APPENDIX F

APPENDIX F

CHI-SQUARED AUTOMATIC INTERACTION DETECTOR

Chi-squared Automatic Interaction Detector *(CHAID)* is a highly efficient statistical technique for segmentation, or tree growing, developed by Kass (1980). The analysis in *CHAID* begins by dividing the population into two or more groups based on the categories of the "best" predictor of a dependent variable. It merges values that are judged to be statistically homogeneous (similar) with respect to the target variable and maintains all other values that are heterogeneous (dissimilar). Each of these groups is then divided into smaller subgroups based on the best available predictor at each level. The splitting process continues recursively until no more statistically significant predictors can be found (or until some other stopping rule is met). The *CHAID* software displays the final subgroups (segments) in the form of a tree diagram whose branches (nodes) correspond to the groups. The segments that *CHAID* derives are mutually exclusive and exhaustive. It also produces a file of associated pseudocode that can be used in SAS®, with minor modifications, to create a SAS® variable for indicating the groups (i.e., the nonresponse adjustment cells).

A node will *not* be split if any of the following conditions is met:

- All cases in a node have identical values for all predictors.

- The node becomes pure; that is, all cases in the node have the same value of the target (or dependent) variable.

- The depth of the tree has reached its prespecified maximum value.

- The number of cases constituting the node is less than a prespecified minimum parent node size.

- The split at the node results in producing a child node whose number of cases is less than a prespecified minimum child node size.

- No more statistically significant split can be found at the specified level of significance.

It should be noted that all but the first two of these rules could be user specified. *CHAID* is not binary; that is, it can produce more than two categories at any particular level in the tree. Therefore, it tends to create a wider tree than do the binary growing methods.

We used the *CHAID* tree growing algorithm to define homogeneous cells or classes for adjustments due to unknown eligibility and nonresponse. Data used to form these classes must be available both for respondents and nonrespondents. In the NSV 2001, the administrative files used to select the List Sample were good sources of information for forming nonresponse adjustment classes for List extended interview data. For the RDD screener nonresponse adjustment, the classes were defined on the basis of the information from the RDD sampling frame. The nonresponse adjustment classes for the RDD extended interview were defined on the basis of the data on the RDD sampling frame and the data collected from the screener survey.

We also used the *CHAID* software to determine the imputation classes for the RDD sample overlap/nonoverlap status for the respondents who did not report Social Security numbers (SSN).

REFERENCES

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,* Vol. 29, pp. 119-127.