ABSTRACT

This paper discusses a customer sampling and reporting system which permits greater accessibility to the contents of an electric utility's customer billing file for the purpose of data analysis and research. The system standardizes the process of obtaining customer extracts, random samples, statistical summary reports, and labels from the customer billing file. SAS is used to manipulate files and data, perform statistical analysis, and generate reports. The system is interactively driven by TSO CLIST's and SAS/FSP to provide an easy-to-use fill-in-the-blank process to select criteria for the extraction, sample, or report. The SAS Macro facility generates SAS code which produces the extract of the population, samples, and reports of the customer file. The primary objective of the system is to transform a billing file into a marketing research tool. Another function of this system is the maintenance of a sampling history file containing previously surveyed customers and to optionally eliminate their records from a sample or to permit longitudinal surveys.

INTRODUCTION AND BACKGROUND

An automated system which creates samples and generates reports from a customer master file is an invaluable tool for the researcher within an electric utility or any other industry. Electric utilities enjoy the benefit of knowing who each of its customers are and how much of its product is sold to each household or customer. Other housing characteristics are also known, such as the type of dwelling and the type of heat. This information allows powerful market research analysis. An automated system provides a method in which customer data can be extracted and aggregated for analysis of trends and relationships.

The idea for an automated sampling and reporting system originated several years ago as a solution to a variety of problems with the existing method of selecting samples and generating reports. The problems had become a major hindrance to the ability to draw samples for anything but the most ordinary of research designs. The existing programs were written in COBOL and were approximately ten years old. At times up to six programs were required to create a sample and generate printed output. Other problems were limitations on the type of selection criteria from which a sample could be drawn, inability to select samples that approached 50% of the population, and the excessive amount of time required to get final output -- up to a week per sample. In addition, the report formats were inflexible, as was access to other data from the customer files.

The objective for the development of the system was to allow greater accessibility to the contents of the customer master file for the purpose of sampling and analytical studies. The three goals of the project were:
1) Provide access to most customer data and allow a wide choice of selection criteria. 2) Generate more reliable samples in a more responsive manner. 3) Create a SAS data sets thus allowing more flexibility in producing reports.

SYSTEM OVERVIEW

The Customer Sampling and Reporting System develops customer extracts from Puget Power's customer master file. Random samples and statistical summary reports are then generated from the extract as required. The customer master file contains such information as mailing and service addresses, telephone numbers, type of business (non-residential), electricity usage, bill amount and other pertinent data relating to the customer, electrical equipment, and consumption history.

CSRS is a menu-driven system written in SAS and TSO/CLIST which provides flexibility to tailor a data file for specific needs. For example, from the hundreds of thousands of customer records in the billing file, a manageable random sample could be obtained by specifying the necessary parameters on the population and sample selection screens. This smaller data file can then be used for statistical analysis, surveys, and other research purposes.

The system runs within the TSO environment through TSO/CLIST's which call SAS interactively. The SAS
interactive program elicits SAS/FSP for the input options, then generates a SAS program to run in the batch mode to access the customer master file or produce samples and reports.

CSRS is divided into three modules:

1) Population Module extracts specified customer records and information from the billing file.
2) Sampling Module creates a random sample from the population extract.
3) Reporting Module organizes data into pre-formatted reports.

POPULATION MODULE

The Population Module involves selecting criteria for the desired extract of the customer master file. Eleven basic parameters from each customer's record can be chosen to narrow the customer master file to a specific population. These categories are:

1) Billing Schedule
2) Revenue Class
3) SIC Code
4) Division
5) County
6) Local Office
7) ZIP Code
8) Billing Cycle
9) Billing Status
10) kWh/KW Selection Criteria
11) Keep Variables

The selections are made through SAS/FSP screens. After all the criteria have been selected for the population extract through SAS/FSP, a TSO file is generated of the selection criteria. The SAS program then reads the TSO file and generates a batch SAS program made up of the selection statements required and other computational macros. This SAS program reads the customer master file, performs the selections and computations and generates a SAS dataset of the selected records and data.

The SAS macro language was utilized to construct the input statement required for the selection and also for selectively running programs to calculate annual and prorated consumption. When the job is completed, output from the population extract consists of: (1) request log; (2) PROC CONTENTS; and (3) a tape containing the extracted SAS datafile.

SAMPLING MODULE

The Sampling Module creates a stratified, systematic, proportionate probability sample from a previously generated extract. Stratification is based on any factor or up to a combination of four factors.

Sampling intervals and random starts for the systematic selection are automatically calculated by SAS. The procedure then produces summary statistics for the universe and the sample.

Creating a sample from the population extract file consists of answering five questions listed below:

1) Is this a survey or research study? If so, a survey, then a three character prefix id is needed.
2) Are previously surveyed customers to be excluded from the sample?
3) Is the sample to be based on a per customer or a per meter basis?
4) What is the sample size?
5) What are the strata variables?

The processing is similar to the population extract in which SAS/FSP is used to determine the specification of the sample. The SAS program then generates the batch SAS program from the specifications. Input is the SAS file created in the population extract step or could be any previously generated SAS file.

Output of the sampling module consists of: 1) Summary report identifying what was specified from the selection screens. 2) Summary report showing internally calculated sampling interval, total number in universe, etc. 3) Summary statistics for the universe and sample. 4) A SAS datafile of those observations which were selected, including all variables which appeared on the population extract file.

REPORTING MODULE

The Reporting Module generates up to five report formats. These reports can be created from any SAS data set and are not limited to output from the population or sample modules.

Standard Customer Report - produces a report which includes customer name, service address, account number, SIC Code, schedule, revenue code, and annual kWh. Sort and page break variables are optional.
Detailed Customer Report - creates a report of up to 15 user-specified variables using the PROC PRINT procedure of SAS. Sort and page break variables are optional.

Customer History Report - creates a report in which each page contains information for one particular meter. There is sufficient space allocated for up to 24 billing histories. Sort variables are optional.

Labels - up to three different sizes of labels may be selected. Any variable on the data file may be printed and each variable occupies one line. Multiple copies may be requested.

Summary Statistics Report - generates a report using the PROC SUMMARY procedure in SAS. The mean, standard deviation, total, minimum value, and maximum value are calculated on KW, kWh, revenue, or total billed amount for revenue month and kWh for calendar month. Also, these statistics can be calculated for annual kWh. PROC SUMMARY uses a CLASS statement to create subgroups; the user may select up to four subgroups.

In order to provide flexibility, a separate macro was developed for each report. This allows additional reports to be added with minimum impact on the system.

CONCLUSIONS

The Customer Sampling and Reporting System has been in use for almost a year. The advantages of the system are quicker turnaround of information and better information for decision making. Although the original intention of the system was in producing samples for survey research, we have found the system has given enough flexibility for other applications. It has been used for studies of rate design and credit evaluation. The system also has potential to provide test samples to test the effect of billing changes. SAS has provided the flexibility needed to generate reports and labels for most any application.

The next major development is to change the input module to access Puget Power's newly redesigned customer master file. The redesigned master file is now organized as an IMS database and contains four times the amount of information as the old version. The next enhancement for CSRS will be to modify the input module to access this additional information. Plans are also being made to add additional standard reports.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the various people who worked on the Customer Sampling & Reporting System at different phases of the project. They are: Lynn Bromaugh, Laurie McCutcheon, Keith Ritland, Doug Sandwich, Ernie Willard, Robert Spencer, and Jim Burridge all from Puget Power. Also the following those from Boeing Computer Services who developed with the requirements and design phase of the project: Robert Jones, Kent Kuiper, Jeff Yaplee, and Nancy Acree.

APPENDIX I

POPULATION SELECTION SCREENS

CUSTOMER SAMPLING & REPORTING SYSTEM (CSRS) Population Selection Module

The following screens can be used: Information needed to produce a sampling plan. There are two main selection methods available: PRO random sampling and PRO stratified sampling. Before allowing the screen, please read through the entire manual carefully. The user must agree to the information before allowing the selection.

Population Selection

For answers to further questions please contact:

Keith Ritland, Doug Sandwich, Ernie Willard, or Robert Jones.

CONCLUSIONS

The Customer Sampling and Reporting System has been in use for almost a year. The advantages of the system are quicker turnaround of information and better information for decision making. Although the original intention of the system was in producing samples for survey research, we have found the system has given enough flexibility for other applications. It has been used for studies of rate design and credit evaluation. The system also has potential to provide test samples to test the effect of billing changes. SAS has provided the flexibility needed to generate reports and labels for most any application.

The next major development is to change the input module to access Puget Power's newly redesigned customer master file. The redesigned master file is now organized as an IMS database and contains four times the amount of information as the old version. The next enhancement for CSRS will be to modify the input module to access this additional information. Plans are also being made to add additional standard reports.
REPORT SELECTION SCREENS