

## Technical Notes: Extending Primary Suppression Rules to Other Common Situations

This appendix contains practices the statistical agencies have found useful when applying disclosure limitation to tables in common situations. The primary and complementary suppression procedures for tables of magnitude data discussed in Chapter IV are based on the assumption that the reported data are strictly positive, and that the published number is the simple sum of the data from all respondents. In some situations published data are not simple sums, and it is not clear how to apply primary and complementary suppression methodology. For example, in this appendix we extend primary suppression rules used for tabular data to tables containing imputed data.

Further, the methods discussed in this paper are implicitly to be applied to every published variable. In practice, simplifying assumptions have been made to reduce the workload associated with disclosure limitation and to improve the consistency of published tables over time.

Section 2 presents the disclosure limitation practices which have been used where there may be some question as to how to apply the standard procedures. Section 3 presents the simplifying assumptions which have been found useful by federal statistical agencies. Both sections are intended as a reference for other agencies facing similar situations.

### 1. Background

The (n,k), pq-ambiguity and p-percent rules described in Chapter IV can all be written in the following form:

$$S(X) = \sum_{i=1}^n x_i - c(T - \sum_{i=1}^s x_i).$$

where the values of n, c and s depend on the specific rule and the parameters chosen, T is the total to be published,  $x_1$  is the largest reported value,  $x_2$  is the second largest reported value, and so on. In this framework, the  $x_i$  are all nonnegative.

### 2. Extension of Disclosure Limitation Practices

#### 2.a. Sample Survey Data

The equation above assumes that all data are reported (as in a census). How can this rule be applied to data from a sample survey? One way of handling this is to let the values of the largest

respondents, the  $x_i$ , be specified by the unweighted reported values, but to let T be the weighted total to be published. (Note: this is a consistent way of stating that there is no disclosure with data from a sample survey when no units are selected with certainty and the sampling fractions are small.)

## **2.b. Tables Containing Imputed Data**

If some data are imputed, disclosure potential depends on the method of imputation.

- a) Imputation for a sample survey is done by adjusting weights: In this case, method 2.a applies (the adjusted weights are used to calculate the weighted total, T).
- b) Imputed values may be based on other respondent's data, as in "hot decking": In this case, the imputed value should not constitute a disclosure about the nonrespondent, so the imputed value (weighted, if appropriate) is included in the estimated total, T. The imputed value is counted as an individual reported value for purposes of identifying the largest respondents only for the donor respondent.
- c) Imputed values may be based on past data from the nonrespondent: If the imputed value were revealed, it could constitute disclosure about the nonrespondent (for example, if the imputed value is based on data submitted by the same respondent in a different time period). The imputed value is included in the estimated total, T, and is also treated as submitted data for purposes of identifying the largest respondents.

## **2.c. Tables that Report Negative Values**

If all reported values are negative, suppression rules can be applied directly by taking the absolute value of the reported data.

## **2.d. Tables Where Differences Between Positive Values are Reported**

If the published item is the difference between two positive quantities reported for the same time period (e.g. net production equals gross production minus inputs), then apply the primary suppression rule as follows:

- a) If the resultant difference is generally positive, apply the suppression procedure to the first item (gross production in the above example).
- b) If the resultant difference is generally negative, apply the suppression procedure to the second item (inputs in the above example.)
- c) If the resultant difference can be either positive or negative and is not dominated by either, there are two approaches. One method is to set a threshold for the minimum number of respondents in a cell. A very conservative approach is to take the absolute value of the difference before applying the primary suppression rule.

## **2.e. Tables Reporting Net Changes (that is, Difference Between Values Reported at Different Times)**

If either of the values used to calculate net change were suppressed in the original publication, then net change must also be suppressed.

## **2.f. Tables Reporting Weighted Averages**

If a published item is the weighted average of two positive reported quantities, such as volume weighted price, apply the suppression procedure to the weighting variable (volume in this example).

## **2.g. Output from Statistical Models**

Output from statistical models, such as econometric equations estimated using confidential data, may pose a disclosure risk. Often the resulting output from the statistical analyses takes the form of parameter coefficients in various types of regression equations or systems of equations. Since it is only possible to exactly recover input data from a regression equation if the number of coefficients is equal to the number of observations, regression output generally poses no disclosure risk. However, sometimes dummy (0,1) variables are used in the model to capture certain effects, and these dummy variables may take on values for only a small number of observations.

One way of handling this situation is provided by the Center for Economic Studies of the Census Bureau. They treat the dummy variables as though they were cells in a table. Using the (n,k) rule, disclosure analysis is performed on the observations for which the dummy variable takes on the value 1.

## **3. Simplifying Procedures**

### **3.a. Key Item Suppression**

In several economic censuses, the Census Bureau employs key item suppression: performing primary disclosure analysis and complementary suppression on certain key data items only, and applying the same suppression pattern to other related items. Under key item suppression, fewer agency resources are devoted to disclosure limitation and data products are more uniform across data items. Key and related items are identified by expert judgment. They should remain stable over time.

### **3.b. Preliminary and Final Data**

For magnitude data released in both preliminary and final form, the suppression pattern identified and used for the preliminary data should be carried forward to the final publication. The final data tables are then subjected to an audit to assure that there are no new disclosures. This conservative approach reduces the risk that a third party will identify a respondent's data from the changes in suppression patterns between preliminary and final publication.

### **3.c. Time Series Data**

For routine monthly or quarterly publications of magnitude data, a standard suppression pattern (primary and complementary) can be developed based on the previous year's monthly data. This suppression pattern, after auditing to assure no new disclosures, would be used in the regular monthly publication.