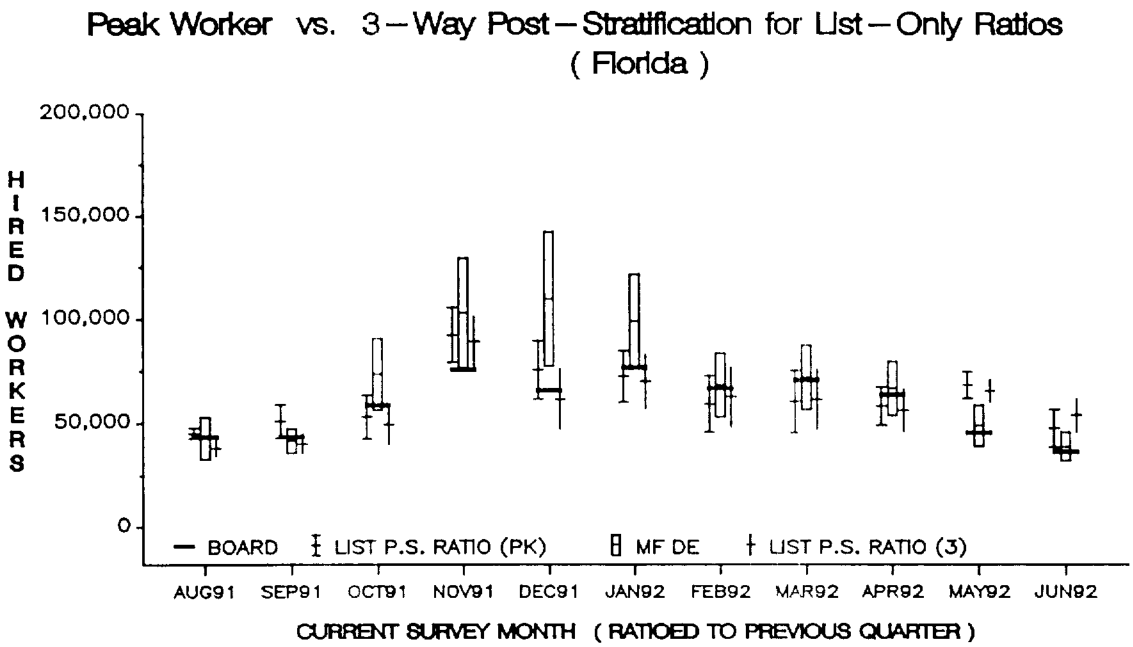


Figure 5b.



(Vertical Symbol Length Represents Two Standard Errors)

Comparison of Peak-Only and 3-Way Post-Stratified List-Only Ratio Estimators. *The 3-Way classification tracks better with the Board number and provides a smaller variance than does the Peak-Only post-classification scheme.*

Survey Design List-Only Ratio - Figures 6a and 6b compare the list-only three-way post-stratified combined ratio expansion with the list-only survey design combined ratio expansion. The post-stratified ratio estimator uses a weighted ratio, accounting for differences in farm numbers across post-strata. The list-only ratio estimator is based on the ratio of all matched useable respondents. It is this difference which makes the post-stratified combined ratio estimator a more accurate estimator than the survey design combined ratio estimator. The two estimators have about the same precision. Again it should be pointed out that the three-way classification is not necessarily optimal for Florida and there is room for further improvement through improved post-stratification.

Turner (1991) evaluated an expanded combined survey design ratio to the previous quarter against a direct expansion estimator. However, Turner's methodology differed in several respects to the one investigated here. Turner's study was conducted to determine the effect of smaller (half-sized) samples during off-quarter months, by respondent type (list or NOL). An expanded ratio estimate of the number of hired workers was obtained for each of eleven seasonal labor states, using matched respondents from half-samples created from the October (current) and July (previous) 1991 labor surveys. These estimates were compared to the eleven half-sample MF DE estimates obtained for October 1991 to evaluate the preferability of either of the two estimators. The ratio estimate was expanded using the full sample MF DE estimate from the previous July quarter. In the study, the half-sample ratio estimator did as well as the half sample direct expansion for either the list or the NOL based on analyses of the eleven states. This is somewhat encouraging since the ratio estimate is a quasi-independent estimator relying only on the rate of change, while the half sample DE estimate is very much related to the current full sample DE estimate. The list-only ratio estimator seemed to perform adequately for California, but was less reliable in Florida and smaller states. Using the previous Board published number instead of the MF DE could add stability to the ratio estimate and make it a more viable option. The use of a separate ratio estimate using all design strata or a lesser number of collapsed strata could also improve the list-only survey design ratio estimate.

Summary of Results.

The combined three-way post-stratified list-only ratio estimator seemed to provide a viable estimate for the total number of hired workers indication. Though the post-stratification model is somewhat complex and would have to be optimized for each state or region, it does fulfill the objective of using a sample which ignores a subgroup, specifically the NOL. An overall comparison of this estimator along with the MF DE, the Board and, for California, the California Administrative Data, can be seen in Figures 7a and 7b.

Figure 6a.

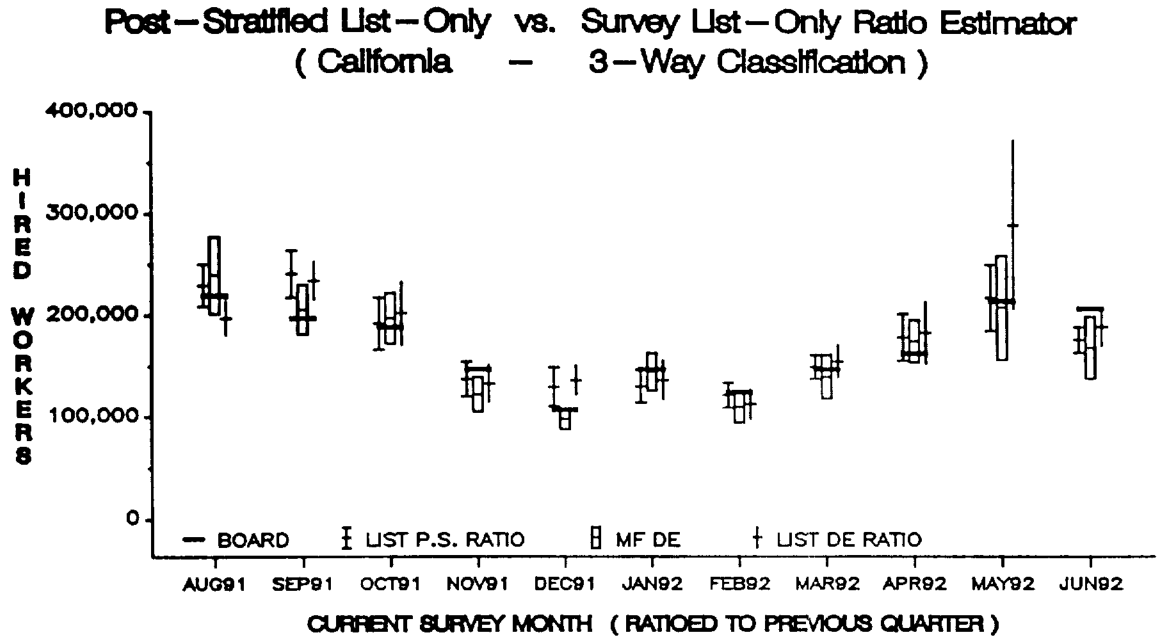
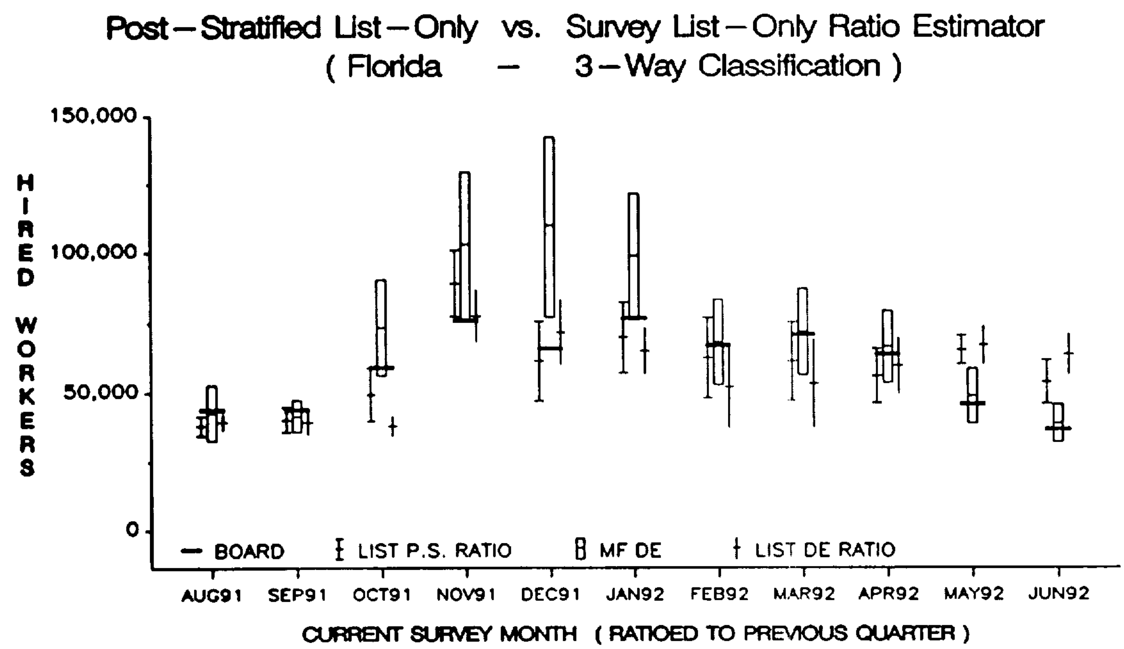


Figure 6b.



(Vertical Symbol Length Represents Two Standard Errors)

Comparison of List-Only Post-Stratified and List-Only Survey Design Ratio Estimators. *The 3-Way post-stratified estimator tracks better with the Board number and generally provides similar if not smaller variances than does the non-post-stratified Survey Design Ratio.*

Figure 7a.

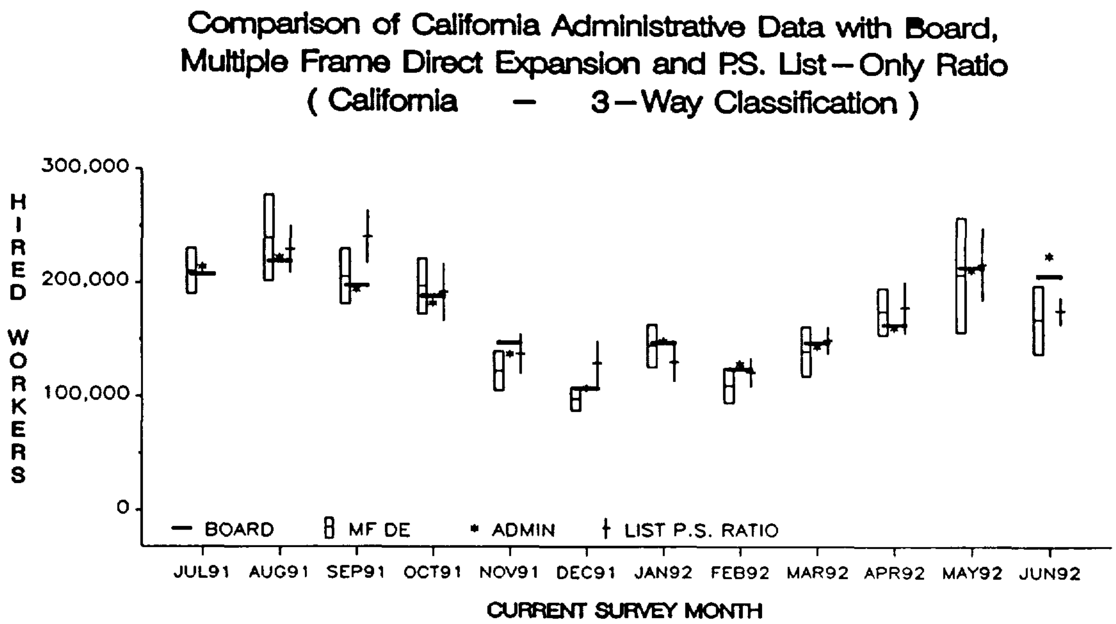
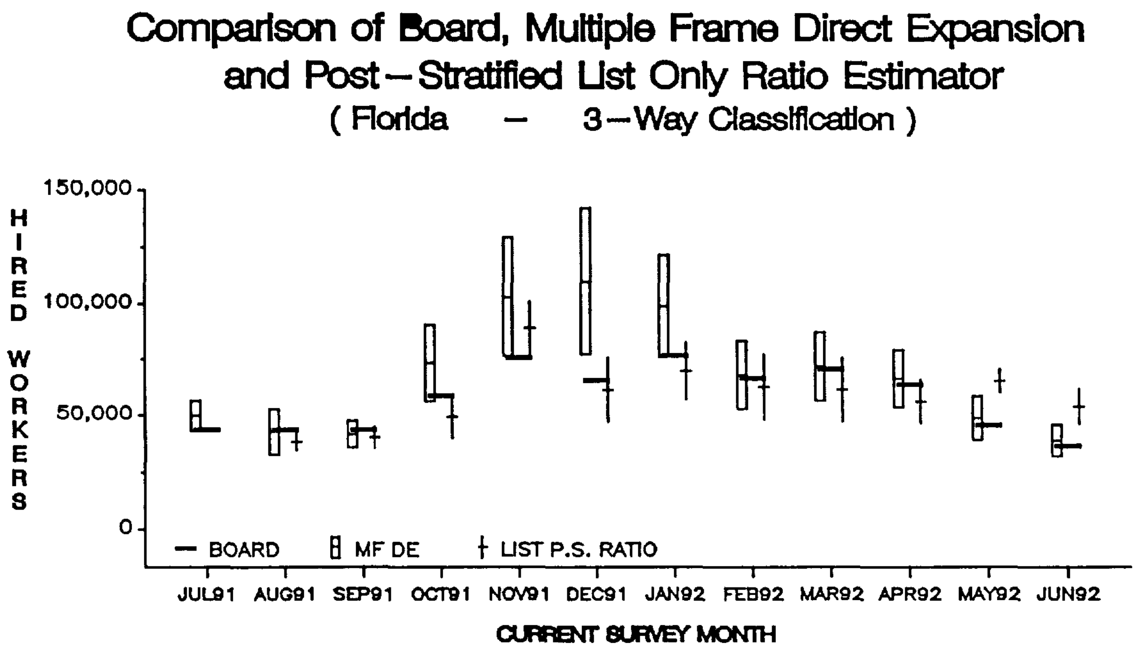


Figure 7b.



(Vertical Symbol Length Represents Two Standard Errors)

Comparison of List-Only Post-Stratified Estimator With Currently Used Indications. *The 3-Way post-stratified estimator tracks well with both the Board number and, for California, the EDD Administrative Data (Admin). All variances are smaller or comparable to the MF DE.*

FUTURE STRATEGIES

Adjusting June Population Counts.

True post-stratum population counts are an integral part of any post-stratification scheme. In the case of farm counts there are outside sources available for the estimate, as well as several estimators used by the Agency. The Bureau of the Census produces a farm count estimate every five years which could be modeled across years. Two estimators are available from the JAS:A, the area weighted and the multiple frame. Beginning in June 1992, five states began collecting the Peak data for JAS list respondents, allowing a MF farm count estimate to be made by post-stratum. It still remains to be seen how much overall variance will be reduced through the use of these post-stratum count estimates.

The Agricultural Statistics Board sets a farm number in June of each year by considering all JAS estimator indications and other inputs, and, so these numbers could be used to set farm population counts. However, any population estimates which are to be used must be partitionable into defined post-strata, either through modeling or through the use of classification variables. The use of population count estimates which have the needed post-classification variables would be preferred, otherwise modeling would be required and this would add more error to the overall estimate. Any population adjustments will primarily aid the usual post-stratified estimators and may not have much influence on the ratio (rate of change) post-stratified estimators.

Alternative Modeling of the NOL to List Respondents.

Since it has been shown that NOL and list respondents, which are identically post-stratified, do not depict similar distributions, some means of relating these differences must be included in any model. Though the ratio expansion estimator seems to minimize this difference, other alternative models are also possible. Two possible approaches being considered or currently under investigation are regression modeling and post-stratification of the NOL only. Regression can be viewed as a generalization of post-stratification. It allows for infinite post-strata for each classification variable without the problem of maintaining adequate counts within post-strata. Post-stratification of the NOL only, while continuing to use the survey design direct expansion from the list sample would maximize post-stratification where it is most effective, and eliminate it where it is least effective. This could improve estimates both in terms of accuracy and precision, though little would be done to reduce respondent burden or variability.

CONCLUSIONS

For the Agricultural Labor Survey, there appear to be differences in mean values of list and NOL respondents within post-strata. Also, the sample design produced some negative correlations between the sample weight and the response within post-strata. These two factors make the unweighted post-stratified estimator biased. Though bias is reduced in the case of the weighted post-stratified estimator, differences between weighted list and NOL respondents still exist within post-strata. Ratio expansion estimators, however, appear to avoid this problem and may have some potential within the NASS framework in addressing the strawman type issues and approaches.

The list-only combined ratio expansion estimator using three-way post-stratification appears to model the NOL adequately, while reducing on average, variances. Thus, both accuracy and precision of this estimator appear comparable, if not improved, when compared to the MF DE. The gains which could be made in reducing respondent burden, survey complexity and estimation variability increase this estimator's chance for success. However, development of post-strata for individual states and regions would be a time consuming job and would involve reworking of the current survey summary system which can become a major task. Also, the Labor Survey is unique in several respects, and it is difficult to draw conclusions beyond this arena. Additionally, an estimator that uses only list respondents will probably be biased and must be cautiously approached and monitored if any list-only estimator were to become operational.

The list-only survey design combined ratio estimator seems to model the NOL fairly well, though not as well as the post-stratified ratio estimator. Perhaps enhancements could be made to this estimator to make it a more viable alternative. Use of this estimator would require only minimal changes to the operational system while allowing the gains from NOL modelling.

One problem with the post-stratified estimators investigated here is the estimated farm counts from the JAS:A. These counts are estimated using the area weighted estimator and tend to be quite variable. The inaccuracies can be corrected to some degree by using the MF population estimate, but any variability in the counts translates to higher overall variances of the post-stratified estimates. The post-stratified ratio estimators reduce the magnitude of this problem, but more accurate population estimates would surely help these estimators also.

The use of a Board number for population counts within post-strata would eliminate the population count variance. Nevertheless, it will have no effect on the error of an estimate using them, since inaccurate Board numbers will lead to biased estimates and the precision of the estimate using them will be over stated.

RECOMMENDATIONS

1. Investigate improvements for the list-only survey design combined ratio estimator to evaluates its feasibility for modelling the NOL without the need for post-stratification. Work could include evaluation of population estimates within design strata, proportioning state (or regional) level Board estimates for the variable of interest within design strata, and possibly collapsing smaller strata and using a separate ratio estimator. Again, bias of any list-only estimator needs to be evaluated with respect to the target population. Any survey design estimator developed should be evaluated against the post-stratified ratio estimator.

2. Compute the list-only survey design combined ratio expansion estimate for agricultural labor at the state level and evaluate its ability to provide viable estimates on alternating quarters in October, January and April.

3. If, after completion of Recommendation 1, post-stratification appears warranted for labor, develop optimum post-strata using the three-way classification for the remaining states and regions. One may have to consider regional estimates for some current state estimates with an allocation scheme back to the state level. Gains made from post-stratification need to be carefully weighed against the time required for developing and maintaining a post-stratification estimator.

4. Evaluate the list-only post-stratification class of ratio expansion estimators on other surveys within NASS, including the current Cattle on Feed series. Evaluate the bias of any list-only estimator with respect to the target population.

5. Investigate other alternatives to the post-stratified estimators such as regression modeling and estimation.

6. Try to improve the overall farm count estimate, or in lieu of that, provide a variable such as total land in farm which can be used to post-stratify the target population. Accurate counts (or other allocation of the target population) are crucial to post-stratification and so any improvements in estimating the population counts would help all post-stratified estimators.

REFERENCES

Flores-Cervantes, Ismael; 1990a; **Simulation Results: A Preliminary Study of the "Strawman" Estimator and Derived Formulas**; November 21, 1991; NASS Internal Correspondence.

Flores-Cervantes, Ismael; 1990b; **Effect of an Incomplete List on the Design Based "Strawman" Estimator**; November 26, 1991; NASS Internal Correspondence.

Flores-Cervantes, Ismael; 1990c; **Estimating the Number of Farms in a Post-Stratum using the June Agricultural Survey**; December 2, 1991; NASS Internal Correspondence.

Kott, Phil; 1990a; **Some Mathematical Comments on Modified "Strawman" Estimators**; NASS Internal Discussion Paper.

Kott, Phil; 1990b; **Comparing "Strawman" and "Weighted Tract" Estimators in June**; NASS Internal Discussion Paper.

Nealon, John Patrick; **Review of the Multiple and Area Frame Estimators**; SF&SRB Staff Report No. 80; Wash., D.C. March 1984.

Perry, Charles, Chhikara, Raj, Deng, Lih-Yuan, Iwig, William and Rumburg, Scot; 1993; **Generalized Post-Stratification Estimators in the Agricultural Labor Survey**; SRB Research Report Number SRB-93-04, Washington, D.C, July 1993.

NASS Program Planning Committee, 1992; **Minutes of the Program Planning Committee of the National Agricultural Statistics Service**; July 14-16, 1992.

Turner, Cheryl L.; **Evaluation of Estimation Options for the Monthly Farm Labor Survey**; SRB Research Report Number SRB-91-12; Washington., D.C., December 1991.

Vogel, Fredric A; 1990a; **"Strawman" Proposal for Multiple Frame Sampling**; October 3, 1990; NASS Internal Memo.

Vogel, Fredric A; 1990b; **"Strawman" Proposal**; November 20, 1990; NASS Internal Memo.

Vogel, Fredric A; 1991; **A New Approach to Multiple Frame Sampling and Estimation**; January 16, 1991; NASS Internal Discussion Paper.

APPENDIX A

Part 1 - An Overview of Post-Stratification Methodology.

Post-stratification permits the use of stratification information obtained during the sampling process to classify respondents into post-strata with known population sizes. By producing a more homogeneous stratification, it provide a way to compensate for under-coverage with respect to these post-stratum groups and thereby reduce the overall variance of the estimates.

Following sampling, all respondents are classified into groups (post-strata) based on information acquired during the survey. Within these post-strata the population units are assumed to be more homogeneous than the target population as a whole. If this occurs, it will produce lower overall variance. Once post-strata are defined and respondents classified, an estimate is produced for the characteristics of interest. These post-stratum estimates, are usually an expansion of averages or proportions based on known post-stratum sizes in terms of a population count of all potential respondents. Estimated totals produced in this manner will be unbiased as long as population size within each post-stratum are known exactly and the post-stratum estimates themselves are unbiased. However, if population sizes are estimated, bias can occur, especially if the expansion sizes (weights) are correlated with the variable of interest. Some forms for a post-stratified estimate of the characteristic of interest Y are:

$$\hat{Y}^{PS} = \sum_{\substack{all\ k \\ post\text{-}strata}} \hat{Y}_k^{PS} = \sum_{\substack{all\ k \\ post\text{-}strata}} \frac{N_k}{\hat{N}_k} \hat{Y}_k = \sum_{\substack{all\ k \\ post\text{-}strata}} N_k \left(\frac{\sum_{i=1}^{n_k} w_i y_i}{\sum_{i=1}^{n_k} w_i} \right) = \sum_{\substack{all\ k \\ post\text{-}strata}} N_k \overline{y_k}\,, \qquad (Eq.4)$$

where

N_k = k^{th} post-stratum known population size,

w_i = i^{th} sample unit weight, and

n_k = k^{th} post-stratum sample size.

All weights (w_i and N_k) must be defined analogously, and nearly always they are defined in the context of members represented in the target population (i.e., the number of members the i^{th} respondent represents is w_i, and the true number of members in post-stratum *k* for the target population is N_k). This however need not always be true. Any weight which can accurately produce a total population value and is known for the unit respondent, can be used. One example would be acreage. Instead of producing a post-stratified estimate for the variable of interest via the target population (i.e., the number of hired workers per farm) one could produce, say, the number of hired workers per acre. One would do this only if it was thought that acreage is correlated with the variable of interest, in this case hired workers, better than the population counts. Note that if y_i is a dichotomous variable (presence or absence, yes or no etc.,) y_k will be a proportion.

In order for any post-stratification scheme to be effective, the following should occur.

1) Subgroups defined by post-stratification should maximize homogeneity of respondents within a post-stratum and minimize homogeneity across post-strata. The better the delineation of respondents into distinct populations within post-strata, the more precise the estimate will be. If no distinct populations exist, then no gain would be expected.
2) Subgroups must be mutually exclusive for respondents and have complete coverage of the target population. This means each respondent must be classified into one and only one post-stratum. In order for a survey response to be useable, it must possess a non-missing value for each post-stratum classification variable. If the value for a specific classification variable is unknown after sampling, it must be imputed before a respondent can be classified into a specific post-stratum.
3) Post-classification responses as obtained during the survey process are often unknown prior to sampling. If such information is known prior to the survey, it is often used in the original survey design to increase sample efficiency. Cases exist, however, where the information is known prior to sampling but the sampling design is unable to make use of the data. Post-stratification can be used in these instances to allow this previously known information to be made use of. The information used to classify respondents into post-strata should be well correlated with the variable of interest. One would like to construct post-strata using variables which create a tight distribution for the variable of interest within each post-stratum.
4) Population counts must be known for each post-stratum. In practice, however, population counts are often not known exactly and must be estimated. This is a disadvantage of post-stratification and often, some inaccuracies in population counts can be expected, introducing bias in the estimate. Additionally, if population size errors occur disproportionately across post-strata and are not compensated for, the estimate will certainly be biased. Population counts are often, though not always, acquired from previous surveys, a census count, or through modeling previous data. Generally, Census Bureau data is used as true population counts when units are defined appropriately. As with any ratio, both the numerator (in this case the population count) and the denominator (sample estimated population count) must be identically defined and represent like units (ie., agricultural operations). Likewise, post-stratum definitions for both population counts and population estimate counts must also be identically defined. Again though, one is not limited to population counts in producing weights for post-stratification. Any number of other variables including total acres, annual sales, proportions by farm type or a host of other possibilities could be employed.

Part 2 - An Overview of Ratio and Ratio Expansion Estimates.

Survey to survey (current to base) ratio estimates can be obtained for any variable(s) of interest provided identically defined variables are included in both surveys. Using only useable matched respondents (sampled units which appear in both surveys and which are both useable) generally produces a better correlation between responses and a less variable estimate overall. A ratio estimate for the rate of change for a characteristic of interest, say population total Y is calculated as:

$$\hat{R} = \frac{\hat{Y}'}{\hat{X}'}, \qquad (Eq. 5)$$

where
$\hat{Y}'$ = current survey estimate for Y,
$\hat{X}'$ = previous survey estimate for Y, and
both $\hat{X}'$ and $\hat{Y}'$ are calculated based only on matched records.

A ratio expansion estimate of the current survey total for the variable of interest can be calculated by taking the product of the rate of change ratio estimate R and an estimate of the survey total for the variable of interest from the base survey X. Its form is:

$$\hat{Y}_{RE} = \hat{X}\hat{R}, \qquad (Eq. 6)$$

where
$\hat{X}$= base survey estimate using all records and
$\hat{R}$= rate of change ratio estimate as defined in Equation 5.

Another ratio expansion estimate for a post-stratified design is the combined ratio of the form:

$$\hat{Y}^{PS}_{RE_1} = \hat{X}^{PS} \cdot \hat{R}^{PS} = \hat{X}^{PS} \cdot \frac{\hat{Y}^{PS'}}{\hat{X}^{PS'}} = \hat{X}^{PS} \cdot \frac{\sum_{\substack{all\ k \\ post\text{-}strata}} \left(\frac{N_k}{\hat{N}'_k} \sum_{i \in m_k} y_i \right)}{\sum_{\substack{all\ k \\ post\text{-}strata}} \left(\frac{N_k}{\hat{N}'_k} \sum_{i \in m_k} x_i \right)}, \qquad (Eq. 7)$$

where
N_k = k^{th} post-stratum population count,
$\hat{N}'_k$ = k^{th} post-stratum estimated population count derived from all matched useable ALS List reporting units,
m_k = set of all matched useable List reporting units in post-stratum k,
y_i = i^{th} matched useable List expanded value from the current survey,
x_i = i^{th} matched useable List expanded value from the previous quarterly survey, and
$\hat{X}^{PS}$ = post-stratification estimator of population total from base survey.

$\hat{X}^{PS}$ is an estimator for the same total as $\hat{X}^{PS\prime}$, but generally is based on additional information. The increased information can come in the form of increased sample size (using all respondents instead of only matched respondents) or in adjustments to the total to account for undercoverage or non-response. Ratio and ratio expansion estimates can be conducted at any level (say within design strata, post-strata or by frame) and summed to produce state, regional or national level estimates where appropriate. However, ratio and likewise ratio expansion estimates are most effective when matched responses are well correlated between surveys and when the use of matched respondents does not reduce the sample size appreciably.

APPENDIX B

Classification Variables and Strategies for Defining Post-Strata.

Three classification variables were selected to be analyzed for their ability to describe distinct populations and estimate the number of hired agricultural workers. For our purposes the possible classification variables were Farm Value of Sales (FVS). Farm Type (FType) and Peak Number of Workers (Peak) expected in the next year. Because of the immensity of analyzing all states, it was decided to concentrate on Florida and California and, for the most part, specifically on the California July ALS.

Basic strategy for optimizing the post-stratum populations involved selecting cutoff values for the three classification variables. These cutoff values produced defining characteristics for each post-stratum. Different cutoff values for a particular classification variable were then compared based on the overall variances of the post-stratified MF estimates computed using the Taylor series linear approximation. Minimum variances suggested optimum cutoff values if adequate sample sizes could be maintained within post-strata.

The number of post-strata was limited to a maximum of 12 in order to meet a goal of at least 20 respondents per post-stratum. This meant that most post-stratum defining categories would be of a binary type - high or low. A cutoff value was selected for an individual classification variable ignoring the remaining two classification variables and all respondents were placed into either of the two (or in some cases three) defined post-strata. Variances were calculated and the cutoff value associated with minimum variance while maintaining adequate sample counts for all post-strata was considered optimum. These cutoff values were then evaluated using pair-wise classification of the variables, and finally three-way classification of the variables. Characteristics of each variable are listed below.

Farm Type.

Farm type (FType) is a categorical variable describing 14 different agricultural farms. A farm may produce several commodities, but the FType variable represents the best description of the predominant commodity produced. Possible FType responses are:

1) Cash Grains	2) Tobacco	3) Cotton	4) Other Field Crops
5) Vegetables, Melons or strawberries	6) Fruits	7) Nuts	8) Nursery or Greenhouse
9) Livestock	10) Poultry	11) Dairy	12) Other Livestock
13) CRP Only	14) Christmas Tree Only		

FType represented the most difficult variable used to define post-strata since the categories are not ordinal. However, proper farm type groupings produced some of the most distinctive populations within post-strata. Three different approaches were selected to be investigated. The first method (F1Type) grouped farm types based on similar commodities. This produced three categories composed of field crops (1, 2, 3, & 4); vegetable, fruit, nut or nursery (5, 6, 7 and 8); and livestock, poultry and dairy (9, 10 & 11). Farm types CRP and Christmas trees only (13 & 14) were placed with the livestock group primarily in an attempt to increase post-strata sample counts.

A second method (F2Type) grouped farm types based on the intensity and seasonality of labor across the survey year. This meant that nursery and greenhouse, poultry and dairy (8, 10 & 11) - types which had a consistent labor force across all months - were placed together. Livestock and field crops (1, 2, 3, 4, 9, & 12) - types which displayed a high spring and fall labor force with minimal winter and summer labor - were combined to form a second group. And lastly, vegetables, fruit and nuts (5, 6, & 7) - intensive short term labor force farm types - were combined. Again, CRP and Christmas trees (13 & 14) were added to the livestock and field crops groups. A third method (RType) was an automated approach which allowed for seasonality. Farm types were selecting based on a computed average ratio of current number of hired workers to peak workers for each farm type sampled in the labor survey. Though this ratio uses the variable of interest (Hired Workers) to post-stratify (a violation of the independence rule between variable of interest and classification variable) it was thought that these ratios could be set and used in the next year's survey for that month where they would be independent of the variable of interest. Once the ratios were calculated for each farm type they were classified as either high or low, forming two FType post-strata. The cutoff between high and low was selected automatically using a clustering algorithm while attempting to maintain adequate farm sample counts within each post-stratum.

The F2Type method of grouping by labor intensity across the survey year appeared to work best, producing the lowest variances both by itself and paired with other classification variables. The F1Type method combined several dissimilar farm types in terms of labor and hence produced a larger variance. The RType method generated no better results than the second and for many farm types the ratio was computed with a small number of farms. The third method may become more attractive as a history of labor survey ratios is developed though computation requirements would also have to be considered.

Peak Number of Workers.

Peak Number of Workers (Peak) is an integer variable which runs from zero to as large as 5,000. Don't Know (DK) was an allowable response. The DK response rate ranged from

5-15%, varying by state and survey month. Two methods of treating DK responses for stratification purposes were investigated. First, they were given their own category, creating three Peak categories overall (DK, less than cutoff, equal or greater than cutoff). The underlying assumption for this method is that most DKs act similarly but unlike the other two Peak post-strata. The second method imputed DKs based on the FVS and FType for the respondent. An average Peak value was computed for each FVS-FType combination sampled. If the average was less than the cutoff the DK respondent was placed in the lower Peak post-stratum, if equal to or larger than the cutoff it was placed in the upper Peak stratum. If no average existed or one of the two defining variables was missing the respondent was placed in the higher Peak stratum. Research into DKs showed that they appeared to be spread uniformly across both FVS and FType. Since imputing for DKs produced a consistently lower variance when respondents were classified strictly on the Peak variable and since little justification existed that DKs acted differently from similar respondents who did know, imputing for DK's was selected as the preferred method. Peak was a required response for the survey and missing values were not a problem. Optimum Peak cutoff was generally 3 or 4. Peak is positively correlated with the third classification variable, Farm Value of Sales. This limited the number of cutoff values possible for one given a cutoff value for the other, if adequate sample counts were to be maintained across all post-strata.

Farm Value of Sales.

Farm Value of Sales (FVS) is an ordinal categorical variable which measures the total farm sales for the year (measured from January through December). Possible FVS responses are:

1) < $1,000	2) $1,000 - 2,500	3) $2,500 - 5,000
4) $5,000 - 10,000	5) $10,000 - 20,000	6) $20,000 - 25,000
7) $25,000 - 40,000	8) $40,000 - 50,000	9) $50,000 - 100,000
10) $100,000 - 250,000	11) $250,000 - 500,000	12) $500,000 +

Farms producing less than $1000 per year are not counted as a farm by definition. However, farms producing less than $1,000 may possibly still be defined as a farm if they possess what amounts to a value of $1,000 or greater through cropland, livestock or other agricultural assets. The Farm Value of Sales response must be answered or imputed at the time of the survey and missing values were not a problem. Optimum cutoffs tended to vary between states with California's set at $50,000 and Florida's at either $40,000 or $50,000.

APPENDIX C

Numerical Evaluations of Estimators.

All estimators were evaluated numerically based on four criteria as follows:

1.) BIAS - A measure of accuracy, which is the expected deviation of an estimator with direction (positive or negative) away from the Board number and is defined as:

$$BIAS = \frac{1}{n_{surveys}} \cdot \sum_{all\,p\,surveys} (Est_p - Board_p).$$

2.) AVERAGE COEFFICIENT OF VARIATION (AVE CV) - A measure of relative precision, which is defined as:

$$\overline{CV} = \left(\sum_{all\,p\,surveys} (SE(Y_p) \,/\, Y_p) \right) / (n_{surveys}).$$

3.) MEAN ERROR (ME) - A measure of both accuracy and precision that accounts for both bias and variation. It is a measure of the expected absolute deviation of an estimator away from the actual Board number for a survey which is defined as:

$$ME = \sum_{all\,p\,surveys} |Est_p - Board_p| \,/\, (n_{surveys}).$$

4.) MAXIMUM ABSOLUTE DEVIATION (MAX) - A measure of the largest expected deviation of an estimate in any direction away from the Board number, which is defined as:

$$MAD = \underset{all\,p\,surveys}{Max} |Est_p - Board_p|.$$

WEIGHTED ESTIMATORS

State	Estimator[1]	Bias	Ave CV	ME	MAX
SURVEY DESIGN ESTIMATORS					
CA	ADMIN(EDD)	1,619	.	7,084	19,056
CA	MF /DE	-7,059	14.20	17,290	41,333
FL	MF /DE	8,860	21.06	16,234	42,775
SURVEY DESIGN RATIO ESTIMATOR					
CA	LIST/DE	6,870	13.29	28,991	72,327
FL	LIST/DE	-3,175	15.03	14,902	25,867
POST-STRATIFIED ESTIMATORS					
CA	MF /PS3	-4,113	17.40	22,887	55,815
FL	MF /PS3	13,165	37.21	17,285	39,009
CA	LIST/PS3	-4,113	15.76	36,026	96,516
FL	LIST/PS3	27,426	17.52	31,922	60,595
CA	NOL /PS3	-86,138	183.32	94,635	158,681
FL	NOL /PS3	-34,701	98.73	38,421	66,590

[1] Explanations for the estimator names follow the table

WEIGHTED ESTIMATORS (Cont'd)

State	Estimator	Bias	Ave CV	ME	MAX
CA	MF /PSK	-28,475	18.13	[illegible]	71,900
FL	MF /PSK	13,990	25.75	[illegible]	61,676
CA	LIST/PSK	13,710	17.27	[illegible]	151,660
FL	LIST/PSK	79,892	24.65	33,135	63,783
CA	NOL /PSK	-88,749	53.14	91,546	146,721
FL	NOL /PSK	-22,691	54.61	[illegible]	60,452
RATIO POST-STRATIFIED ESTIMATORS					
CA	MF /PS3	-280	11.85	14,769	29,638
FL	MF /PS3	-2,763	24.65	9,315	14,239
CA	LIST/PS3	-1,017	11.34	[illegible]	40,555
FL	LIST/PS3	-1,332	16.35	[illegible]	18,386
CA	MF /PSK	7,164	13.79	[illegible]	51,530
FL	MF /PSK	-2,558	21.97	[illegible]	15,335
CA	LIST/PSK	28,067	13.72	49,303	128,518
FL	LIST/PSK	1,893	16.50	10,559	21,270

UNWEIGHTED ESTIMATORS

State	Estimator	Bias	Ave CV	ME	MAX
SURVEY DESIGN RATIO ESTIMATOR					
CA	LIST/DE	17,506	12.13	17,381	42,282
FL	LIST/DE	25,603	17.08	10,534	22,074
POST-STRATIFIED ESTIMATORS					
CA	MF /PS3	191,819	19.24	201,504	302,026
FL	MF /PS3	49,180	28.14	56,133	98,804
CA	LIST/PS3	219,441	18.21	234,448	387,133
FL	LIST/PS3	63,242	16.41	72,067	134,902
CA	NOL /PS3	18,682	117.61	100,647	235,504
FL	NOL /PS3	-31,082	282.75	[illegible]	61,383
CA	MF /PSK	108,713	19.18	[illegible]	356,067
FL	MF /PSK	17,551	25.22	17,551	61,676
CA	LIST/PSK	173,856	20.65	[illegible]	750,194
FL	LIST/PSK	27,999	26.26	27,799	63,783
CA	NOL /PSK	-25,484	70.36	53,886	130,196
FL	NOL /PSK	-22,691	162.16	[illegible]	60,452

Estimator Naming Convention.

For post-stratified and ratio post-stratified estimators the first part of the name defines the type of respondents used:

MF - Multiple Frame (Both List and NOL Respondents)
LIST - Only List Respondents
NOL - Only NOL Respondents.

The second part defines the type of Stratification (or Post-Stratification) scheme employed:

DE - Direct Expansion Survey Design
PS3 - Three-Way Post-Stratification
PSK - Peak Worker-Only Post-Stratification