United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research Division

SRB Research Report
Number SRB-95-04

July 1995

An Evaluation of Four Potential Problems with the Agricultural Survey Substitution Procedures

Mike Fleming

An Evaluation of Four Potential Problems with the Agricultural Survey Substitution Procedures. Mike Fleming, Survey Research Branch, Research Division, National Agricultural Statistics Service, United States Department of Agriculture, Washington, D.C. 20250-2000. July 1995. Research Report No. SRB-95-04.

ABSTRACT

To account for the natural attrition in a sample selected for a NASS Agricultural Survey, provisions are made to substitute out-of-business original operators with new operators who had no other chance of selection. According to the methodology used by NASS for selecting samples from multiple frames, a loss of qualified operators may occur from the time of the base survey reference date to later surveys in the year. Because the events that lead to a substitution are rare and because farms whose operators qualify to be substituted are probably being operated at minimum capacity, missed substitutions cause at most a -0.7 percent relative error in the expansions with a 99 percent coefficient of confidence. Under the current procedures, it is possible to select a substitute more than once. It was found from the research that the error caused by the multiple selection of replacement operators was essentially zero. However, making substitutions unnecessarily may result in a significant positive bias in the expansions. The effect on the expansions due to improperly including substitute operators in the sample and erroneously coding operators as no longer in business is nearly ten times greater than the error caused operationally by missing both full and partial substitutions. These other sources of error overshadow the effect that missed substitutions may have on the Agricultural Survey expansions.

KEY WORDS

Substitution of Respondents; Attrition of Sampled Operators; Survey Quality Assurance.

ACKNOWLEDGMENTS

The author wishes to thank Ron Bosecker and Dale Atkinson for recommending the special use of the SPS indications, Jeff Bailey for his extensive technical assistance and John Amrhein for his advice on collecting the data. The cooperation of the participating SSO's is greatly appreciated; the quality of their data was excellent.

TABLE OF CONTENTS

SUMMARY

The National Agricultural Statistics Service Working Group to Standardize Survey Procedures requested research in February 1993 to investigate certain potential problems with substitution procedures. These problem areas include:

1. Occurrences of missed full substitutions in which sampled operators who were in business at the time of the base reference date of a survey later transfer their entire operations to someone not contained in any frame.

2. Occurrences of missed partial substitutions in which sampled operators who were in business at the time of the base reference date of a survey later transfer a part of their operations, but not all of it, to someone not contained in any frame.

3. Occurrences of multiple chances of selection caused by the formation of new operations after the base reference date from parcels of land acquired from two or more former operators.

4. Occurrences of substitutions made unnecessarily.

The first and fourth cases would only pose a problem if current procedures are not followed correctly; the others are not addressed by the substitution procedures. Even if the procedures are followed perfectly, there is opportunity for negative bias by ignoring new operators that would result from partial substitutions. An offsetting positive bias is possible by failing to account for multiple selections of substitute operators. In spite of sincere intentions, mistakes do occur in following the substitution procedures as they are written. Each case, therefore, describes a potential source of error in the estimates because of changes in farm operations following base survey specification of the list sampling frame.

Based upon this research the following were found:

1. The upper bound on the relative error in the expansions due to missed partial substitutions is 0.35 percent.

2. The upper bound on the relative error in the expansions due to the combination of both missed full and missed partial substitutions is 0.7 percent.

3. The relative error attributable to multiple contacts is essentially zero.

4. An over substitution may affect the expansions considerably, while in general the errors discovered for missed substitutions were relatively minor.

Research also found evidence of the problems survey statisticians encounter when submitting records involving substitutions. Some records were, with permission from Headquarters, incorrectly coded in order to pass them successfully through the SPS edit.

INTRODUCTION

It is an established practice in the National Agricultural Statistics Service (NASS) to substitute one operator with another in certain special cases. The use of substitution is an attempt to maximize coverage of the farming population. Without complete coverage, a sample may not represent the entire population. As a result, an estimate of the total quantity of an agricultural commodity could be understated. In the continuing effort to eliminate the bias which could be introduced into the estimates if coverage was not maximized, survey statisticians throughout NASS follow a set of rules known as the substitution procedures.

The current substitution procedures, however, are not comprehensive in order to avoid further complications in survey procedures. The extent of bias which these shortcomings might be causing in an estimate has not been previously measured. It is on this account that this research on substitution was brought about. The research does not dispute the original purpose of substitution to increase coverage as described in Bosecker [3], nor does it assess the costs and benefits of using any alternate set of procedures. Instead, it answers the questions posed by the Working Group to Standardize Survey Procedures [4] concerning the effects on the expansions of missed substitutions, over substitutions, and multiple chances of selection through the substitution procedures.

Because the substitution procedures do not address cases where a new operator acquires only a part of some operation, the existence of missed partial substitutions was suggested as a possible reason for the observed downward trend over time in some estimates. For the Quarterly Agricultural Surveys, the decline in the expansions extends across quarters from June to March. The decline in the survey expansion for number of farms, for instance, could be explained by the attrition in the sample of operators who had gone out of business.

In practice, interviews accomplish two major objectives. They are used to collect data and they are used to screen the sample to identify bona fide operators. These operators are people who were making the day to day decisions on June 1 for land that had a potential for producing an agricultural product. June 1 is used in NASS as the base reference date for the Agricultural Surveys in determining the admissibility of an operator's data to the process of making estimates. Only those operators who were in business on June 1 qualify as bona fide operators for the June Agricultural Survey. Though the actual operating status of an operator is usually not known when a sample is drawn, the correct operating status is learned during the interviews. Consequently, at the completion of the June Agricultural Survey, the sample consists of bona fide operators; some of them having been selected from the list frame which is a list of names, addresses, and other data and some from the area frame.

The area frame completely covers the land mass of the United States and by so doing it covers the entire population of farm operators. This property of complete coverage by the area frame makes it indispensable when drawing a sample for a survey from which estimates of the total quantity of an agricultural commodity are made. Because the area frame is utilized in selecting a sample of operators in June, the June Agricultural Survey is called the base survey, and the reference date of the June

Survey, June 1, is called the base reference date. Once the sample has been successfully screened by the interviewing process for bona fide operators, the resulting set of sampled operators provides data for the estimation of various statistics. At this point in the survey process, the issue of substitution arises. From then on, the sample may be depleted by the disappearance of operators who divest their operations to others who were not elements of either frame on June 1.

Since no sample is permanent, then from the time of the base survey in June to the commencement of the interviews in September or after, one would expect that some of the operators who were in business on June 1 will cease operating. Not only must an operator claim to be no longer farming, but he must be removed from agricultural activity on land he controls in order to change his operating status to no longer in business. In turn, if a person acquires at least a part of that operation and was not in business anywhere else in the State on June 1, then the frames, in effect, would no longer cover the entire farming population, since the new operator will not have been in either frame. Unless a mechanism for substituting new operators for the old ones is implemented, information about that newly acquired operation will be lost and a negative bias from the resulting attrition in the sample will appear in the expansions.

Because the frames are not revised after the base reference date for the purposes of sampling, the attrition in the sample would otherwise go unchecked if there were no mechanism to replenish the sample with substitutes. In the 1993 Quarterly Agricultural Surveys, the proportion of the total usable reports represented by substitute operators from June to September was $\frac{72}{67004}=0.1\%$. With an application of Bayes Formula, the June to December proportion was found to be $\left(1-\frac{72}{67004}\right)\frac{115}{68848}=0.2\%$. The order of magnitude shows substitutes are indeed a very small portion of total reports. If it were not for the presence of extreme operators, large expansion factors, variations in the data, and unknown rates at which survey statisticians commit errors in a multitude of scenarios, the issue of missed substitutions could be answered at this time. Nonetheless, if one could estimate an upper bound on the relative error in the expansions by using worst cases, and if that upper bound were subsequently found to be very small with a high coefficient of confidence, then one could give a definite answer to the problem of missed substitutions.

Sometimes during a survey, a survey statistician must make a decision when an operator reports that he no longer operates a farm. If an operator divests his entire operation after June 1, the appropriate survey questionnaire is coded as no longer in business and usually no further action is required by the survey statistician. However, if the person who acquired that operation had no chance of being sampled because he was not contained in any frame, then the data about the operation can only be obtained by substituting for the old operator with the new operator.

According to the methodology that NASS uses in selecting samples from multiple frames, every farming operation is assumed to have a chance of being selected. By using an area frame in conjunction with a list

frame, every operator is eligible for selection in at least one frame at the time of the base reference date of a survey. For example, when sampling for the June Agricultural Survey, if an operator is not eligible for list frame selection, the area frame provides a way by which he may be selected. In spite of the imperfect coverage of the list frame, all bona fide operators stand a chance of being selected for the June Survey. NASS uses June 1 to mark the reference date of the area frame enumeration. Thereafter, the contents of the frames remain unrevised for sampling purposes until the next base reference date.

The set of operators who were selected from the area frame for the June Survey and are not eligible for list frame selection is used as another, but disjoint, list frame for selecting samples for subsequent quarterly surveys. From these two frames, samples are drawn for the September, December, and March Agricultural Surveys. No complete area frame enumeration is conducted for these quarters, so that if an operator whose name does not appear on the list frame starts farming after June 1, then it would be impossible for this operator to be selected for the other three quarterly surveys.

In order to mitigate the attrition of sampled operators from the time of the base reference date to the time of the interviews, rules are followed to substitute for the previous operator with the new operator who had taken over a sampled operation after June 1 and who did not have a chance of selection in either sampling frame. The set of rules for making that determination can be found in Section 6 of the Supervising and Editing Manual (S & E manual) [6] for the Agricultural Surveys. To make the rules easier to follow, only those operators who divest their entire operations are addressed by the substitution procedures. This simplification creates two potential problems. Cases involving operators who divest only a part of their operation and provisions guarding against multiple selection of operators are not addressed.

The Working Group to Standardize Survey Procedures asked whether the substitution procedures should also try to account for operators who divest only a portion of their operation. The issues of substitution were studied by Ned Jones [5] in 1988. He suggested that the conditions for making substitutions may rarely exist and that missing either full or partial substitutions would not be an important loss of information. Several years later, the downward trend in the expansions across quarters was shown by Jeff Bailey [1,2] in 1994 to be primarily caused by the presence of non-operators counted in the base survey and by erroneous coding of records indicating that bona fide operators are no longer in business at the time of the current survey.

With the current substitution procedures, a new operator may be selected more than once. When a new operator who became a bona fide operator after June 1 acquires several parcels of land, each from separate operators who had completely divested their operations, the probability exists that the new operator may be contacted through the substitution process for each of those acquisitions. With each acquired farm whose previous operator had gone out of business, a substitute record would have been made; thus the new operator could report for his consolidated operation with each corresponding substitution. In effect, the new operator would be reporting the same

information too often and would introduce a positive bias into the expansions each excessive time.

Another concern associated with the problem of missed substitutions involves the opposite problem; that is, the problem of making substitutions unnecessarily. Over substitutions were discovered accidentally when a few operators were contacted again for the research on missed substitutions. Measuring the extent to which over substitutions affect the expansions added yet another dimension to the research.

In accordance with the request for research from the Working Group to Standardize Survey Procedures, research on substitution was conducted in four categories:

1. The effects on the survey expansions when a sampled operator who was in business at the time of the base reference date transfers his entire operation to somebody else who cannot be selected for the survey.

2. The effects of missing data when a sampled name turns over part but not all of the operation to someone else.

3. The effects on the expansions of multiple chances of selection.

4. The effects caused by substituting for an operator erroneously.

The following examples illustrate common situations governed by the substitution procedures. They show the reliance on two important concepts used for determining the correct disposition of an operator's record. In each example, the survey statistician must correctly identify the bona fide operator of the land and he must determine whether that operator could have been selected for the survey. Deciding whether or not to make a substitution may not be easy as the knowledge of the circumstances may be insufficient or they may be clouded with misunderstanding. Though the following simple examples are contrived and use fictitious data, they portray what might happen if a survey statistician had to address every aspect of the substitution problem.

Example 1. Joe Smith sold his entire operation to William Jones on August 10. He now lives on a one acre lot and is no longer farming. For the September Agricultural Survey, Joe was contacted but informed the telephone enumerator that he had already retired and sold his farm to William Jones. The name of William Jones does not appear on the list frame. When he was contacted for an interview, it was learned from William Jones that he had not operated any land on June 1. The survey statistician coded the William Jones questionnaire as the substituted record and coded Joe Smith's questionnaire to indicate that a substitution had been made. A substitution was correctly made.

Example 2. Roy Adams rented land from David Banwell before June 1. In October, Roy returned the farm to the owner and quit farming. When Roy was contacted in December for the Agricultural Survey, he said that he was no longer the operator. The survey statistician entered the reporting unit code, 921=9 (out of business) on Adams' questionnaire and did not make a substitution because David Banwell qualified for list frame selection. Coincidentally, David Banwell was contacted in June and was

coded a *known zero*; hence, he was not interviewed in any subsequent quarter. In February, unknown to the survey statistician, David Banwell, the landlord, rented the farm to another person. This new operator had never operated a farm before then and the knowledge of that transfer would have led the survey statistician to make a substitution only if David Banwell had been interviewed; consequently, a substitution was missed.

Example 3. During the June Agricultural Survey, Ed Barnum reported that he operated 100 acres. In July, Ed sold five acres to his son who never before farmed. In September, Ed reported that he operated 95 acres; however, since the substitution procedures do not address partial substitutions, no attempt was made to contact Ed's son. A partial substitution was missed.

Example 4. A wealthy financier, Maximilian Doe, decided to take an early retirement to pursue his interest in breeding thoroughbred horses. He found a suitable location to launch his new vocation and bought five separate farms in October. By chance, each of the five former farm operators were contacted in December for the Agricultural Survey and reported that they were no longer farming. The name, Maximilian Doe, did not appear on the list frame and in each of the five interviews he indicated that he never farmed before June 1. In this instance of multiple selection, Mr. Doe would have reported for the same farm five times.

Being selected for a sample is of course a random event that in turn makes the positive bias caused by multiple chances of selection a random variable. In compliance with the method adopted for doing the research on missed substitution by which an upper bound is determined from studying the worst cases, it was assumed that an operator would be contacted the maximum number of possible times. Although that event may be unlikely to happen, the assumption provides a way to find the largest possible bias in the expansions from multiple chances of selection. The maximum bias under that assumption can be measured by counting the number of parcels of land that an operator acquires from different people regardless of their operating status to ensure the worst case scenario.

Example 5. Mrs. Amos Price responded in her husband's behalf for the September Agricultural Survey. She reported that her husband operated 100 acres. In December, Amos Price, himself, responded and said that he didn't operate the land, but Joe Ramos who lives in Chicago operates it. The survey statistician could not find Joe's name in the list frame and therefore he made a substitution. Even though the survey statistician could not find the name Joe Ramos in the list frame, Joe nonetheless operated a hobby farm somewhere else in the same State on June 1. Because a subtract entry was made in the Amos Price record for the substitution and because Joe Ramos was a bona fide operator on June 1, an over substitution was deemed to have been made.

DATA COLLECTION

Sufficient information to measure the effects of missed substitution already lies in the edited Quarterly Agricultural Survey data, if the rate at which errors are made in following the substitution procedures can be determined. This was later estimated by contacting a sample of operators. But, as will be seen in the Method section of this

report, if we look at worst case scenarios, the information contained in the edited survey data is sufficient to set an accurate upper bound on the effects.

Before a substitution can be made, certain criteria stated in the editing manual must be satisfied. If the substitution procedures are followed correctly then specific codes will be used in item code boxes 923 and 941 on the questionnaire. There are nine possible combinations of codes that relate to substitution. An error can be committed in any of them, but the ones that are checked in Table 1 correspond to the situations which were examined in the research on missed substitutions.

Let the probability of committing an error in coding those boxes for any operator be denoted by ϵ. When a mistake of omission is made in following the substitution procedures, a negative bias is introduced into the expansions because the sample loses an observation. The bias occurs in the present quarter and carries forward into all subsequent quarters in the survey cycle. Given enough of these mistakes, a noticeable downward trend from June to March will develop.

Suppose $\epsilon=0$, then one could say that the survey statisticians did a perfect job in coding for substitution. A value of $\epsilon=.50$ would correspond to a coin flip decision in making substitutions. To arrive at a precise estimate of ϵ, operators had to be interviewed to learn their correct operating status and to obtain very accurate statistics on the amount of land they operated in June, September, and December. By contacting operators, it was found that $\epsilon=1/9$ for the December Quarterly Survey.

Estimating ϵ was the most important reason for contacting previously interviewed operators. Operators in Kansas, Kentucky, and South Carolina who divested at least a part of their operation between June and December 1993 were contacted again in

Table 1.
Combination of Action Codes for Change of Operator and Substitution

		941=0	Substitution 941=1	No Substitution 941=2
No Change in Operator	923=0		✓	
Change in Operator	923=1	✓	✓	✓
Mistake in Operator's Name	923=2			✓

January 1994 to determine if they actually were in business in June, September, or December 1993 and to verify whether or not a substitution should have been made. Kansas, Kentucky, and South Carolina were chosen for the research study because collectively they provided a large enough population of operators who went out of business and who had been interviewed in at least two quarters. These States had participated in an acreage reconciliation study [2], and they are located in dissimilar geographic regions of the United States.

At least 225 operators had to be contacted in order to verify the validity of the predicted upper bound on the relative error with a precision of ± 5 percent,. The Survey Quality Research Section selected a sample of 279 operators of which 104 operators had gone out of business and 175 operators were still in business but had divested a part of their land since June 1.

Table 2 contains the size of the sample and the number of successful contacts for each State. A survey statistician in each of the three States coordinated the collection of the data and edited the specially designed questionnaires according to specific procedures that are stated in Appendix II. (All names and identifying information used in the examples are fictitious). In addition, a copy of the project proposal shown in Appendix III was given to each State Statistical Office (SSO) as background for doing the interviews. Each State received the same set of questions for use by the SSO telephone enumerators. They interviewed operators by telephone and recorded the results of the interviews on the furnished questionnaires. The completed questionnaires were returned to the Research Division for analysis.

A total of 243 operators was successfully contacted by the SSO telephone enumerators. The numbers of missed substitutions and over substitutions are shown in Table 3. The estimated probability of missing a substitution is the quotient of the number of missed substitutions divided by the number of operators requiring substitution, as verified from contacting the operators. The estimated probability of making an over substitution is the quotient of the number of

Table 2.
Number of Operators Selected to be Interviewed in January 1994

State	Number of Full Divestitures Sampled	Number of Partial Divestitures Sampled	Number of Successful Contacts in Full Divestiture Sample	Number of Successful Contacts in Partial Divestiture Sample
Kansas	43	59	36	58
Kentucky	29	48	26	47
South Carolina	32	68	14	62
Total	104	175	76	167

Table 3.
Number of Cases Involving Substitution in Kansas, Kentucky, and South Carolina from the September and December 1993 Agricultural Surveys

Month	Number of Reported Substitutions	Number of Required Substitutions	Number of Over Substitutions	Number of Missed Full Substitutions
September	10	10	1	1
December	11	9	3	1

over substitutions divided by the number of reported substitutions. Three over substitutions and one missed substitution were discovered in the December 1993 Agricultural Survey data.

Also in the December 1993 data, six valid substitutions were purposely coded incorrectly at the survey's clean date, when all records must be submitted, in order to overcome a problem with the Survey Processing System (SPS) edit. These records had been correctly coded for substitutions in a previous quarter; however, because of the problems with passing them through the SPS edit, their original coding was changed with no adverse effect on the expansions. Except for one case, all of these records corresponded to substitutes who were wives, sons, or estates of deceased operators. Four new operators could have qualified for providing substitute records, if the operational substitution procedures had contained provisions for partial substitutions.

The estimates of the probability of missing a full substitution ϵ and of making an over substitution, as shown in Table 4, seem to be constant across quarters; moreover, the likelihood of over substituting appears to be about the same as missing a substitution. However, it will be noted later that the effects caused by the observed over substitutions on the expansions were significantly greater than those caused by missed substitutions by a factor of ten. A table containing the detailed data of those individual records involved with substitution coding for December 1993 can be found in Appendix I. The number of records by list and NOL[1] for September and December 1993 are contained in Table 5.

The data from the interviews indicated that an average of 1.085 parcels of land were acquired per operator whenever an acquisition was made. Each parcel is a connected unit of land. If a substitute is selected more than once, a positive error is introduced into the expansions each excessive time. The number of times that a substitute can be selected is at most the number of parcels of land that he has acquired. The size of a parcel, therefore, is immaterial. The number of parcels coming from different

[1]Let $\mathcal{L}$ be the set of operators found in the list frame and let $\mathcal{A}$ be the set of operators selected from the area frame as of the survey's base reference date, June 1, then NOL$=\mathcal{A}-\mathcal{L}$.

Table 4.
Estimated Probabilities in Committing Errors in Following the Substitution Procedures
Based on the Interviews in Kansas, Kentucky, and South Carolina

Month	Estimated Probability of Missing a Full Substitution	Estimated Probability of Making an Over Substitution
September	1/10	1/10
December	1/9	3/11

landowners determines the extent to which multiple chances of selection affect the expansions. If new operators acquire an average of 1.085 parcels, then they could over report through multiple chances of selection by about 8.5 percent. Suppose that with every required substitution multiple selection of the new operators occurs to the maximum possible extent. Then the maximum relative error which multiple chances of selection could cause, assuming that the predicted 0.7 percent is a good upper bound, would be approximately 0.06 percent.

METHOD

The relative error that would be introduced into an estimate by missing a full substitution was first studied by Ned Jones. He used an empirical method to measure the size of the bias in the estimates from the Quarterly Agricultural Surveys. According to his approach, the relative error caused by missed full substitutions is -0.5 percent. A more theoretical method was developed for this research on missed substitutions. It provided a way to predict a bound on the relative error in the expansions due to missed substitutions

Table 5.
Number of List and NOL Operators Involved with Substitution
Among the Sampled Operators in Kansas, Kentucky, and South Carolina

Kind of Substitution	List	NOL
Valid Substitution	5	3
Over Substitution	2	1
Missed Full Substitution	0	1
"Missed" Partial Substitution	3	1
No Substitution Required	177	50

using only edited survey data. The bound on the absolute value of the relative error was predicted to be 0.7 percent with a coefficient of confidence of 99 percent. By August 1993, estimates of the effects of missed full and partial substitutions had been made from the 1992 Quarterly Agricultural Survey data and the results from both methods were found to be consistent.

Neither method could provide definitive answers to the questions posed by the Working Group without additional information. To measure the size of the effects on the expansions due to missed substitutions with a high degree of accuracy requires the knowledge of the rate at which errors in following the substitution procedures are committed.

When a missed substitution occurs, information is lost and cannot be used in the expansion. Let $X_j=0$ if substitution is not done correctly with probability $J\epsilon$ where J is the probability that a sampled operator warrants substitution and ϵ is the probability that a substitution is missed. Let $X_j=1$ if substitution is done correctly for operator j with probability $1-J\epsilon$. .

By means of the indicator variable X_j, corresponding terms in the estimator $\hat{t}$ for measuring the total quantity of an agricultural commodity disappear when a missed substitution occurs. Explicitly, the estimator can be written as:

$$\hat{t} = \sum_{\substack{h \in strata \\ list}} \frac{N_h}{u_h} \sum_{\substack{j \in U_h \\ list}} y_j X_j + \sum_{\substack{h \in strata \\ area}} \sum_{k \in substrata} e_{h_k} \sum_{\substack{j \in U_{h_k} \\ area}} w_j y_j X_j \qquad (1)$$

Except for the indicator variable, equation 1 is the basic estimator used by NASS in computing the expansions for the commodities such as the amount of land in all farms, total number of hogs, and the total quantity of corn stocks found in a Quarterly Agricultural Survey. Variants of (1) are also used by NASS but, for the purposes of the research on missed substitutions, (1) is sufficient for deriving an upper bound on the relative error due to missed substitutions.

The estimator given by (1) consists of two parts. The first part is associated with the sample of operators selected from the list frame; the second part is associated with the sample of operators taken from the area frame which had no chance of list frame selection. By construction, the set of sampled operators of the list part and the set of sampled operators of the area part of (1) are disjoint. The quantity y_j is a reported (but modified) value for commodity y given by operator j. The reported response given by each operator is adjusted by a data adjustment factor (DAF) determined by his operating arrangement, that ensures that the reporting unit properly represents the sampled unit. Individual operating arrangements reported by an operator are referred to and represented in the data file as subtracts. The quantity y_j of either part of (1) is the sum of the product of the recorded responses by the DAF over all subtracts for operator j. In June, when the full area frame is used, $w_j=1$, but, in subsequent quarters when only a subsample of NOL tracts is used, $w_j=\frac{t_j}{l_j}$ for the NOL operators where l_j is the amount of land in the operation excluding any government pasture or grazing land, and t_j is the amount of land reported for the area tract in June.

In the first part of (1), $\frac{N_h}{u_h}$ is the reciprocal of the probability of selecting a unit from stratum h, adjusted for nonresponse. Equivalently, it is the list frame expansion factor for stratum h where N_h is the size of the population of stratum h and u_h is the size of the set U_h of usable sampled operators drawn from stratum h of the list frame. In the second part of (1), e_{h_k} is the area frame expansion factor for substratum k of stratum h and U_{h_k} is the corresponding set of usable sampled operators from the area frame. For most items in NASS surveys all area records are made usable through either manual or automated imputation.

The indicator variable X_j provides a way by which the effects of missed substitutions may be introduced into the estimator $\hat{\tau}$. Whenever a substitution is missed, $X_j=0$ and the observation y_j disappears from the computation. Currently, under the rules for making a substitution, partial divestitures are ignored, hence $\epsilon=1$. In this case, whatever part of the expansions that could have been attributed to the divested portion of an operation is completely lost. That effect is manifest in (1) by $X_j=0$ for the divested part of the operation. In cases involving full divestiture of an operation, on the other hand, all of the related data may be saved provided that the substitution procedures are followed perfectly. In practice, some full substitutions are missed so that the probability ϵ of missing these will vary between 0 and 1.

A potential substitution can be assigned to one of three categories. It will either be associated with a partial divestiture, a full divestiture, or no divestiture. Accordingly, the terms in (1) were grouped by category and the conditional expectation taken. Assuming that over substitutions do not occur, an estimate of the relative error in the expansion, given a sample of matched usables, becomes:

$$R_o = \left|\frac{E[\hat{\tau}]-\tau}{\tau}\right| = \frac{J}{\tau} a_p + \frac{J\epsilon}{\tau} a_f \qquad (2)$$

where a_p is the expansion of a commodity from those operators who made a partial divestiture, a_f is the expansion of a commodity from those operators who made a full divestiture, and τ is the expansion from the sample of all matched usables. Notice that, in the first term of (2), $\epsilon=1$. R_o can be generalized to include the effects of over substitutions on the expansions. A discussion of its derivation can be found in Appendix IV.

The first term in (2) estimates the relative error that is attributable to missed partial substitutions, and the second term provides an estimate of the relative error due to missed full substitutions. The quantities J, τ, and the a_k's can be readily estimated from the data in the edited Agricultural Survey data sets, leaving ϵ as the only unknown. We cannot infer what it is because there is nothing in the survey data that allows us to measure the proficiency of the survey statisticians in following the substitution procedures. By contacting the operators in Kansas, Kentucky, and South Carolina, an estimate of ϵ was obtained.

We want R_o to behave reliably when it is used to predict the relative error in the expansions from the edited survey data. If it is to be useful in evaluating the problems with the substitution procedures, R_o should provide a way to set limits on the errors in

the expansions over a class of circumstances that one would expect to occur from year to year. Accordingly, the applicability of R_0 must be restricted in scope to conform with our expectations of the kind of circumstances in which the survey process will function. Hence, the following assumptions were made in the derivation of R_0:

1. There exists a continuity of ownership of real estate in the United States.

2. In the event that a valid substitution is missed, the reported values given by the other operators are representative of what would have been the substituted value. In other words, it is assumed that the missed valid substitution is not some kind of outlier.

3. The probability of committing a procedural error is constant across strata. Missing an extreme operator is just as likely as missing an operator in any other stratum.

4. The agricultural industry is stable so that drastic changes in the rate at which farmers go out of business do not occur. The results from this research would not be valid, for instance, in times of a natural disaster, war, or an economic depression.

Given that the assumptions are valid, then the probability J that a sampled operator warrants substitution will be stable from one year to the next. Although J may be estimated by the ratio of the number of reported substitutions to the number of operators who went out of business when no substitutions are missed, not knowing the true number of required substitutions makes J depend on ϵ. To see this, suppose $\epsilon=\frac{1}{3}$, then one out of three potential substitutions will have been missed by the survey statisticians causing the number of reported substitutions to be $\frac{2}{3}$ the true number of required substitutions.

Let $J_0=\frac{|repsub|}{|potsub|}$ where $|repsub|$ is the number of reported substitutions from the Quarterly Agricultural Survey data and where $|potsub|$ is the number of potential substitutions or equivalently the number of operators who have gone out of business. Taking into account the dependency of J on ϵ, an estimate of J can be written as:

$$\hat{J}=\frac{J_0}{1-\epsilon} \qquad 0\le\epsilon<1-J_0 \qquad (3)$$

If the survey statisticians make no errors in following the substitution procedures, then $\epsilon=0$ and $\hat{J}$ can be calculated directly from the survey data since the reported substitutions are exactly the required ones. Unfortunately, ϵ is unknown because there is no information available in the survey data with which to measure the proficiency of the survey statisticians. But, in the worst case, with substitution decisions made at random, $\epsilon=\frac{1}{2}$. By using (3) the estimate for the probable upper value for J would be $\hat{J}=2J_0$. Accordingly, we will scale ϵ so that as ϵ varies from 0 to $1-J_0$, $\hat{J}$ will vary from its minimum value of J_0 to its upper value of $2J_0$. The functional relation between J and ϵ shown in (3) can be rewritten to

incorporate the probable upper value of $\hat{J}$ as:

$$\hat{J}=\frac{J_0}{\left(1-\frac{\epsilon}{2(1-J_0)}\right)} \qquad 0\le\epsilon<1-J_0 \tag{4}$$

One can test (4) by letting $\epsilon=0$ and $\epsilon=1-J_0$ and observing that $\hat{J}=J_0$ and $\hat{J}=2J_0$, respectively. The approximate variance of $\hat{J}$ can be written as:

$$var(\hat{J}) \doteq \frac{var(J_0)}{\left(1-\frac{\epsilon}{2(1-J_0)}\right)^2} \qquad 0\le\epsilon<1-J_0 \tag{5}$$

Numerical values of $\hat{J}$ and $var(\hat{J})$ follow immediately. By using the December 1993 data for example, $J_0=\frac{115}{1194}=0.096$ and with $\epsilon=\frac{1}{9}$, the value of ϵ determined by contacting previously interviewed operators, $\hat{J}=0.108$ and $var(\hat{J})=8.08x10^{-5}$. Therefore, for the Agricultural Surveys, approximately one out of ten sampled operators who go out of business require substitution.

In keeping with the goal of establishing an upper bound on the bias in the expansions due to missed substitutions collectively for all commodities, it is necessary to estimate var(R_o) from which the upper bound, based on three standard deviations from R_o, can be obtained with an approximate coefficient of confidence of 99 percent. To that end, the estimate of $var(\hat{J})$ and the estimates of $var(a_p)$ and $var(a_f)$ found from bootstrapping methods readily provide the upper bound as a function of ϵ. Specifically, by applying a Taylor series approximation to (2), the variance of R_o becomes:

$$var(R_o)= var(\hat{J})\left(\frac{a_p}{\tau}\right)^2+\left(\frac{J}{\tau}\right)^2 var(a_p)+ var(\hat{J})\left(\frac{a_f\epsilon}{\tau}\right)^2+\left(\frac{J\epsilon}{\tau}\right)^2 var(a_f) \tag{6}$$

It follows at once that the upper bound on the relative error due to missed substitutions at an approximate coefficient of confidence of 99 percent is $R_o+3\sqrt{var(R_o)}$; clearly, both R_o and its upper bound are functions of ϵ.

The worst case involving the use of the substitution procedures corresponds to $\epsilon=1$. At that point, $R_o+3\sqrt{var(R_o)}$ reaches its maximum value. It is the upper bound on the relative error in the expansions, if the substitution procedures are omitted. The upper bound is easier to describe when R_o and $R_o+3\sqrt{var(R_o)}$ are plotted as in Figure 1 against ϵ. By letting ϵ vary over all possible values, a picture can be drawn of how the relative error in the expansions varies as more mistakes leading to missed substitutions are made. We can draw that picture starting from the point of complete and perfect compliance with the procedures when $\epsilon=0$ to the total disregard of them when $\epsilon=1$.

FINDINGS

Without having to contact any operator, the graph shows how the bias in the expansions changes as the probability of missing a substitution varies over all possible values. While it is informative to see the relationship of the bias as a function of the probability of missing a substitution, the upper bound shown in Figure 1 accounts for the worst situations involving substitution. It is

obvious that the observed absolute relative errors, designated by ★'s, found by contacting operators in Kansas, Kentucky, and South Carolina fall well below the predicted upper bound and that the proficiency 1-ϵ of the survey statisticians in following the substitution procedures is greater than 90 percent. That is, in this study they committed an error in following the procedures, collectively, less than 10 percent of the time.

The line labeled *Operational* in Figure 1 shows the predicted relative error in the estimate of the amount of land in farms under the current substitution procedures, as it changes with the probability of missing a substitution. If the substitution procedures were augmented to address partial divestitures of an operation, then the predicted relative error due to missed substitutions would follow the line labeled *Proposed*. At most, missed partial substitutions introduce an error of 0.35 percent into the expansion for the amount of land in farms. All of the observed error in South Carolina and 80 percent of the observed error in Kansas were attributed to missed partial substitutions. No errors were observed in Kentucky.

The observed relative errors found in Kansas, Kentucky, and South Carolina lie well below the predicted upper bound. Since they had very low values of ϵ, they do not give an indication of the accuracy of the predicted upper bound as ϵ approaches 1, the case in which the substitution procedures are omitted. Noticing in Table 4 that the probability of missing a substitution and the probability of making an over substitution are about the same, the number of over substitutions should approximately equal the number of substitutions that are missed. Therefore one may hypothesize that the number of reported substitutions in the survey data is the same as the number of required substitutions. If that is the case, then the expansions from only those records for which substitutions were reportedly made could be used to estimate the bias which would be generated in the expansions if the substitution procedures were omitted. By comparing the expansions from the reported substitutions to the state level expansions from all records, a way was found to assess the accuracy of the predicted upper bound at $\epsilon = 1$.

The Survey Processing System was used to compute the expansions for this approach in assessing the accuracy of the predicted upper bound. However, because information concerning partial substitutions could not be captured in the SPS indications, they were inflated by 0.28 percent to adjust for missed partial substitutions. The adjustment is the same size as the predicted relative error for land in farm corresponding to $\epsilon = 0$ in Figure 1. At that point where $\epsilon = 0$, the substitution procedures are being followed without errors hence the bias in the expansions comes from missing just the partial substitutions. By making the adjustment for missed partial substitutions to the relative errors corresponding to the state level SPS indications and to the national level SPS indications, a useful comparison can be made between the relative errors predicted by theory and those derived empirically via the SPS indications.

Using each item in all the States from the 1993 December Agricultural Survey, a total of 4,648 relative errors were computed as if the substitution procedures were omitted. They appear in Figure 2 to form an empirical cumulative distribution function of the

Relative Error in the Expansion for the Amount of Land in All Farms

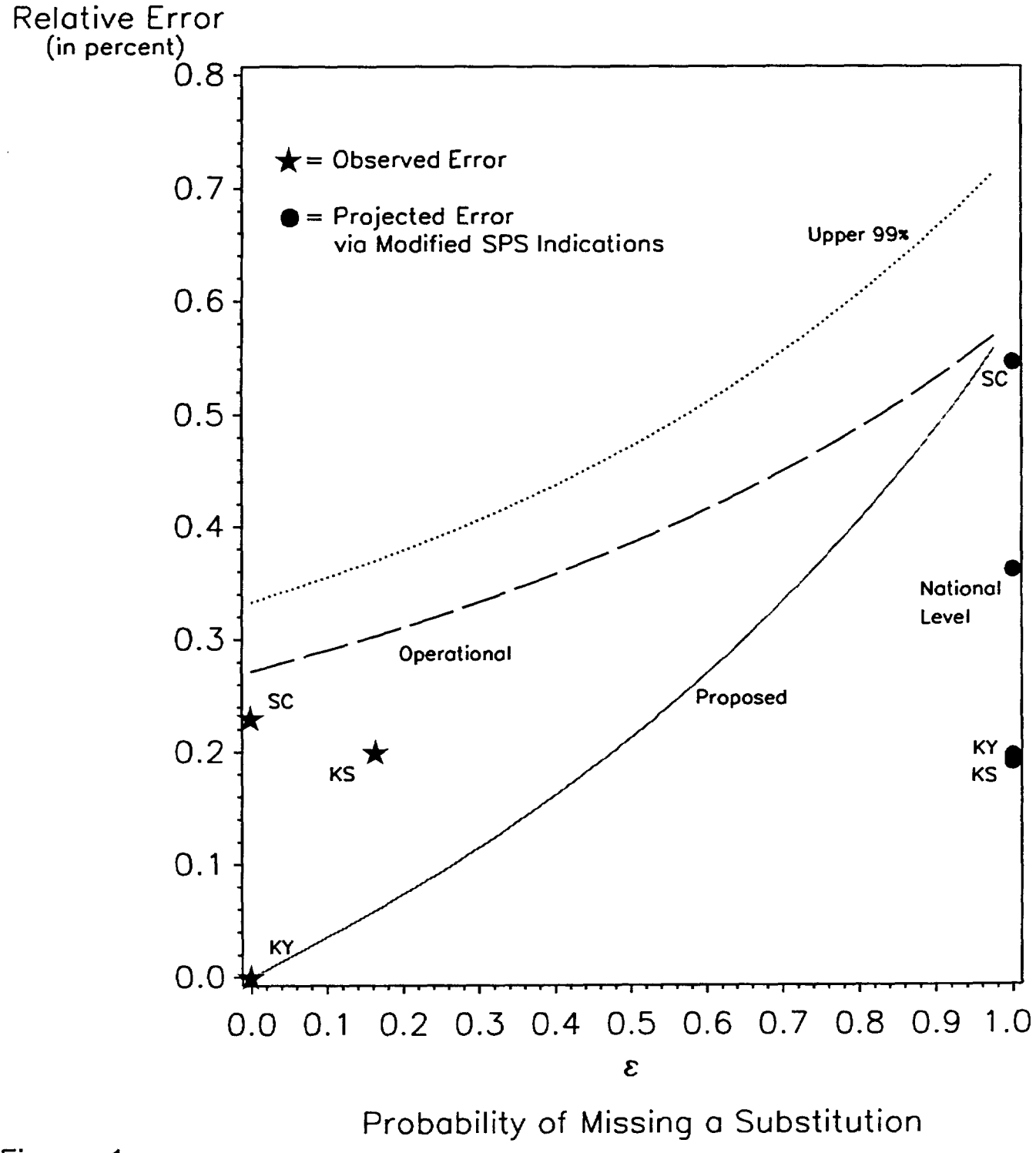


Figure 1

relative error. The upper 99 percent confidence limit of 0.7 percent for the predicted relative error when $\epsilon = 1$ occurs in Figure 2 at the 92nd percentile. Similarly, 327 relative errors corresponding to each item at the national level appear in Figure 3. At the national level, the upper 99 percent confidence limit of 0.7 percent for the predicted relative error when $\epsilon = 1$ occurs at the 99th percentile. Although the percentile for the state level indications falls short of the predicted upper bound's level of significance, its high value provides strong evidence that the upper bound is a very good one. The state level comparisons of the SPS land in farm indications for those records for which a substitution was made to the expansions for all records appear in Figure 1 as ●'s for Kansas, Kentucky, and South Carolina at $\epsilon = 1$.

A tabulation of items for which there was a relative error in the special SPS indications at the national level greater than 0.29 percent when adjusted for missed partial substitutions appears in Appendix V. A similar tabulation appears in Appendix VI relating to the special state level SPS indications of those items for which there was a relative error greater than or equal to 0.7 percent. In both appendices, the three items, land in farm, hogs, and corn stocks, which were the focus of the research on missed substitution are flagged by a solid dot. None have significant relative errors at the national level shown in Appendix V, but at the state level for Kansas, Kentucky, and South Carolina there are instances of significant relative errors associated with those items. Each instance, however, corresponds to an over substitution.

Even if the assumption that equal likelihoods of missing a substitution and of making an over substitution implies that the number of reported and required substitutions are the same, since larger relative errors were observed with over substitutions, the distribution of relative errors is probably skewed to make the upper tail heavier. As a result, the observed percentiles, shown in Figures 2 and 3, for the predicted upper bound are probably smaller than they should be. But in view of how high they are, we may conclude that the predicted upper bound on the relative errors from missed substitutions is valid.

In order to discern whether a missed full or partial substitution occurred, it was necessary in doing the research to match records across quarters from the Agricultural Survey data by operator. The observed relative errors caused by incorrectly following the substitution procedures are based on the expansions shown in Table 6 for land in farm, hog inventory, and corn stocks from these matched records. The items that were studied were chosen to represent an attribute of every farm, the amount of land operated; a relatively rare item, hogs; and an important item that in general does not correlate very well with either land or livestock, corn stocks.

The observed missed substitution that was found affected only the expansion for land in farm. The expansions for hogs and for corn stocks were unaffected, since the new operator did not have either of these items. Eleven reported substitutions in Kansas, Kentucky, and South Carolina were examined; three of them were over substitutions. As can be seen in Tables 7, 8, and 9, they introduced a significant positive error into the expansions for hogs and corn stocks but not for land in farm. In contrast to the effects caused by missed substitutions,

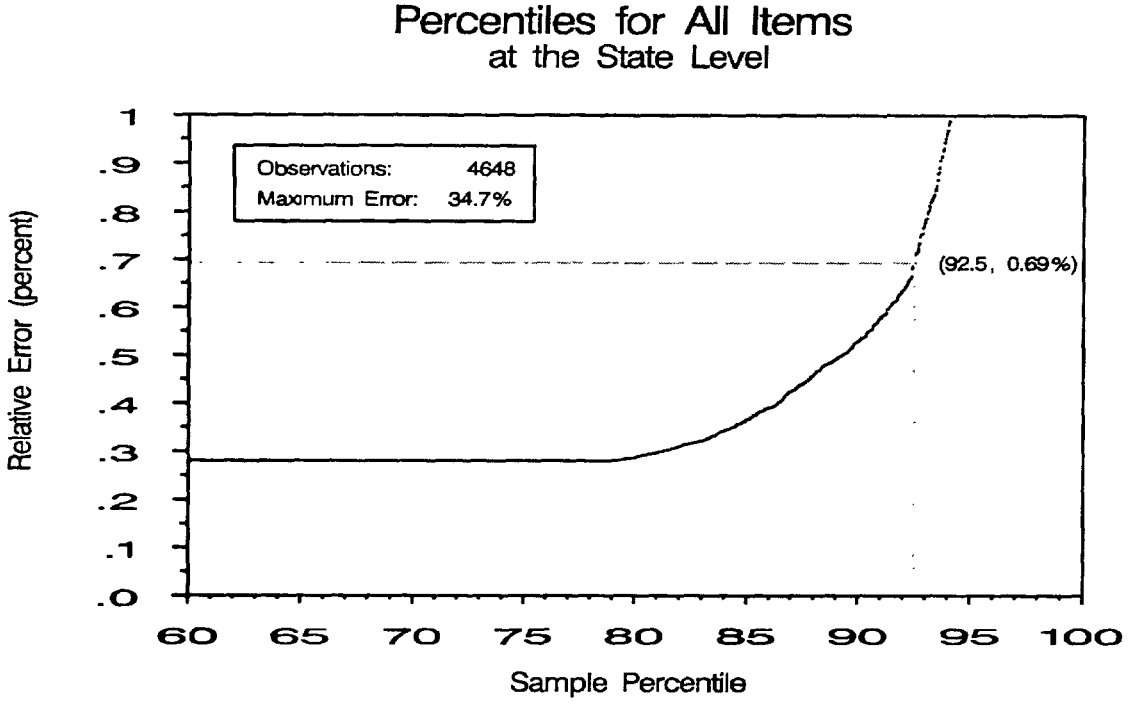


Figure 2. The expansions from only those records for which a substitution occurred were compared to the expansions from the complete set of records at the state level in the December 1993 Agricultural Survey. Each of the 4,648 items over all 45 States provided an estimate of the relative error if the substitution procedures were omitted. The predicted upper bound of 0.7 percent on the relative error occurs at the 92nd percentile for the state level expansions.

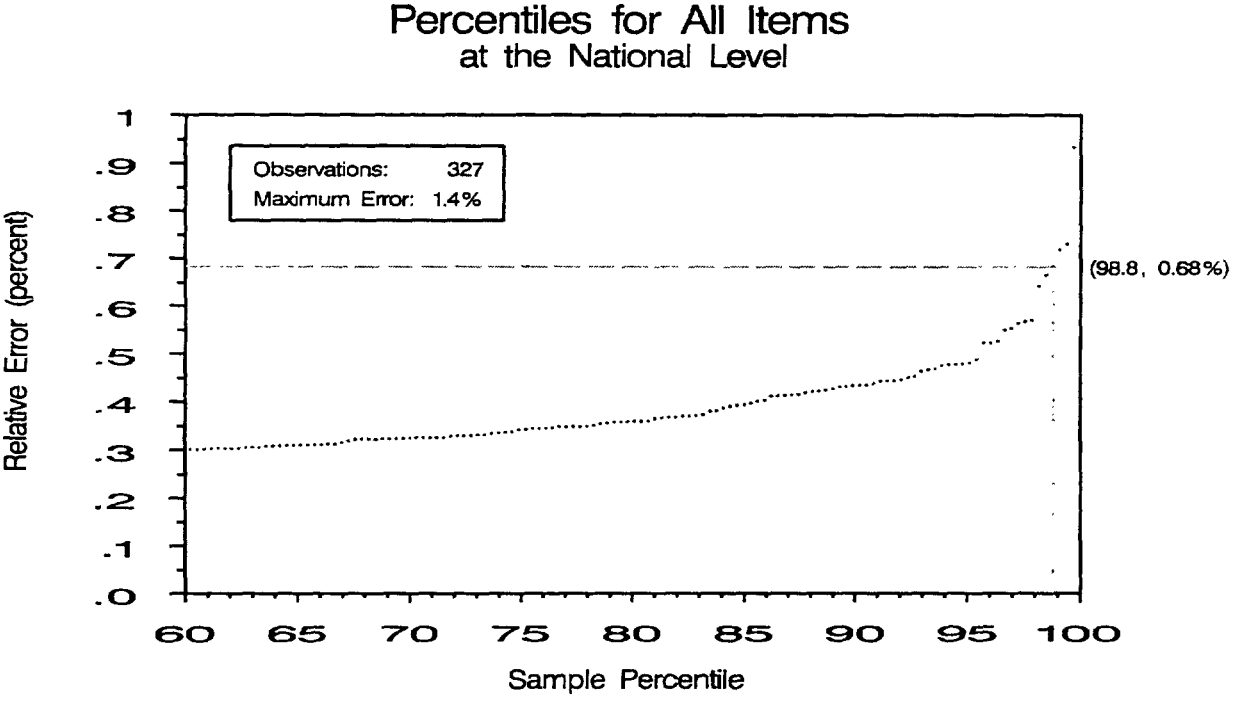


Figure 3. The relative errors in the expansions at the national level provided by 327 items from only those records for which a substitution occurred show that the predicted upper bound of 0.7 percent occurs at the 99th percentile if the substitution procedures were omitted.

Table 6.
Expansions at the State Level of All Matched Records from the December 1993 Agricultural Survey

State	Land in Farm (acres)	Hog Inventory	Corn Stocks (bushels)
Kansas	33,118,841	832,761	25,774,351
Kentucky	12,571,665	523,844	38,577,025
South Carolina	6,502,807	275,169	1,830,639

the effects of over substitutions were substantially larger in items other than land in farm by a factor of ten. This is probably because at the time that the over substitutions were made, the respective operations were operating at full capacity while those qualifying for substitution were operating with minimal agricultural activity. From the tables, the over substitution in Kansas caused a positive bias in the expansion for hogs of +5.11 percent whereas missed substitutions caused in the worst case among the three commodities an error of about -0.3 percent.

Perhaps it may be a normal practice for operators who intend to go out of business to reduce the activity of their operation gradually. By the time an operator reports that he is no longer farming, the operation is likely to be dormant or operating at minimum capacity. It may be typical for the new operators, in turn, not to have the financial resources to bring a recently divested operation to full capacity. Consequently, when the rare event does occur for requiring a substitution, the effects of missing it on the expansions produce the small error which was observed in the research. In fact, the estimate of the hog inventory would have been unaffected and the estimate of the amount of corn stocks would have been only slightly affected in Kansas, Kentucky, and South Carolina by eliminating the substitution procedures altogether. But in the case of over substitutions, the operations probably will be already operating at full capacity and the expansions, for whatever commodities they might have, will probably be significantly affected.

Concerns had been expressed over the effects of missing an extreme operator (EO). As with operators classified in other strata, the effects due to over substitutions or missed substitutions of extreme operators on the expansions propagate to subsequent quarters, unless the incorrect coding of the records is changed in a later quarter. Only one of two extreme operators was found during the research to cause an error in the expansions. That one extreme operator, an over substitution, produced a large positive error in the estimate of corn stocks. Other smaller operators listed in Appendix I produced or could have produced, if valid substitutions had been missed, comparable errors in the expansions to those of the extreme operators. Whether or not an operator is an extreme operator, the size of the relative error in the expansions tends to be about the same as that produced by operators in any other strata. This is because the expansion factors tend to

Table 7.
Observed Relative Error in Land in Farm Expansion

State	Multiple Contacts	Over Substitution	Missed Full Substitution	Valid Substitution	Missed Partial Substitution	Observed Relative Error via Sub. Procedures	Relative Error if Sub. Procedures Were Omitted
Kansas	0%	+0.02%	-0.06%	-0.19%	-0.14%	-0.18%	-0.39%
Kentucky	0%	0%	0%	-0.05%	0%	0%	-0.05%
South Carolina	0%	+0.40%	0%	0%	-0.23%	+0.17%	-0.23%

Table 8.
Observed Relative Error in Hog Expansion

State	Multiple Contacts	Over Substitution	Missed Full Substitution	Valid Substitution	Missed Partial Substitution	Observed Relative Error via Sub. Procedures	Relative Error if Sub. Procedures Were Omitted
Kansas	0%	+5.11%	0%	0%	0%	+5.11%	0%
Kentucky	0%	0%	0%	0%	0%	0%	0%
South Carolina	0%	+0.7%	0%	0%	0%	+0.7%	0%

Table 9.
Observed Relative Error in Corn Stocks Expansion

State	Multiple Contacts	Over Substitution	Missed Full Substitution	Valid Substitution	Missed Partial Substitution	Observed Relative Error via Sub. Procedures	Relative Error if Sub. Procedures Were Omitted
Kansas	0%	0%	0%	0%	0%	0%	0%
Kentucky	0%	0%	0%	-0.51%	0%	0%	-0.51%
South Carolina	0%	+1.91%	0%	0%	0%	+1.91%	0%

equalize the effect of operators in the various strata.

If the substitution procedures had been omitted before this research began, the resulting errors from missing all substitutions would be those indicated in the last column in Tables 7, 8, and 9. They represent the maximum error that the substitution procedures could have possibly eliminated in Kansas, Kentucky, and South Carolina in 1993. All three operators who were substituted unnecessarily had been farming before June 1. The errors due to these over substitutions in the expansions are also shown in Tables 7, 8, and 9. If additional probing had been done during the interviews, the SSO survey statisticians would have learned that these substitutions were not needed.

CONCLUSIONS

To explain the apparently small effects on the expansions due to missed substitutions, Ned Jones suggested that the necessary conditions for making substitutions rarely exist. Although an accurate measurement of the rate at which errors are made in following the substitution procedures was not available to Ned Jones, the existence of missed substitutions was known. It offered an explanation for the downward trend that is usually observed in the calculated number of farms and the amount of land in farms from June to March. It was this observed trend that led the Working Group to Standardize Survey Procedures to request research on missed substitutions. Extensive research done by Jeff Bailey, in the meantime, found a 2 percent decline from June to December and indicated that the downward trend is probably caused by:

1. The initial inclusion of non-operators in the sample, primarily in the June sample.

2. The erroneous coding of records indicating that a bona fide operator is no longer in business.

When compared to the observed relative error of -0.2 percent in the expansions for December due to missed full and partial substitutions, it becomes apparent that these overshadow missed substitutions in importance.

The results of this research on substitution substantiate the idea proposed by Ned Jones [5] that the conditions for requiring a substitute record are rare. Moreover, if in the worst case where the substitution procedures are omitted altogether, the absolute value of the relative error in the expansions will be less than 0.7 percent with a coefficient of confidence of 99 percent. By establishing an upper bound on the relative error rather than addressing every possible situation involving substitution, a comprehensive way was found to account for all of the error associated with missed substitutions.

This research on missed substitutions affirmed the research performed by Jeff Bailey [1,2], measured the reduction of the bias in the expansions if partial substitutions were made, revealed the importance of over substitutions on the expansions, confirmed that the substitution procedures are being followed accurately, indicated that missing all substitutions would have a relatively minor effect on the expansions, and showed that the problem of multiple chances of selection is unimportant.

In summary, the answers to the questions asked by the Working Group can be stated as follows:

1. The effect on the expansions from multiple chances of selection in the substitution process is essentially zero.

2. The effect on the expansions from over substitution can be significant as the statistics in Tables 7, 8, and 9 suggest.

3. The effect on the expansions due to missed partial substitutions is less than 0.35 percent in absolute value with a coefficient of confidence of 99 percent.

4. The effect on the expansions if the substitution procedures are omitted is less than 0.7 percent in absolute value with a coefficient of confidence of 99 percent.

5. Unlike their effects on the estimates of variances, extreme operators affect the survey indications in situations involving substitution to the same degree as do operators from other strata.

6. Missed substitutions constitute an insignificant component in causing the observed downward trend across quarters in a survey cycle.

RECOMMENDATIONS

1. It is recommended that the survey complications associated with new rules on partial substitutions be weighed against the potential bias reduction before implementing additional rules.

2. It is recommended that no provisions be introduced into the substitution procedures for addressing multiple chances of selection.

3. It is recommended that a remedy be found for making the SPS edit readily accept records involved in substitution or whose substitution coding needs to be revised.

4. It is recommended that measures be taken to reduce over substitutions. The capability to defend survey coverage by allowing new operations, or those with changed names, to enter the sampled universe should be balanced with operational capabilities for preventing duplication due to erroneous substitution.

REFERENCES

[1] Bailey, Jeffrey T., (1993), "Time Related Coverage Errors and the Adjustment Factor (DAF)", SRB-93-03, National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, D.C.

[2] Bailey, Jeffrey T., (1994), "Results of the December 1993 Acreage Reconciliation Study", SRB-94-07, National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, D.C..

[3] Bosecker, R. R (1984), "SRS List vs. Area Overlap Determination: List Dominant and Frozen Domain Procedures", Statistical Reporting Service, U. S. Department of Agriculture, Washington, D.C..

[4] Dillard, Dave, (1993), "Design Specification to Standardize Survey Procedures", National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, D.C..

[5] Jones, Ned (1988), "QAS Substitution Analysis", SMB-88-04, National Agricultural Statistics Service, U. S. Department of Agriculture, Washington, D.C..

[6] NASS, (1993), "Agricultural Surveys, Supervising and Editing Manual", National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C..

APPENDIX I

Individual Records Used for Finding the Relative Errors in the Expansions from the December 1993 Agricultural Survey

State	Stratum	ID	Tract	Land in Farm	Expanded Land in Farm	Number of Hogs	Expanded Number of Hogs	Amount of Corn Stocks	Expanded Amount of Corn Stocks	Explanation
KS	11	2149	4	1120	19675	0	0	0	0	Valid Substitution*
KS	11	3082	7	11	9014	0	0	0	0	Valid Substitution*
KS	64	800098540	1	642	15410	0	0	0	0	Valid Substitution*
KS	63	801220650	1	306	11416	0	0	0	0	Valid Substitution*
KS	95	824000360	1	8000	8000	0	0	0	0	Valid Substitution
KY	40	1267	31	2.6	780	0	0	0	0	Valid Substitution*
KY	75	790139250	1	1157	3502	0	0	45000	168000	Valid Substitution*
KY	80	790451370	1	413	2693	0	0	4000	27375	Valid Substitution
KS	76	801213250	1	720	4787	603	67853	0	0	Over Substitution
SC	97	790219690	1	6757	6757	2505	2505	71250	71250	Over Substitution
SC	40	2283	1	60	19425	0	0	0	0	Over Substitution
KS	11	1134	8	568	19192	0	0	0	0	Missed Full Substitution
KS	61	800801580	1	33	7233	0	0	0	0	"Missed" Partial Substitution
KS	63	80173680	1	200	10036	0	0	0	0	"Missed" Partial Substitution
SC	40	1332	14	14	645	0	0	0	0	"Missed" Partial Substitution
SC	66	907113820	1	350	8109	0	0	0	0	"Missed" Partial Substitution

* These are records that were correctly coded for substitution in September. In December, their coding was purposely changed in order to pass them through the SPS edit. No adverse consequences were experienced in the expansions by the revised though incorrect coding.

APPENDIX II

Procedures for Conducting the Substitution Research Study
For January 1994

General:
The Kansas, Kentucky, and South Carolina State Statistical Offices were selected and agreed to participate in a Survey Quality Research Study on determining the effect of missed substitutions on the Quarterly Agricultural Survey expansions. A detailed study of the effects of missed substitutions requires information which cannot be obtained other than by reinterviewing out of business operators, interviewing new operators, or both. The three states which were asked to participate were selected because collectively they can provide a sample of about 225 matched out of business operators who had been interviewed in at least two quarters, are already participating in an acreage reconciliation research project, and are geographically dissimilar.

By parcels we mean parcels of land that came from different people. For example, an enumerator may learn that a new operator bought the out of business operation and also rents from two other landowners. The new operator will have therefore acquired three parcels. The size of a parcel is not important; rather, it is important that the land has a potential for agriculture. Complicated land transfers probably will not occur.

Certain key aspects of the substitution research study will stand out during the conduct of it. Every sampled operator and probably every new operator in this study may have to be interviewed in order to ascertain:

1. The time when a change in operating status occurred; that is, whether land was sold, rented, or leased before or after June 1.
2. To whom the operation of the land was transferred.
3. Whether the recipient operator began farming individually before or after June 1.
4. How many parcels of land the recipient operator acquired after June 1.

Misreporting of acreage and miscoding the operating status are other key aspects that may be encountered during the study. Depending on the outcome of the reconciliation study, only those operators that had a real change in land operated will be used in the substitution study sample. For operators in the study sample, we need to learn about the disposition of their operations, i.e., what happened to the divested land; did a change really occur; when and to whom was the land given. There may be some out of business operators in the substitution research sample which should have been coded out of business in June and should not have been contacted in a subsequent quarter. The initial operating status therefore of every out of business operator in the substitution research sample should be accurately determined. Finally, the most important aspect of doing the research study is the request to provide an explanation of the investigative process. A response from each operator should be described so that somebody who is not cognizant of the

telephone interview can correctly decide whether the operator qualifies for a substitute record by reading the enumerators' notes.

Taking into account completions, refusals, and inaccessibles, allow about 10-15 minutes (based on the Kentucky pilot study) to complete a questionnaire. The stratum, substratum, ID, tract, September CASEID number, December CASEID number, and reported acres of land in farm will be given to each state by the Survey Quality Research Section. The sample will be sorted in ascending order to help in matching a research sample operator with the CATI review sheet or paper questionnaire. Making a paper copy of the CATI review will make interviewing much easier. Every operator will have been in at least two quarterly surveys. No December refusals will be included in the substitution research sample. Some September refusals may be in the research sample provided that they cooperated in the December survey. All operators in the research sample will have been contacted in either the June and September, September and December, June and December, or June, September and December surveys.

Five examples of completed study questionnaires are attached with this proposal. They illustrate instances of correcting the land in farm acreage, the use of the questionnaire as a work sheet for recording names, telephone numbers, and reminders, and the kinds of comments that would be appropriate for the narrative section. Among the examples, there is a case where an out of business operator was found to have been out of business before June (example 1); other cases where some questions could not be answered due to lack of information (examples 2 and 4); and a case involving a missed partial substitution (example 5). These examples should describe the common situations that a telephone enumerator will encounter.

Steps to Follow:
If the SSO has assigned a sequence number to a paper questionnaire where a CASEID number is missing, then the sequence number should be written on the research study questionnaire. That will help in identifying a research questionnaire with the corresponding state questionnaire in case a clarification is needed after the study is completed. If a "-1" appears in the land in farm acreage column on the study questionnaire, then that operator was not in that month's sample. In Example 1, the operator was not in the June sample, so that land in farm acreage was not reported and appears as a "-1" on the study questionnaire. Some information, like names of new operators, can be obtained immediately from the enumerator notes on the paper questionnaires and CATI review sheets. Sometimes an enumerator will discover during an interview that the original operator will know the new operator well enough to say whether the new operator had been farming before or after June 1. The out of business operator may comment, for instance, that the new operator has been cutting hay on his land for years and that the new operator does not operate any other land. In this case, there would be no need to contact the new operator because we would know that the new operator already began operating before June 1 and that the new operator did not acquire any parcels after June 1.

APPENDIX II

Note: Completing a study questionnaire may require a reinterview of the previous operator or an interview of the new operator or both.

Mention that NASS is doing a quality control study on how well our survey procedures are working.

Response is voluntary and all information will be kept confidential.

The notes which will appear in the Action and Decision Narrative section of the research study questionnaire will be very important. The telephone enumerators should make a special effort to take comprehensive notes so that someone else can accurately answer the questions on the questionnaire just by reading the notes. If multiple interviews are made, clearly mark which information goes with which respondent. If there is insufficient space left on the study questionnaire for all of the notes, then use another sheet of paper and staple it to the study questionnaire. A blank questionnaire will be provided to make extra copies if needed.

Question number 5 on the questionnaire is used to determine if a new operator is NOL. If a new operator who indicated that he began farming after June 1 should appear on the list frame then determine whether or not the operator is NOL.

Care should be exercised in verifying the original coding of a substitute record because it is possible that a substitution may have been made unnecessarily. Care should also be exercised when probing an out of business operator to learn when the change in operating status occurred. Quite possibly, a reported out of business operator was already out of business before June 1. For instance, a neighbor of an out of business operator may rent the operator's land annually, but in June the operating status was incorrectly coded as being in business due to a misunderstanding by the respondent about the meaning of the word "operate" which he may have construed to mean "own". Learning when the new operator took over may require persistent probing by an enumerator.

The research questionnaires will be distributed to the participating states in early January 1994. They should require about 30 manhours per state to complete. Evening call backs will be common and refusals will occur. A reasonable number of attempts to contact "no answers" should be made.

After the questionnaires have been completed by the telephone enumerators, the survey statistician should review them for having legible and sufficient notes. Good notes will be greatly appreciated. Return the research study questionnaires to the Survey Quality Research Section no later than 30 days after receiving them. Keep the CATI review sheets and paper questionnaires at hand in case there is a need to clarify something in the study questionnaire. Once the data from

APPENDIX II

the research study have been tabulated and deemed complete, then each state will be notified that the CATI review sheets and paper questionnaires will no longer be needed for the substitution research study.

Questions on how to conduct the research study can be made to Mike Fleming in the Survey Quality Research Section at 703-235-5213 ext.170 from 0700-1600 EDT. Please send the completed study questionnaires to:

Mike Fleming
USDA/NASS
Research Division
Room 305
3251 Old Lee Highway
Fairfax, Virginia 22030-1501

APPENDIX II

Example 1

STATE	DECEMBER CASEID	SEPTEMBER CASEID	STRATUM	SUBSTRATUM	ID	TRACT	Land in Farm JUNE	SEPT	DEC
KY	0186	0067	71	0	800859910	1	-1	~~300~~ Ø	0

1.Was there a change in acres operated
after June 1? ..... ☐YES ☑NO

1a. If no, then correct land in farm.

1b. If yes, who is now operating the land?
List names, addresses, and telephone numbers.

Name: JAY SMITH
Address:
City: State:
Zip: Phone: 509-5289

2.When did the new operator start
farming?..... ☑Before June 1 ☐After June 1 ☐UNKNOWN

3.How many parcels did the new
operator acquire after June 1?... [2]

4.What 923 item code entry was originally
made on CATI or paper questionnaire? ☐1 ☐2 ☑Blank

5.Is the new operator on the
list frame?..... ☐YES ☑NO ☐UNKNOWN

6.Number of interview attempts: Out of business operator [5] NO ANSWER
New operator [2]

ACTION AND DECISION NARRATIVE:

1) The new operator's name, Jay Smith, was found on the September CATI review sheet and was called because the out of business operator could not be contacted.
2) Smith's wife said that Jay has been renting this land for two years so the out of business operator was already out of business in June.
3) Smith was not on the List Frame.
4) Smith rents two other places since June 1.
5) No substitution is required.

APPENDIX II

Example 2

STATE	DECEMBER CASEID	SEPTEMBER CASEID	STRATUM	SUBSTRATUM	ID	TRACT	Land in Farm JUNE	SEPT	DEC
MINN	0607	0589	62	0	887021080	1	50	0	-1

1.Was there a change in acres operated
after June 1? ..... ☑YES ☐NO

1a. If no, then correct land in farm.

1b. If yes, who is now operating the land?
List names, addresses, and telephone numbers.

Name:
Address: ?
City: State:
Zip: Phone:

2.When did the new operator start
farming?..... ☐Before June 1 ☐After June 1 ☑UNKNOWN

3.How many parcels did the new
operator acquire after June 1?... ☐ DONT KNOW

4.What 923 item code entry was originally
made on CATI or paper questionnaire? ☐1 ☐2 ☑Blank

5.Is the new operator on the
list frame?..... ☐YES ☐NO ☑UNKNOWN

6.Number of interview attempts: Out of business operator ☐1
New operator ☐

ACTION AND DECISION NARRATIVE:

1) The operator reported that he no longer rents the land and does not own any.
2) The landlord lives in Chicago and his name was not recorded on the CATI review sheet and the operator didn't remember his name or phone number.
3) The operator had cut hay on this land last year and the landlord found somebody else to cut the hay in July. Didn't know his name.
4) Could not determine how many parcels the new operator acquired since June 1.
5) The need for a substitute record could not be determined.

APPENDIX II

Example 3

STATE	DECEMBER CASEID	SEPTEMBER CASEID	STRATUM	SUBSTRATUM	ID	TRACT	Land in Farm JUNE	SEPT	DEC
MISS	0513	0512	40	10	1433	7	175	200	0

1.Was there a change in acres operated
after June 1? ..... ☑YES ☐NO

1a. If no, then correct land in farm.

1b. If yes, who is now operating the land?
List names, addresses, and telephone numbers.

Name: MARK BRAD
Address: RT 1
City: ALES State:
Zip: Phone: 318-9681

2.When did the new operator start
farming?..... ☐Before June 1 ☑After June 1 ☐UNKNOWN

3.How many parcels did the new
operator acquire after June 1?... [2]

4.What 923 item code entry was originally
made on CATI or paper questionnaire? ☐1 ☐2 ☑Blank

5.Is the new operator on the
list frame?..... ☐YES ☑NO ☐UNKNOWN

6.Number of interview attempts: Out of business operator [1]
New operator [4]

ACTION AND DECISION NARRATIVE:

1) CATI did not have address and telephone number of new operator.
2) The out of business operator was contacted and gave the name of the new operator and the town in which the new operator resides. He said that he sold his entire farm after June 1 and is retired.
3) New operator was eventually contacted and indicates that he did not farm before June 1. He also rents from one other farmer.
4) New operator does not appear on the List Frame.
5) Substitution is required.

APPENDIX II

Example 4

STATE	DECEMBER CASEID	SEPTEMBER CASEID	STRATUM	SUBSTRATUM	ID	TRACT	Land in Farm JUNE	SEPT	DEC
KY	2013	2150	84	0	80011980	1	600	600	~~100~~ 200

1.Was there a change in acres operated
after June 1? ..... ☑YES ☐NO

1a. If no, then correct land in farm.

1b. If yes, who is now operating the land?
List names, addresses, and telephone numbers.

Name: HARRIS ZERO
Address: RT 2
City: HOPE State: KY LANDLORD
Zip: Phone: 421-0088

2.When did the new operator start
farming?..... ☑Before June 1 ☐After June 1 ☐UNKNOWN

3.How many parcels did the new
operator acquire after June 1?... ☐ DON'T KNOW

4.What 923 item code entry was originally
made on CATI or paper questionnaire? ☐1 ☐2 ☑Blank

5.Is the new operator on the
list frame?..... ☑YES ☐NO ☐UNKNOWN

6.Number of interview attempts: Out of business operator [1]
New operator [8] NO ANSWER

ACTION AND DECISION NARRATIVE:

1) The operator was re-interviewed. He indicated that in the winter he does not rent land.
2) He reported that the landlord was an operator prior to June 1.
3) The landlord of the 400 acres is on the list.
4) This would not be a partial substitution case.

APPENDIX II

Example 5

STATE	DECEMBER CASEID	SEPTEMBER CASEID	STRATUM	SUBSTRATUM	ID	TRACT	Land in Farm JUNE	SEPT	DEC
MINN	1512	0072	12	4	1421	5	30	30	25

1.Was there a change in acres operated after June 1? ..... ☑YES ☐NO

1a. If no, then correct land in farm.

1. If yes, who is now operating the land?
List names, addresses, and telephone numbers.

Name: MAX HILL
Address: RT 1
City: WINCHESTER State: MINN
Zip: Phone: 253-5238

2.When did the new operator start farming?..... ☐Before June 1 ☑After June 1 ☐UNKNOWN

3.How many parcels did the new operator acquire after June 1?... [1]

4.What 923 item code entry was originally made on CATI or paper questionnaire? ☐1 ☐2 ☑Blank

5.Is the new operator on the list frame?..... ☐YES ☑NO ☐UNKNOWN

6.Number of interview attempts: Out of business operator [2]
New operator ☐

ACTION AND DECISION NARRATIVE:

1) The operator was re-interviewed. He reported that he sold 5 acres to his son after June 1.
2) He also reported that his son was not farming prior to June 1 and that his son didn't buy any more land and doesn't rent land.
3) The son is not on the list frame.
4) This would be a partial substitution case.

APPENDIX III

Substitution Research Study
For January 1994

Author: Mike Fleming
Survey Quality Research Section

Introduction

A part of the problem of studying the effects of missed substitutions lies with the complexity of the rules under which the operational procedures fall. Although substitutions can be missed for various reasons such as misreporting, encountering uncooperative operators, and miskeying the raw data, the complexity of the decision making rules which govern the substitution procedures stands out as the most likely reason for explaining why most missed substitutions occur. From the perspective of researching the effects of missed substitutions on the total expansions, the rarity of cases which warrant substitution makes the problem difficult in two ways. First, an adequate estimate of the incidence of potential substitutions can only be done with a large data set that is comprised of many states; second, missed substitutions produce in expectation a very small effect on the total expansions and are obscured by non-sampling errors.

In preparation for the January substitution research study, a pilot study was conducted in the Kentucky State Statistical Office during the last week in September to pretest the research study questionnaire. As a result of the pilot study, the design of the research study questionnaire was simplified, some important observations were made for determining the January study sample size, and the relative errors due to missed partial substitutions, multiple contacts, and over substitutions were observed and compared with their predicted values.

Recommendations for Doing Substitution Research

The Working Group to Standardize Survey Procedures recommended that research on substitution be conducted in two categories:

1. The effects of missed substitution when the sampled name was inbusiness during part of the survey reference period and turns over all of the operation to someone who has no chance of being sampled during the survey period.

2. The effects of missing data caused when a sampled name turns over part but not all of the operation to someone else.

Once the spring classify process for revising the list frame is completed in February or March, no further revisions to the list frame are allowed for the remainder of the survey year for sampling purposes. If an operator who is on the list frame goes out of business during the survey year, the new operator may be designated his substitute according to a specific set of rules found in Section 6 of the Supervising and Editing Manual. A substitution may not be made,

APPENDIX III

as a result perhaps of some kind of procedural mistake or loophole in the procedures. That lost record, caused by missing a substitution, generates a negative error in the total expansions now and in subsequent surveys throughout the survey year.

Conceivably, all previously active out of business operators could warrant a substitute record. On the other hand, if the survey process is deemed absolutely reliable, then the reported substitutions in the final edited data sets would account for every required substitution. In the latter case, the error rate of missing a substitution would be zero. Missed substitutions probably do occur, however, since we know that mistakes in coding for substitution are made. It was found, for instance, that during the 1992 survey year, 235 substitutions were reported, but in 19 of them the substitution code was changed in a subsequent quarter to a code that indicated no substitution had been required in the first place. Evidently there is some difficulty in making the right decision; otherwise, the coding would not have been changed. The extent to which missed substitutions occur is unknown. It is the purpose of this research study to quantify how many substitutions are missed in the September and December 1993 Agricultural Surveys and how they affect the total expansions.

The other issue which the Working Group to Standardize Survey Procedures raised concerns the effect which missed partial substitutions have on the total expansions. Currently, all cases involving a partial divestiture of an operation result in missed data because the survey procedures do not address this situation. Consequently, even when there is complete compliance with the substitution procedures, the current total expansions are inherently biased, due to missed partial substitutions.

In order to estimate the size of the effects due to missed substitutions, matched records for land in farm between quarters from the 1992 Agricultural Survey were expanded to the state level. Summed across all states, this national total expansion was compared to the corresponding total expansion of land in farm which in expectation was lost due to missed substitutions. The relative difference between these two expansions estimates the relative error in the total expansion of land in farm due to missed substitutions.

Table 1
Substitution and Usable Counts for Land in Farm for 1992 by Month

Month	Number of Usables	Number of Matched Usables	Number of Potential Substitutions	Number of Reported Substitutions	Number of Matched Substitutions
September	64,471	*35,505	*1,157	43	*23
December	68,320	43,015	1,231	60	30
March	65,588	41,775	1,542	132	51

(* used later in equation 4)

APPENDIX III

To qualify for the between quarter match, records in the edited data set had to meet the same usability criteria for land in farm which can be found in the Survey Processing System summary. They had to be in at least two quarters, and they had to have non-zero acreage operated for their first reported occurrence in order to guarantee that the relative divestiture for some operator would not be undefined. Consequently, a record for which an operator reported no land in farm in one quarter was not admitted for analysis in subsequent quarters. Operators that met these conditions are called *restricted matched usables* or matched usables for short. A match may occur between any two quarters or possibly among all three quarters. The number of such records in the edited Agricultural Survey data file are shown in Table 1 by month for the 1992 survey year.

Table 2
Expanded Acres Divested from September 1992 Data

Divestiture Category	Percent Divested	Number of Matched Usables	Number of Expanded Acres a_k	Number of Acres of Divested Land
0	0	18,232	320,843,986	0
1	10	3,309	55,007,238	5,500,724
2	20	1,593	26,724,586	5,344,917
3	30	1,015	18,349,600	5,504,880
4	40	622	13,029,711	5,211,884
5	50	512	14,075,000	7,037,500
6	60	345	6,738,387	4,043,032
7	70	266	4,636,973	3,245,881
8	80	204	4,845,764	3,875,012
9	90	233	5,467,751	4,920,976
10	100	1,157	15,939,477	15,939,477
Number of Matched Divestitures				*9,256
Total Acres of Partial Divestitures from Matched Usables				44,684,806
Total Acres of Divested Land from Matched Usables				60,624,283
Total Acres from Matched Usables				594,347,406
Total U.S. Summary Expansion				928,794,607

(* used later in equation 4)

APPENDIX III

Missed Substitutions

The relative divestiture of an operation is the same as the percentage of the operation which the operator turns over to someone else by way of a sale, lease, or some other transfer where the day to day decisions are made by the new operator. For example, if an operator reported 50 acres in September and 30 acres in December, then that operator divested 20/50=40% of his land and would be placed in divestiture category 4. The divestiture categories provide the index for the summations that are used in the estimators as in (3), for example.

Table 2 shows the expanded land in farm to the national level by each divestiture category. In category 8 for instance, all usable operators that were sampled in June and September operated 4,845,764 acres initially in June and by September operated 970,752 acres for a divestiture of 3,875,012 acres. In reality, however, the amount of divested land is affected by misreported data. Based on observations made in the Kentucky pilot study, actual divestitures are much smaller than the reported values, so that the amount of divested land shown in Table 2 is probably greater than the actual amount and will provide an upper limit on the relative error due to missed partial substitutions for the research study.

When a missed substitution occurs, that information is lost and can not be used in the total expansion. Let

$$X_j = \begin{cases} 1 & \textit{if substitution is done correctly for operator } j \\ & \textit{with probability } 1\text{-}J\epsilon \\ 0 & \textit{if a required substitution is missed for operator } j \\ & \textit{with probability } J\epsilon \end{cases} \qquad (1)$$

J is the probability that a sampled operator is eligible for substitution and ϵ is the probability that a substitution is missed. By means of the indicator variable, X_j, corresponding terms in the total expansion estimator disappear when a missed substitution occurs. Explicitly, the total expansion estimator can be written as:

$$\hat{t} = \sum_{\substack{strata \\ list}} \frac{N_h}{u_h} \sum_{\substack{j \in U_h \\ list}} y_j X_j + \sum_{\substack{strata \\ area}} \sum_{substrata} mexpfctr_{h_k} \sum_{\substack{j \in U_{h_k} \\ area}} y_j X_j \qquad (2)$$

Equation 2 is a Horvitz-Thompson estimator for the total expansion of a commodity, y, that is in a Quarterly Agricultural Survey. The estimator consists of two parts. The first part is associated with the list frame sample of operators; the second part is associated with the area frame sample of operators who are not also elements of the list frame. By construction, the set of sampled operators of the list part and the set of sampled operators of the area part of (2) are disjoint. The quantity, y_j, is a reported but modified value for commodity, y, given by operator, j. The reported responses given by each operator is recorded in subtract levels to which a data adjustment factor determined by his operating arrangement is assigned. The quantity, y_j, of either part of (2) is the sum of the product of the recorded responses by the data adjustment factor over all subtracts for operator, j.

APPENDIX III

In the first part of (2), $\frac{N_h}{u_h}$ is the reciprocal of the product of the probability of selecting a sample unit in stratum h by the corresponding response rate, or equivalently it is the list frame expansion factor for stratum h where N_h is the population size of stratum h and u_h is the size of the set, U_h, of usable sampled operators taken from the list frame and assigned to list stratum h. In the second part of (2), $mexpfctr_{h_k}$ is the area frame expansion factor for substratum k of stratum h and U_{h_k} is the corresponding set of usable sampled operators from the area frame. As noted earlier, this set is by construction disjoint from the set of list frame sampled operators.

The effects of missed substitution occur only when an operator divests some or all of his holdings and the substitution procedures are not properly followed. Under the current operational procedures, two kinds of divestitures are recognized: full divestiture where an operator goes completely out of business, and partial divestiture where an operator turns over part but not all of the operation. To gain a more refined picture of the effects of missed substitutions on the total expansions, partial divestitures were divided into eight additional categories. Every matched operator was therefore assigned to one of ten divestiture categories according to how much land the operator had divested from the previous quarter. Accordingly the terms in (2) were grouped by divestiture category. Then taking the expectation, an estimate of the relative error in the total expansion, given a sample of matched usables, becomes:

$$R_o = \left|\frac{E[\hat{\tau}]-\tau}{\tau}\right| = \frac{J}{\tau}\sum_{k=1}^{9}\frac{k}{10}a_k + \frac{J\epsilon}{\tau}a_{10} \qquad (3)$$

where a_k is the total expansion of land in farm of those operators in the k^{th} divestiture category and τ is the total expansion from the sample of all matched usables. The first term in (3) estimates the relative error that is attributable to missed partial substitutions, and the second term estimates the relative error due to missed full substitutions. The quantities, J and the a_k's can be easily estimated from the data in the edited Agricultural Survey data sets, but ϵ is unknown. We cannot infer what it is because there is nothing in the survey data that allows us to measure the proficiency of a state survey statistician in following the substitution procedures. We can, instead, let ϵ vary over all possible values to give a picture of how the relative error varies as more mistakes are made, that is from complete and perfect compliance with the procedures for an error rate of $\epsilon = 0$ to total disregard of them for $\epsilon = 1$.

The results of the preliminary data analysis are shown in Figures 1 and 2 which are attached to the back of this report. The line labelled *Operational* in Figure 1 represents the relative error due to missed substitutions under the current system, as the probability of missing a substitution, ϵ, varies from 0 to 1. Partial substitutions are immediately lost as there is no provision in the operational procedures to identify them. The line labelled *Proposed*, therefore, represents what the relative error would be if partial substitutions were included as part of the operational procedures. The difference between the two lines represents the contribution to the relative error in the total expansion of land in farm from missing partial substitutions. Above and below these

APPENDIX III

two lines are the upper and lower 99% confidence region limits. To make a comparison between the effect on the land in farm total expansion with another commodity, the effect of missed substitutions on the total U. S. hog expansion was added to make Figure 2.

The information which is missing in these graphs, however, is that point which indicates the relative error which presently exists in the total expansions. A survey statistician who simply flips a coin when deciding whether or not to code for a substitution would be committing an error of $\epsilon=0.5$ which in turn would cause a relative error of 0.38%, according to Figure 1, in the total expansion from land in farm under the current system. In Figure 2, if 10% of the out of business operators require substitute records but are missed, then in expectation a 1.8% relative error will be introduced into the total expansions for land in farm and 1.2% into the hog expansion. Estimating ϵ is one of three principal objectives of the January study. The other two are estimating the rate of over substituting and the number of different parcels of land that a new operator acquires to estimate the effects of multiple contacts.

There exists a one to one correspondence with each matched usable with an enumerator contact, an interview, a report, and a decision by the survey statistician about the correct disposition of that record regarding substitution. From the 1992 survey year, 123,828 records were admitted to the research study analysis of which 104 required a substitute record. As a percentage, substitute records represent 0.08% of the matched usables. What is more interesting is an estimate of the probability that an operator warrants either kind of substitution. If from Tables 1 and 2, we take the proportion of operators who divest some or all of their operation and of those the proportion of reported substitutions to potential cases for a substitute of either kind then we can estimate the probability that an operator requires a substitute record to be:

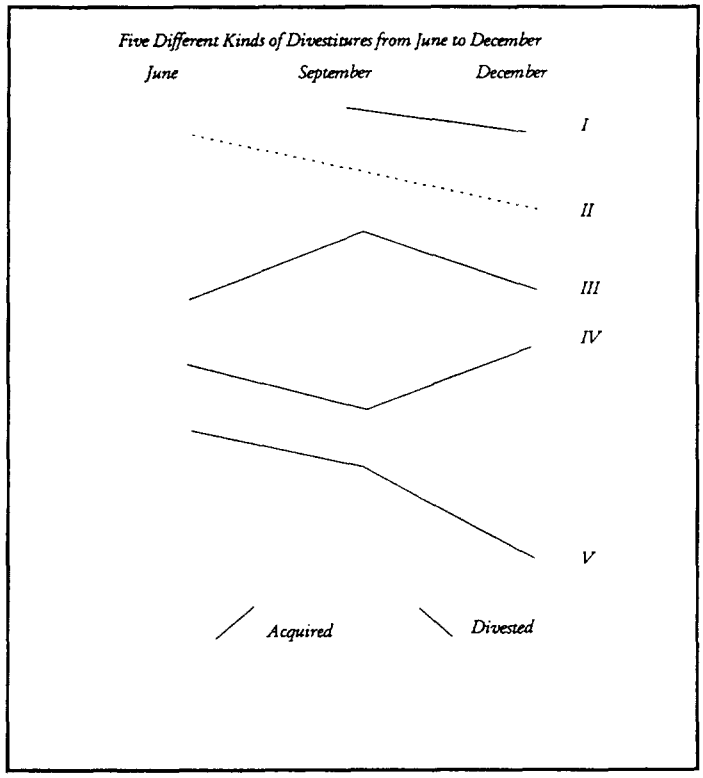
$$(9256/35505)(23/1157)=0.52\% \qquad (4)$$

That is to say, if all the substitutes are missed, then the total expansion will be deficient by about 0.52%.

Preliminary results of the 1992 data analysis indicate that an upper 95% confidence limit for the relative error in the total expansion of land in farm is about 0.7% by the end of the December survey of which that part attributable to missed partial substitutions is at most 0.34% as shown by the upper confidence limit in Figure 1. At $\epsilon=0$, no required substitutions are missed, but because partial substitutions are automatically missed, the relative error at that point is due only to missed partial substitutions. At the other end of the range, where $\epsilon=1$, all potential substitutions are missed, both full and partial substitutions, and the relative error is at a maximum. Because of how the operational procedures address substitution, the relative error is the sum of two parts. The first part in (3) measures the contribution due to missed partial substitutions, and it corresponds to that point in Figure 1 where $\epsilon=0$. At that point where $\epsilon=0$, the second term in (3) disappears, so that the relative error due to missed partial substitution is the product of the likelihood, J, that an out of business operator warrants a substitute record by

APPENDIX III

the proportion of the sum total of acres divested to the total expansion from matched usables, τ. To compute the relative error due to missed partial substitutions for December, is unfortunately a complicated task, because the divestiture of land can occur in five different ways. Schematically as shown on the right, they can occur when an operator who is sampled in September divests by December (I); an operator sampled in June but not in September divests by December(II); an operator sampled in June, September, and December divests between September and December (III) or between June and September (IV) or both (V). For March there are 19 possible ways that divestitures can occur. For simplicity, since divestiture can only occur one way between June and September, we can use September's data to show that the error due to missed partial substitutions is close to 0.3%.



As cited previously in Table 1, there were 23 reported substitutions out of 1157 potential substitution cases in the population of matched reports in September, accordingly, $J=23/1157$ by definition. Assuming that no mistakes had occurred in coding in 1992 for substitution, then $\epsilon=0$ and the last term in (3) disappears. The product of the likelihood that an out of business operator should be substituted, J, and the proportion of acres lost due to missed partial substitutions to total acres of matched usables is:

$$(23/1157)(44684806/594347406)=(0.020)(0.075)=0.0015\approx 0.2\% \qquad (5)$$

From this example, based upon the September data, the relative error due to missed partial substitutions is close to that value of 0.3% which we see in Figure 1 at $\epsilon=0$ for December. In essence, the graphs and the numerical examples show that the relative error due to missed partial substitutions is practically negligible.

Over Substitutions

Mistakes in coding for substitutions do occur. Substitutions may be missed or a substitute record may be made unnecessarily. An unnecessary substitute record is a case of over substitution, and it introduces a positive error into the total expansions. Let ω be the proportion of substitute records that should not have been made to the number of reported substitutions.

APPENDIX III

Conversely, 1-ω represents the proportion of reported substitutions that are valid. The existence of over substitutions adds another dimension to the problem of evaluating the effect on the total expansion due to flaws in the substitution procedures and errors in following them.

The relative error in the total expansion due to over substitutions under the current substitution procedures is:

$$O(\epsilon,\omega) = \omega J(1-\epsilon)\frac{a_{10}}{\tau} \tag{6}$$

where ϵ and J are defined in (1). Referring to (3), τ is the total expansion from all matched usables and a_{10} is the expanded land divested by those operators who have completely gone out of business.

Incorporating the presence of over substitutions, the estimator for the relative error due to missed substitutions formulated in (3) now becomes a function of two terms. The first term corrects the relative error for the over substitutions and the second term is an interaction term between missed substitutions and over substitutions. Together the two terms provide an over all expression for the relative error from missed substitutions:

$$R(\epsilon,\omega) = (1-\omega)R_o + \left(\frac{\epsilon J}{1-J(1-\epsilon)}\right)\frac{\omega}{\tau}\sum_{k=1}^{10} a_k \tag{7}$$

where R_o is the relative error due to missed substitutions in the absence of over substitutions; it coincides with the same quantity that is defined by (3).

Multiple Contacts

The effect of missed substitution and the effect of multiple contacts of a substitute operator on the total expansions are directly related. Missed substitutions produce a negative relative error, $-R(\epsilon,\omega)$, in the total expansion. If a substitute operator is selected more than once, then a positive relative error, $R(\epsilon,\omega)$, is produced each excessive time. If a new operator acquires parcels from three different operators, for example, then he could be selected three times, if each out of business operator qualifies for a substitute record. If this should happen, then his data will be counted two times too much.

The number of times a substitute operator can be contacted is in one to one correspondence with the number of parcels of land that he has acquired from different people. Let λ be the number of parcels that a new operator acquires, then the relative error due to multiple contacts is:

$$M(\epsilon,\omega,\lambda) = (\lambda - 1)R(\epsilon,\omega) \tag{8}$$

Composite Relative Error

From each of the three sources of error mentioned above, the total expansion is affected to differing degrees. Multiple contacts and over substitutions both make positive contributions whereas missed substitutions make a negative contribution to the total expansions. The resultant

APPENDIX III

effect on the total expansions from missed substitutions, over substitutions, and multiple contacts is $T(\epsilon,\omega,\lambda) = -R(\epsilon,\omega) + O(\epsilon,\omega) + M(\epsilon,\omega,\lambda)$ or combining terms:

$$T(\epsilon,\omega,\lambda) = \omega J(1-\epsilon)\frac{a_{10}}{\tau} + (\lambda - 2)R(\epsilon,\omega) \qquad (9)$$

Pilot Study Results

The Kentucky pilot study provided data from which the estimate of $T(\epsilon,\omega,\lambda)$ could be compared with an observed relative error. From that study, it was found that $\lambda = 1.2$. Even though it is a first estimate of the average number of parcels that a new operator acquires, it supports the idea that typically someone will acquire one parcel of land when first entering the agricultural domain.

No missed full substitutions were observed in the course of doing the pilot study, so that $\epsilon = 0$; two missed partial substitutions were observed, and one over substitution out of five reported substitutions was found for $\omega = 1/5$. Upon this information, an estimate of each component of the relative error in the total expansion of land in farm together with the observed relative errors is shown in Table 3. Although the estimates closely agree with the observed relative errors, their sizes indicate, more importantly, that for all practical purposes each component on the average have an insignificant effect on the total expansions. Substantiating what the Kentucky pilot study suggests in general will depend on the outcome of the January research study.

Table 3
Relative Error in the Total Expansion of Land in Farm for Kentucky
September 1993

Component	Estimated error	Observed error
Multiple contacts	+0.001	0
Over substitutions	+0.001	+0.0006
Missed partial substitutions	-0.005	-0.0010
Missed full substitutions	0	0
Total error	-0.003	-0.0004

Sample Size

By using the September and December 1992 Agricultural Surveys as models for anticipating what might happen this year, the size of the research study sample was set and three states were found which collectively reported a population of potential substitutions large enough to accommodate the study sample. Based on the pilot study that was conducted in Kentucky during the last week in September and the outcome of the reconciliation study that was done in December 1993, 282 operators from Kansas, Kentucky, and South Carolina were drawn to form the research study sample. Table 4 shows the research study sample allocation by state.

APPENDIX III

All qualified matched out of business operators who did not refuse an interview in December are in the sample; that includes some operators who had already gone out of business in September together with those who went out of business between September and December. The sample was drawn from both September and December populations. The research study sample size will provide a bound of $\pm 2\%$ on the proportion of missed substitutions to non-substituted out of business operators. The sample from the population of operators who turn over part of their operations was drawn from a sub-population of those operators who divested part of their holdings, reported data in June, September, and December, and had been screened in the reconciliation research study for valid changes in land operated.

Table 4
Sample Size by State for January 1994

State	Number of Potential Substitutions in 1992	Final Sample Size	
		Out of Business	Partially Divested
Kansas	77	56	46
Kentucky	88	55	25
South Carolina	80	50	50

Scope of the Research Study

This research study consists of four primary activities:

1. To find the number of missed substitutions from the sample of matched out of business operators.
2. To find the number of partial substitutions from the sample of matched operators who divest only a part of their operation.
3. To find the number of over substitutions.
4. To find the average number of parcels which a new operator acquires after June 1.

APPENDIX III

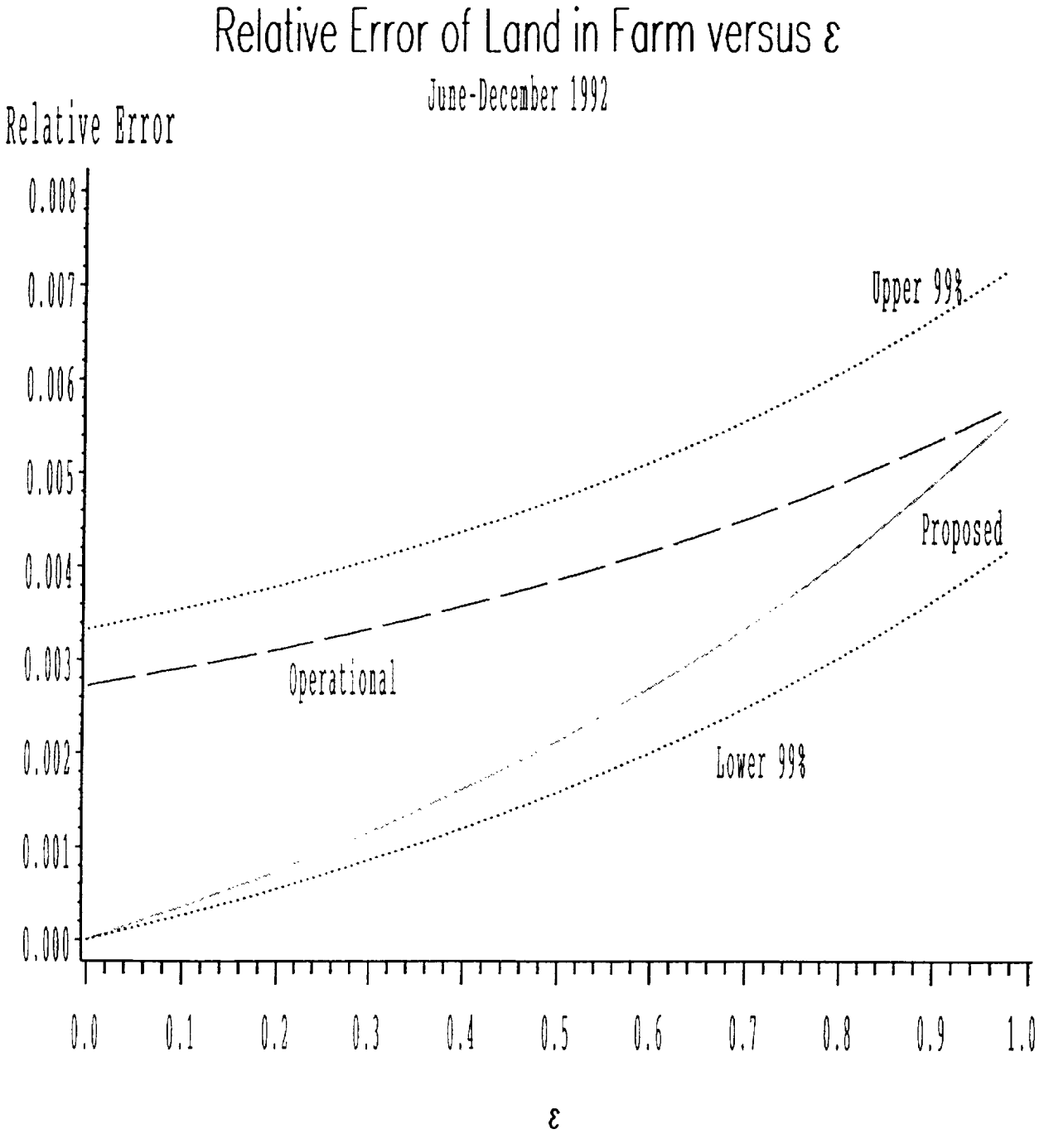


Figure 1

APPENDIX III

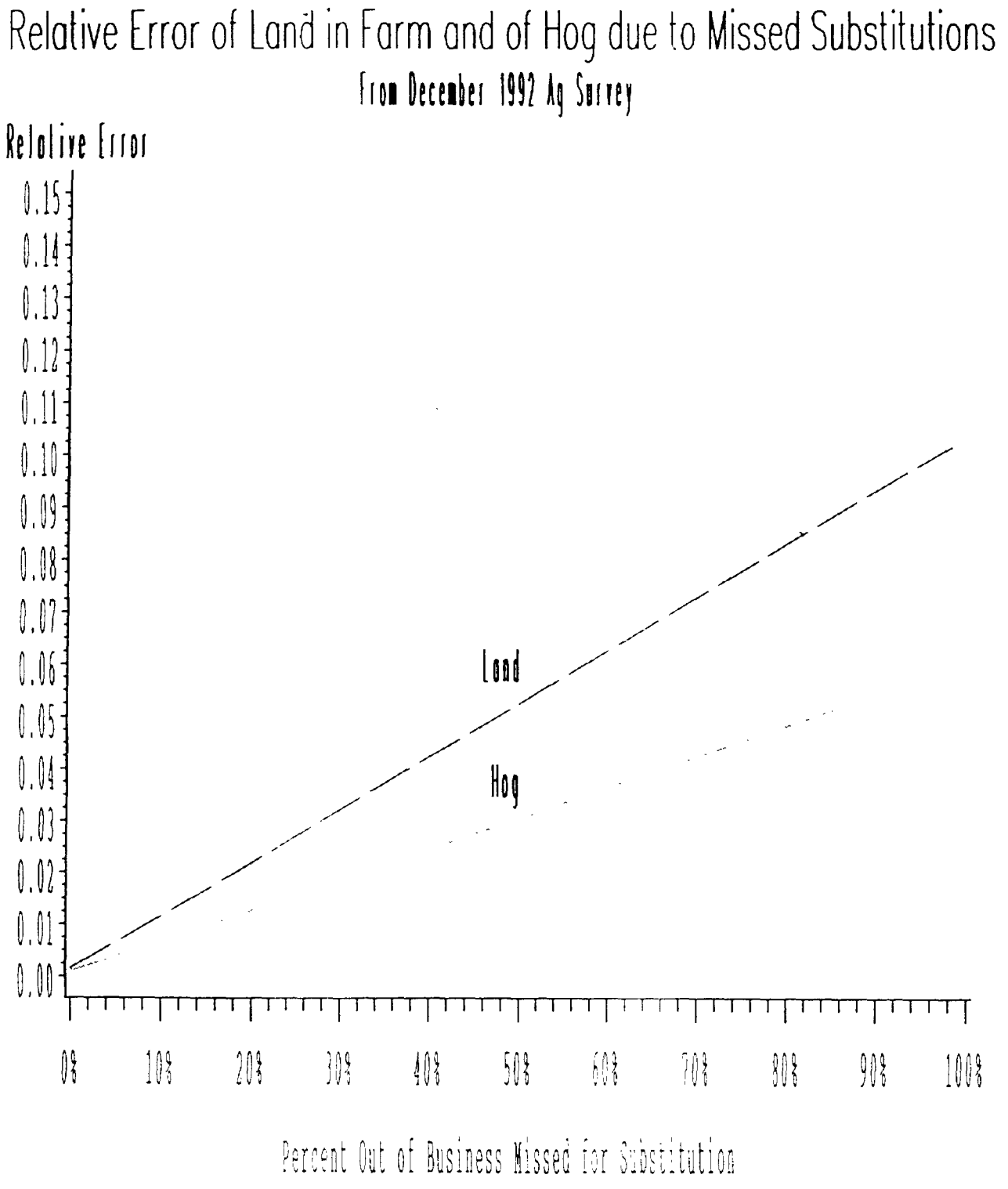


Figure 2

APPENDIX IV

Generalization of R_0 to Include Over Substitutions

The effects of missed substitutions and over substitutions demonstrate the non-linear ramifications that the problems with the substitution procedures have on the estimates. From the derivation of R_0, it can be seen already that missing a substitution has a hyperbolic effect on the relative error. Generalizing R_0 to include over substitutions will enable us to see the rather obscure dependence of the bias on both missed substitutions and over substitutions. As usual, the derivation begins with a lengthy section of definitions and preparatory work, then concludes quickly with the main result.

Let P be the set of operators who have gone out of business. As we know, each element of P is a potential candidate for full substitution. We are assuming that the probabilities of requiring full and partial substitutions are the same. An element of P will qualify to be substituted only if the corresponding new operator cannot be selected for the survey. From P, operators were chosen to be contacted again as a part of the research on substitution. We will call that set S. Let $T \subseteq S$ be the set of operators for which a substitution was made. Of course, some substitutions may have been missed; also, some reported substitutions could have been made unnecessarily. Let $O \subseteq S$ be that set of operators containing all over substitutions.

We will count the number of missed substitutions and over substitutions by using two indicator variables. Define

$$z_i = \begin{cases} 1 & \text{if operator } i \in S' \text{ was missed} \\ 0 & \text{otherwise} \end{cases}$$

$$o_i = \begin{cases} 1 & \text{if operator } i \in T \text{ was an over substitution} \\ 0 & \text{otherwise} \end{cases}$$

where $S' = S-(T-O)$. It is the sample minus reported valid substitutions.

The probability J that a sampled operator for a survey warrants substitution and the probability ϵ that a substitution was missed in a survey can now be expressed in terms of two quantities. They are:

$$\xi = \frac{\sum_{i \in S'} z_i}{|S'|} \qquad (1)$$

$$\omega = \frac{\sum_{i \in T} o_i}{|T|} \qquad (2)$$

where $|G|$ is the cardinality of G. These two statistics measure the proportion of operators who were missed in the sample of operators contacted in conjunction with this research and the proportion of the reported substitutions which were found to have been made unnecessarily.

Let us note that $|O| = \sum_{i \in T} o_i$ and from (2), $|O| = \omega |T|$. Since $|S'| = |S| - |T-O|$ and $|T-O| = (1-\omega)|T|$, we can write $|S| = |S'| + (1-\omega)|T|$. Define $J_o = \frac{|T|}{|S|}$. By definition, J is the quotient of the number of valid substitutions divided by the number of potential substitutions. In a series of steps, we will write J in terms of ξ and ω:

APPENDIX IV

$$J=\frac{\sum_{i\in S'} z_i+|T-O|}{|S|}$$

$$J=\frac{\xi\,|S|+|T-O|}{|S|}$$

$$J=\frac{\xi\,(|S|-|T-O|)+|T-O|}{|S|}$$

$$J=\frac{\xi\,|S|+(1-\xi)(1-\omega)|T|}{|S|}$$

$$J=\xi+(1-\xi)(1-\omega)J_o \qquad (3)$$

An estimate of the probability ϵ of missing a substitution is the quotient of the number of missed substitutions divided by the number of required valid substitutions in the sample S. Again we can follow the same line of reasoning to write ϵ in terms of ξ and ω as follows:

$$\epsilon=\frac{\sum_{i\in S'} z_i}{\sum_{i\in S'} z_i+|T-O|}$$

$$\epsilon=\frac{\xi\,|S|}{\xi\,|S|+(1-\omega)|T|}$$

$$\epsilon=\frac{\xi\,(1-(1-\omega)J_o)}{\xi+(1-\xi)(1-\omega)J_o} \qquad (4)$$

Combining (3) and (4) gives the simple result:

$$J=\frac{(1-\omega)}{(1-\epsilon)}\,J_o \qquad 0\le\epsilon\le 1-(1-\omega)J_o \qquad (5)$$

We will mention parenthetically that if the rate of making over substitutions is the same as the rate of missing substitutions, then $J=J_o$. By coincidence, the rates observed during the research were about the same, so that the special use of the SPS indications was justified in testing the validity of the predicted upper bound at $\epsilon=1-(1-\omega)J_o$.

The relative error in the expansions due to missed substitutions is a function of ϵ and ω. Taking the conditional expectation of $\hat{\tau}$ given a sample of matched records gives:

$$R=\frac{J}{\tau}a_p+\frac{J\epsilon}{\tau}a_f \qquad (6)$$

where

$$\hat{\tau}=\sum_{\substack{h\in strata\\ list}}\frac{N_h}{u_h}\sum_{\substack{j\in U_h\\ list}} y_j X_j+$$

$$\sum_{\substack{h\in strata\\ area}}\sum_{k\in substrata} e_{h_k}\sum_{\substack{j\in U_{h_k}\\ area}} w_j y_j X_j$$

and a_p is the expansion of the divested portion of the commodity from those operators who made a partial divestiture and a_f is the expanion of a commodity from those operators who made a full divestitute. A complete discussion of $\hat{\tau}$ can be found on page 10.

To write $R(\epsilon,\omega)$ in terms of R_o, we will substitute J and ϵ in (6) with (3) and (4) which will give:

$$:,\omega)=(\xi+(1-\xi)(1-\omega)J_o)\frac{a_p}{\tau}+\xi\,(1-(1-\omega)J_o$$

Setting $\omega=0$ gives an expression for R_o:

$$R_o=J_o\frac{a_p}{\tau}+\xi\,(1-J_o)\frac{(a_p+a_f)}{\tau} \qquad (7)$$

Rearranging (4) and simplifying with (3) allows us to write:

$$\xi=\frac{\epsilon J}{1-(1-\epsilon)J} \qquad (8)$$

With some algebraic manipulations and the use of (7) and (8), the main result follows:

$$R(\epsilon,\omega)=(1-\omega)R_o+\frac{\epsilon\,\omega\,J}{1-(1-\epsilon)J}\left(\frac{a_p+a_f}{\tau}\right)$$

$$where\ \ 0\le\epsilon\le 1-(1-\omega)J_o$$

APPENDIX V

Top Ranking in Descending Order of the Relative Error
Due to the Absence of the Substiution Procedures
for All December 1993 SPS Indications at the National Level

Commodity Description	EST[1]	Expansions from Reported Substitutions	SPS Indications	Relative Error (per cent)	Adjusted for Partial Substitutions
STOCKS RICE LONG GRAIN .	I	325329	29741403	1.09	1.37
HOGS UNDER CONTRACT .	R	290998	44417202	0.66	0.94
RICE LONG GRAIN HARV AC .	I	17991	3981820	0.45	0.73
RICE LONG GRAIN PLTD AC .	I	17991	4092379	0.44	0.72
RICE LONG GRAIN PROD .	I	825713	204320551	0.40	0.68
SOYBEANS IRR DBL CROP PROD .	I	243827	62987140	0.39	0.67
GREEN CHOP/ALL HAY PRODUCTION .	I	59412	16339570	0.36	0.64
SOYBEANS IRR DBL CROP HARV AC .	I	6776	2327575	0.29	0.57
SOYBEANS IRR DBL CROP PLTD AC .	I	6776	2348615	0.29	0.57
SOYBEANS N-IRR SING CROP HARV AC	I	21614	7601900	0.28	0.56
SOYBEANS N-IRR SING CROP PLTD AC	I	21614	7864360	0.27	0.55
SOYBEANS N-IRR SING CROP PROD .	I	460861	169904480	0.27	0.55
STOCKS SOYBEANS ALL UNHARV AC .	I	5606	2267553	0.25	0.53
SOYBEANS IRR SING CROP PROD .	I	395061	162315250	0.24	0.52
SOYBEANS N-IRR DBL CROP PROD .	I	143256	58861075	0.24	0.52
SOYBEANS N-IRR DBL CROP HARV AC.	I	5786	2762570	0.21	0.49
HAY GRAIN IRR PRODUCTION .	I	2049	1011272	0.20	0.48
SOYBEANS IRR SING CROP HARV AC .	I	10418	5199705	0.20	0.48
SOYBEANS N-IRR DBL CROP PLTD AC.	I	5786	2910845	0.20	0.48
SOYBEANS IRR SING CROP PLTD AC .	I	10418	5253795	0.20	0.48
WHEAT WINTER IRR SEEDINGS AC .	I	9742	4943664	0.20	0.48
PIGS SOLD OR SLAUGHTERED SEP-NOV	R	35927	18856690	0.19	0.47
HOG DEATHS SEP-NOV .	R	13604	7212658	0.19	0.47
GREEN CHOP/ALL HAY HARV ACRES .	I	5957	3218277	0.19	0.47
MARKET HOGS & PIGS 180+ .	R	114514	65877096	0.17	0.45
SOYBEANS ALL IRR PROD .	I	720667	419249557	0.17	0.45
PEANUTS IRR PLTD AC .	I	1046	623333	0.17	0.45
SOWS FARROWED SEP-NOV .	R	36079	21821732	0.17	0.45
SOWS EXPECT TO FARROW MAR-MAY .	R	37824	22896186	0.17	0.45
CANOLA ALL HARVESTED AC	I	567	345766	0.16	0.44
CANOLA ALL PLANTED AC	I	567	355779	0.16	0.44
HAYLAGE & GREENCHOP PROD .	I	63791	40872817	0.16	0.44
MARKET HOGS & PIGS 60-119 LBS .	R	149774	96159739	0.16	0.44
STOCKS SOYBEANS ALL UNHARV PROD.	I	71968	46477445	0.15	0.43
SOWS, GILTS, & YG GILT FOR BREED	R	75507	49310279	0.15	0.43
SOYBEANS ALL IRR HARV AC .	I	19348	12751382	0.15	0.43
SOYBEANS ALL IRR PLTD AC .	I	19351	12998487	0.15	0.43
RICE MEDIUM GRAIN HARV AC .	I	2230	1532182	0.15	0.43
RICE MEDIUM GRAIN PLTD AC .	I	2230	1548488	0.14	0.42
SOWS EXPECT TO FARROW DEC-FEB .	R	32453	22663487	0.14	0.42
PIGS ON HAND SEP-NOV .	R	219536	157238750	0.14	0.42
●TOTAL HOGS & PIGS .	R	577560	420348395	0.14	0.42
COTTON UPLAND NON-IRR HARV AC .	I	21758	15994887	0.14	0.42
MARKET HOGS & PIGS 120-179 LBS .	R	103447	76849515	0.13	0.41
CORN ALL HI-MOIST SHELL DRYWT PD	I	219314	163604916	0.13	0.41
HAY GRAIN IRR HARV ACRES .	I	513	385253	0.13	0.41
COTTON UPLAND NON-IRR PLTD AC .	I	22522	18010657	0.13	0.41
BOARS & YOUNG MALES FOR BREEDING	R	4547	3699796	0.12	0.40
SUDAN & SORGXSUDAN PLTD AC .	I	3105	2607075	0.12	0.40
RICE MEDIUM GRAIN PROD .	I	115392	99983936	0.12	0.40
SUDAN & SORGXSUDAN N-IRR PLTD AC	I	2293	2015891	0.11	0.39
PEANUTS ALL PLTD AC .	I	3913	3467883	0.11	0.39
COTTON UPLAND NON-IRR PROD .	I	14821	13596567	0.11	0.39
WHEAT WINTER NON-IRR SEEDINGS .	I	18408	17933364	0.10	0.38

APPENDIX V

Commodity Description	EST	Expansions from Reported Substitutions	SPS Indications	Relative Error (per cent)	Adjusted for Partial Substitutions
MARKET HOGS & PIGS UNDER 60 LBS.	R	129176	127086205	0.10	0.38
PEANUTS ALL PROD .	I	6594519	6985283385	0.09	0.37
SOYBEANS ALL N-IRR HARV AC .	I	31989	34395337	0.09	0.37
SOYBEANS DBL CROP PROD .	I	413841	450449324	0.09	0.37
CORN ALL HI-MOIST SHELLED HV AC.	I	1512	1650228	0.09	0.37
SOYBEANS ALL N-IRR PLTD AC .	I	32027	35749050	0.09	0.37
HAYLAGE & GREENCHOP HARV AC .	I	6794	7619304	0.09	0.37
SOYBEANS ALL N-IRR PROD .	I	782374	888438241	0.09	0.37
PEANUTS ALL HARV AC .	I	2834	3294528	0.09	0.37
●LAND MINUS PIGA	R	1656558	2030471676	0.08	0.36
HAY ALFALFA ALL PROD .	I	128254	159101361	0.08	0.36
HAY OTHER (EXCL GRAIN) ALL PROD.	I	111007	137766590	0.08	0.36
LAND ALL IRR AC .	I	11157	14016134	0.08	0.36
HAY ALL PRODUCTION .	I	248088	311845878	0.08	0.36
CORN NON-IRR OTHER USES AC .	I	2217	2796172	0.08	0.36
HAY OTHER (EXCL GRAIN) ALL HV AC	I	47228	61404696	0.08	0.36
SOYBEANS DBL CROP HARV AC .	I	13519	17933278	0.08	0.36
SOYBEANS DBL CROP PLTD AC .	I	13529	18898779	0.07	0.35
COTTON UPLAND ALL HARV AC .	I	26287	37021123	0.07	0.35
CORN ALL DRY GRAIN PROD .	I	540023	771655754	0.07	0.35
COTTON UPLAND ALL PLTD AC .	I	27420	39398172	0.07	0.35
HAY ALFALFA IRR PROD .	I	24891	35954268	0.07	0.35
OATS NON-IRR SEEDINGS .	I	1004	1456854	0.07	0.35
STOCKS HAY .	I	135869	200418279	0.07	0.35
CORN ALL DRY GRAIN HARV AC .	I	5034	7588658	0.07	0.35
CROPLAND AC .	I	540254	821511656	0.07	0.35
HAY ALFALFA IRR HARV AC .	I	5665	8682735	0.07	0.35
HAY ALFALFA ALL HARV AC .	I	31620	49486148	0.06	0.34
HAY GRAIN ALL PRODUCTION .	I	7473	11719412	0.06	0.34
CORN NON-IRR GRAIN PROD .	I	2027340	3434868152	0.06	0.34
STOCKS OATS ALL .	I	215807	368896395	0.06	0.34
HAY OTHER EXCL GRAIN N-IRR HV AC	I	5945	10415322	0.06	0.34
HAY GRAIN ALL HARV AC .	I	3070	5564758	0.06	0.34
TOBACCO FLUE CURED PROD .	I	1064877	2010546847	0.05	0.33
HAY ALFALFA NON-IRR PROD .	I	8964	16935797	0.05	0.33
CORN NON-IRR GRAIN HARV AC .	I	24608	47957304	0.05	0.33
COTTON UPLAND ALL PROD .	I	23016	45357215	0.05	0.33
HAY GRAIN NON IRR HARV ACRES .	I	572	1135159	0.05	0.33
CORN NON-IRR PLTD AC .	I	28885	57839324	0.05	0.33
STOCKS FARM GRAIN CAPACITY .	I	10805241	22790037749	0.05	0.33
CORN IRR GRAIN PROD .	I	2935780	6220488894	0.05	0.33
HAY WILD N-IRR HARV AC .	I	1203	2564314	0.05	0.33
CORN IRR GRAIN HARV AC .	I	24066	51336953	0.05	0.33
WHEAT WINTER ALL FALL SEEDED AC.	I	85643	183201945	0.05	0.33
STOCKS BARLEY .	I	271748	584432417	0.05	0.33
SORGHUM ALL OTHER USES AC .	I	1203	2643073	0.05	0.33
CORN IRR PLTD AC .	I	24769	55482372	0.04	0.32
TOBACCO FLUE CURED HARV AC .	I	402	918090	0.04	0.32
HAY OTHER EXCL GRAIN N-IRR PROD.	I	9293	21238298	0.04	0.32
HAY ALFALFA NON-IRR HARV AC .	I	2872	6579303	0.04	0.32
SOWS & BOARS NOT FOR BREEDING .	R	596	1365771	0.04	0.32
CORN IRR OTHER USES AC .	I	403	924840	0.04	0.32
HAY WILD N-IRR PROD .	I	1354	3138246	0.04	0.32
SOYBEANS ALL SING CROP HARV AC .	I	35241	89367500	0.04	0.32
SOYBEANS ALL SING CROP PLTD .	I	35243	94716741	0.04	0.32
COTTON UPLAND IRR PLTD AC .	I	3247	9728225	0.03	0.31
GREEN CHOP ALFALFA HARV AC .	I	745	2240125	0.03	0.31
SORGHUM ALL PLTD AC .	I	15571	48602412	0.03	0.31
SORGHUM ALL GRAIN HARV AC .	I	14285	44669662	0.03	0.31

APPENDIX V

Commodity Description	EST	Expansions from Reported Substitutions	SPS Indications	Relative Error (per cent)	Adjusted for Partial Substitutions
CORN NON-IRR SILAGE HARV AC	I	2061	6474619	0.03	0.31
TOBACCO BURLEY HARV AC	I	202	642868	0.03	0.31
TOBACCO BURLEY PROD	I	423604	1385139563	0.03	0.31
GREEN CHOP ALFALFA PROD.	I	4068	13375893	0.03	0.31
COTTON UPLAND IRR HARV AC	I	2880	9615915	0.03	0.31
COTTON UPLAND IRR PROD	I	4088	13776120	0.03	0.31
SOYBEANS ALL SING CROP PROD	I	950760	3227236511	0.03	0.31
SORGHUM NON-IRR GRAIN HARV AC	I	9086	34457471	0.03	0.31
CORN ALL SILAGE HARV AC	I	10226	38862496	0.03	0.31
CORN ALL SILAGE PROD	I	117677	457010783	0.03	0.31
SORGHUM NON-IRR PLTD AC	I	9159	36683270	0.02	0.30
HOGS&PIGS BUTCH ON LAND OPER	R	88	366745	0.02	0.30
HAY GRAIN NON-IRR PRODUCTION	I	522	2182942	0.02	0.30
SORGHUM ALL GRAIN PROD	I	633945	2682039868	0.02	0.30
CHICK EGG ROOSTERS & MALES	R	395	1694072	0.02	0.30
SOYBEANS ALL BEANS HARV AC	I	62741	274289190	0.02	0.30
SOYBEANS ALL PLTD AC	I	63056	285902084	0.02	0.30
CORN ALL OTHER USES AC	I	3926	18196974	0.02	0.30
SORGHUM NON-IRR GRAIN PROD	I	414941	1951955744	0.02	0.30
STOCKS WHEAT ALL	I	388459	1858012420	0.02	0.30
CORN NON-IRR SILAGE PROD	I	8467	40666459	0.02	0.30
SOYBEANS ALL PROD	I	1857274	8976763314	0.02	0.30
STOCKS WHEAT DURUM ALL	I	19676	99270589	0.02	0.30
CORN ALL PLTD AC	I	83950	441152178	0.02	0.30
STOCKS SORGHUM	I	78682	419814354	0.02	0.30
CORN ALL GRAIN PROD	I	7111825	38363798709	0.02	0.30
CANOLA ALL PROD	I	84984	461995906	0.02	0.30
CORN ALL GRAIN HARV AC	I	69798	383468854	0.02	0.30
STOCKS WHEAT SPRING	I	222321	1232834723	0.02	0.30
●STOCKS CORN DRY GRAIN	I	25335	141681966	0.02	0.30
SORGHUM NON-IRR SILAGE PROD	I	1049	5883721	0.02	0.30
SAFFLOWER ALL PROD	I	133492	880246225	0.02	0.30
SUGARBEETS ALL PROD	I	7198	49165795	0.01	0.29
STOCKS WHEAT WINTER	I	47436	346322458	0.01	0.29
SAFFLOWER ALL HARVESTED AC	I	63	469298	0.01	0.29
STOCKS CORN ALL UNHARV CALC PROD	I	60948	459051216	0.01	0.29
SORGHUM NON-IRR SILAGE HARV AC	I	73	593429	0.01	0.29
STOCKS CORN ALL UNHARV AC	I	666	5508335	0.01	0.29
RYE ALL SEEDINGS	I	673	5946383	0.01	0.29
SAFFLOWER ALL PLANTED AC	I	77	690170	0.01	0.29
SUGARBEETS ALL PLTD AC	I	278	2717798	0.01	0.29
SUGARBEETS ALL HARV AC	I	273	2675350	0.01	0.29
GREEN CHOP ALL OTHER HARV AC	I	83	829214	0.01	0.29

[1]EST denotes the kind of estimator used to compute the SPS indicdations:
I=Imputed Modified + Weighted NOL
R=Reweighted + Weighted NOL

A ● identifies one of the three commodities, land in farm, total hogs and pigs, and corn stocks that was chosen to represnt a typical commodity of every farm.

Only those commodities (156 out of 327) that had a relative error greater than 0.01% before adjusting for the effects of any missed partial substitutions are included in this list. To adjust for the effects of missed partial substitutions, since the SPS indications do not account for them, the predicted relative error of 0.28% from missed partial substitutions was added to the relative error computed from the SPS indications.

APPENDIX VI

Top Ranking in Descending Order of the Relative Error
Due to the Absence of the Substiution Procedures
for Selected December 1993 SPS Indications at the State Level

ST	Commodity Description	EST[1]	Expansions from Reported Substitutions	SPS Indications	Relative Error (per cent)	Adjusted fo Partial Substitutio
AR	STOCKS SOYBEANS ALL UNHARV AC .	I	4920	65315	7.53	7.81
NJ	SOYBEANS ALL OTHER USES AC .	I	109	1465	7.45	7.73
AR	STOCKS SOYBEANS ALL UNHARV PROD.	I	52090	916779	5.68	5.96
KS	SOWS, GILTS, & YG GILT FOR BREED	A	8439	153004	5.52	5.80
✖SD	TOTAL HOGS & PIGS .	A	98708	1849269	5.34	5.62
OK	BOARS & YOUNG MALES FOR BREEDING	A	185	3562	5.19	5.47
SD	SOWS, GILTS, & YG GILT FOR BREED	A	10911	211997	5.15	5.43
●KS	TOTAL HOGS & PIGS .	A	67853	1327468	5.11	5.39
SC	CORN ALL SILAGE PROD .	I	3886	86138	4.51	4.79
KS	BOARS & YOUNG MALES FOR BREEDING	A	450	10491	4.29	4.57
SD	BOARS & YOUNG MALES FOR BREEDING	A	509	13470	3.78	4.06
KY	STOCKS SOYBEANS ALL UNHARV PROD.	I	16425	475749	3.45	3.73
NE	BOARS & YOUNG MALES FOR BREEDING	A	995	28906	3.44	3.72
OK	SOWS, GILTS, & YG GILT FOR BREED	A	1848	53928	3.43	3.71
SD	CORN IRR GRAIN PROD .	I	495484	14805225	3.35	3.63
IL	SOWS, GILTS, & YG GILT FOR BREED	A	27827	854097	3.26	3.54
✖NE	TOTAL HOGS & PIGS .	A	131325	4277061	3.07	3.35
NY	CORN ALL OTHER USES AC .	I	319	10419	3.07	3.35
SD	CORN IRR GRAIN HARV AC .	I	4317	143782	3.00	3.28
✖IL	TOTAL HOGS & PIGS .	A	154442	5338150	2.89	3.17
NM	COTTON UPLAND IRR PROD .	I	2370	88221	2.69	2.97
NM	COTTON UPLAND ALL PROD .	I	2370	88221	2.69	2.97
PA	SOWS, GILTS, & YG GILT FOR BREED	A	2385	89168	2.67	2.95
SD	CORN IRR PLTD AC .	I	4582	173309	2.64	2.92
KY	STOCKS SOYBEANS ALL UNHARV AC .	I	548	21315	2.57	2.85
WI	SOWS, GILTS, & YG GILT FOR BREED	A	3486	150156	2.32	2.60
NE	SOWS, GILTS, & YG GILT FOR BREED	A	10944	472072	2.32	2.60
NY	BOARS & YOUNG MALES FOR BREEDING	A	25	1076	2.31	2.59
NM	COTTON UPLAND IRR HARV AC .	I	1161	50904	2.28	2.56
NM	COTTON UPLAND ALL HARV AC .	I	1161	50904	2.28	2.56
UT	SOWS, GILTS, & YG GILT FOR BREED	A	102	4639	2.19	2.47
NM	COTTON UPLAND IRR PLTD AC .	I	1161	53102	2.19	2.47
SC	CORN ALL SILAGE HARV AC .	I	268	12409	2.16	2.44
NM	COTTON UPLAND ALL PLTD AC .	I	1161	53977	2.15	2.43
✖WI	TOTAL HOGS & PIGS .	A	24053	1189388	2.02	2.30
AR	SOYBEANS IRR DBL CROP PROD .	I	243827	12597428	1.94	2.22
●SC	STOCKS CORN ALL .	I	71250	3737773	1.91	2.19
WI	BOARS & YOUNG MALES FOR BREEDING	A	232	13805	1.68	1.96
✖PA	TOTAL HOGS & PIGS .	A	22260	1327662	1.68	1.96
NY	SOWS, GILTS, & YG GILT FOR BREED	A	166	9987	1.66	1.94
AR	WHEAT WINTER ALL FALL SEEDED AC.	I	17939	1096076	1.64	1.92
AR	SOYBEANS DBL CROP PROD .	I	387083	24369644	1.59	1.87
NY	CORN ALL HI-MOIST SHELL DRYWT PD	I	218522	13973881	1.56	1.84
NJ	CORN ALL SILAGE HARV AC .	I	300	19732	1.52	1.80
NM	BOARS & YOUNG MALES FOR BREEDING	A	5	362	1.48	1.76
UT	WHEAT WINTER ALL FALL SEEDED AC.	I	2132	145117	1.47	1.75
PA	BOARS & YOUNG MALES FOR BREEDING	A	119	8129	1.47	1.75
AR	SOYBEANS IRR DBL CROP HARV AC .	I	6776	465515	1.46	1.74
AR	SOYBEANS IRR DBL CROP PLTD AC .	I	6776	469723	1.44	1.72
MD	CORN NON-IRR SILAGE HARV AC .	I	1100	76825	1.43	1.71
MD	CORN ALL SILAGE HARV AC .	I	1100	76975	1.43	1.71
AR	SOYBEANS N-IRR SING CROP HARV AC	I	21614	1520380	1.42	1.70
✖NY	TOTAL HOGS & PIGS .	A	1191	83924	1.42	1.70
AR	SOYBEANS ALL IRR PROD .	I	638888	45060478	1.42	1.70
AR	SOYBEANS N-IRR SING CROP PLTD AC	I	21614	1572872	1.37	1.65

APPENDIX VI

ST	Commodity Description	EST	Expansions from Reported Substitutions	SPS Indications	Relative Error (per cent)	Adjusted for Partial Substitutions
AR	SOYBEANS ALL PROD	I	1243005	90813589	1.37	1.65
AR	SOYBEANS N-IRR SING CROP PROD	I	460861	33980896	1.36	1.64
NY	CORN ALL GRAIN PROD	I	754546	56869037	1.33	1.61
AR	SOYBEANS ALL N-IRR HARV AC	I	27400	2072894	1.32	1.60
NY	CORN ALL DRY GRAIN HARV AC	I	4993	377890	1.32	1.60
AR	SOYBEANS ALL N-IRR PROD	I	604117	45753111	1.32	1.60
NY	CORN ALL DRY GRAIN PROD	I	536024	40747894	1.32	1.60
AR	SOYBEANS ALL SING CROP PROD	I	855922	66443945	1.29	1.57
AR	SOYBEANS ALL N-IRR PLTD AC	I	27400	2155041	1.27	1.55
AL	CORN ALL SILAGE HARV AC	I	322	25374	1.27	1.55
NC	CORN ALL OTHER USES AC	I	898	71196	1.26	1.54
AR	SOYBEANS ALL SING CROP HARV AC	I	32032	2560321	1.25	1.53
AR	SOYBEANS ALL BEANS HARV AC	I	44595	3578350	1.25	1.53
NM	CORN ALL SILAGE PROD	I	6800	549509	1.24	1.52
AR	SOYBEANS DBL CROP HARV AC	I	12563	1018029	1.23	1.51
AR	SOYBEANS ALL SING CROP PLTD	I	32032	2623631	1.22	1.50
AR	SOYBEANS IRR SING CROP PROD	I	395061	32463050	1.22	1.50
AR	SOYBEANS N-IRR DBL CROP PROD	I	143256	11772215	1.22	1.50
NY	CORN ALL GRAIN HARV AC	I	6497	535141	1.21	1.49
AR	SOYBEANS ALL PLTD AC	I	44595	3675523	1.21	1.49
AR	SOYBEANS DBL CROP PLTD AC	I	12563	1051892	1.19	1.47
SD	CORN IRR OTHER USES AC	I	235	20034	1.17	1.45
IL	BOARS & YOUNG MALES FOR BREEDING	A	1391	120306	1.16	1.44
✖OK	TOTAL HOGS & PIGS	A	3433	297670	1.15	1.43
AR	SOYBEANS ALL IRR HARV AC	I	17195	1505456	1.14	1.42
NY	CORN ALL HI-MOIST SHELLED HV AC	I	1504	132002	1.14	1.42
NJ	CORN ALL SILAGE PROD	I	3000	264198	1.14	1.42
AR	SOYBEANS ALL IRR PLTD AC	I	17195	1520482	1.13	1.41
NM	CORN ALL SILAGE HARV AC	I	340	30116	1.13	1.41
✖OR	TOTAL HOGS & PIGS	A	443	39454	1.12	1.40
TX	CORN NON-IRR GRAIN PROD	I	624815	57395549	1.09	1.37
MI	CORN ALL SILAGE PROD	I	31856	3014648	1.06	1.34
AR	SOYBEANS N-IRR DBL CROP HARV AC	I	5786	552514	1.05	1.33
WI	SOYBEANS ALL OTHER USES AC	I	153	14678	1.04	1.32
AR	STOCKS SOYBEANS	I	68223	6629427	1.03	1.31
AR	SOYBEANS IRR SING CROP HARV AC	I	10418	1039941	1.00	1.28
TX	CORN NON-IRR GRAIN HARV AC	I	8443	849357	0.99	1.27
AR	SOYBEANS N-IRR DBL CROP PLTD AC	I	5786	582169	0.99	1.27
AR	SOYBEANS IRR SING CROP PLTD AC	I	10418	1050759	0.99	1.27
NJ	CORN ALL PLTD AC	I	966	102468	0.94	1.22
TX	CORN NON-IRR PLTD AC	I	8443	904937	0.93	1.21
TX	CORN ALL GRAIN HARV AC	I	16887	1833305	0.92	1.20
SD	CORN ALL OTHER USES AC	I	2246	244222	0.92	1.20
TX	WHEAT WINTER IRR SEEDINGS AC	I	9742	1070306	0.91	1.19
TX	CORN ALL GRAIN PROD	I	1922495	211216246	0.91	1.19
SD	CORN NON-IRR OTHER USES AC	I	2011	224188	0.90	1.18
✖NJ	LAND MINUS PIGA	R	7929	901038	0.88	1.16
MI	CORN ALL SILAGE HARV AC	I	2210	256433	0.86	1.14
TX	CORN ALL PLTD AC	I	16887	1963707	0.86	1.14
TX	CORN IRR GRAIN HARV AC	I	8443	983948	0.86	1.14
TX	CORN IRR GRAIN PROD	I	1297680	153820697	0.84	1.12
NJ	CORN ALL GRAIN HARV AC	I	666	80437	0.83	1.11
OK	SOYBEANS ALL PROD	I	39000	4781710	0.82	1.10
TX	CORN IRR PLTD AC	I	8443	1058770	0.80	1.08
ME	CORN ALL OTHER USES AC	I	1	89	0.79	1.07
NY	CORN ALL PLTD AC	I	8320	1066204	0.78	1.06
OK	SOYBEANS ALL BEANS HARV AC	I	1500	202444	0.74	1.02
✖MT	LAND MINUS PIGA	R	412177	56727750	0.73	1.01
✖TX	TOTAL HOGS & PIGS	A	3275	465001	0.70	0.98

APPENDIX VI

ST	Commodity Description	EST	Expansions from Reported Substitutions	SPS Indications	Relative Error (per cent)	Adjusted fo Partial Substitutio
SD	CORN ALL GRAIN PROD .	I	1134416	161835898	0.70	0.98
TX	COTTON UPLAND NON-IRR HARV AC .	I	21758	3185921	0.68	0.96
OK	SOYBEANS ALL PLTD AC .	I	1500	220791	0.68	0.96
SC	STOCKS SOYBEANS .	I	15900	2364304	0.67	0.95
NY	WHEAT WINTER ALL FALL SEEDED AC.	I	729	108511	0.67	0.95
●SC	TOTAL HOGS & PIGS .	A	2505	377185	0.66	0.94
TX	COTTON UPLAND NON-IRR PROD .	I	14821	2234093	0.66	0.94
NJ	CORN ALL GRAIN PROD .	I	48649	7681270	0.63	0.91
TX	COTTON UPLAND NON-IRR PLTD AC .	I	22522	3674083	0.61	0.89
✖UT	LAND MINUS PIGA .	R	48980	8026827	0.61	0.89
NJ	SOYBEANS ALL PLTD AC .	I	1021	170300	0.60	0.88
OR	SOWS, GILTS, & YG GILT FOR BREED	A	32	5289	0.60	0.88
AR	CORN ALL GRAIN PROD .	I	37943	6469059	0.59	0.87
✖SD	STOCKS CORN ALL .	I	824670	143592537	0.57	0.85
SD	CORN IRR SILAGE PROD .	I	433	77818	0.56	0.84
SC	BOARS & YOUNG MALES FOR BREEDING	A	22	3979	0.55	0.83
✖AR	LAND MINUS PIGA .	R	92525	16848031	0.55	0.83
SC	SOWS, GILTS, & YG GILT FOR BREED	A	250	45641	0.55	0.83
NJ	SOYBEANS ALL BEANS HARV AC .	I	912	168836	0.54	0.82
TX	WHEAT WINTER ALL FALL SEEDED AC.	I	28151	5287484	0.53	0.81
AR	CORN ALL GRAIN HARV AC .	I	412	79238	0.52	0.80
NJ	SOYBEANS ALL PROD .	I	25013	4970199	0.50	0.78
SD	CORN ALL GRAIN HARV AC .	I	12740	2553479	0.50	0.78
AR	CORN ALL PLTD AC .	I	412	84823	0.49	0.77
SD	CORN ALL PLTD AC .	I	15547	3271584	0.48	0.76
SD	STOCKS SOYBEANS .	I	78066	16524501	0.47	0.75
TX	WHEAT WINTER NON-IRR SEEDINGS .	I	18408	4217178	0.44	0.72
SD	CORN NON-IRR GRAIN PROD .	I	638932	147030673	0.43	0.71
VA	CORN ALL GRAIN PROD .	I	78799	18174008	0.43	0.71
TX	COTTON UPLAND ALL HARV AC .	I	21758	5048739	0.43	0.71

[1]EST denotes the kind of estimator used to compute the SPS indications:
A=Adjusted List + Weighted NOL
I=Imputed Modified + Weighted NOL
R=Reweighted + Weighted NOL

A ● identifies a commodity and state which were studied as representative commodities of a farming operati by contacting operators in Kansas, Kentucky, and South Carolina.

An ✖ identifies the same commodities but associated with different states.

The selected commodities listed in this appendix are:
CORN
COTTON
HOGS
LAND IN FARM
SOYBEANS
WINTER WHEAT

The top 144 commodity and state combinations from a possible 4,648 total combinations are shown.