ABSTRACT

Using the best features of many different pieces of software, RAGS (Regression Analysis and Graphics System) forms a powerful, yet user-friendly tool for use in freight tariff negotiations. In this paper, the motivations for RAGS' development and design will be discussed, along with examples of how RAGS is used in tariff negotiations with carriers who transport Air Products and Chemicals' products. The paper will be concluded with a discussion of how and why such diverse software technologies were integrated, and the savings to the company by using RAGS.

Highlights:

- RAGS provides rapid access to SAS\textsuperscript{TM} regression and graphics software, and through the use of ISPF\textsuperscript{TM} menus, requires no knowledge of the SAS\textsuperscript{TM} commands needed to execute the SAS\textsuperscript{TM} procedures successfully.
- RAGS utilizes an IDMS\textsuperscript{TM} (A Cullinet Corp. product) database to efficiently store regression model data. Data is transferred between SAS\textsuperscript{TM} and IDMS\textsuperscript{TM} via FORTRAN and physical sequential files.
- RAGS paid for itself within the first three months after implementation.

INTRODUCTION

RAGS (Regression Analysis and Graphics System) is a user-friendly system which performs simple and piecewise linear regression analysis. It was born out of a need for the Chemicals Distribution group at Air Products and Chemicals, Inc. (APCI) to perform regression analysis during tariff negotiations with common carriers in a timely fashion. There was also a need to store these regression models in such a way as to facilitate ease of retrieval for future analysis. Finally, the system had to minimize the need for MIS ongoing support, and of course, be built as inexpensively as possible. In order to meet all of these requirements, RAGS combines the best aspects of many different pieces of software to enable the user to enter raw data into the database, perform regression analysis on the data, and graph the results. The user is never required to enter any SAS\textsuperscript{TM} commands.

HOW RAGS IS USED

RAGS' primary area of use is in tariff negotiations. All tariffs are periodically re-negotiated and proposals are given to the Chemicals Distribution group in tabular form. There are often hundreds of data points in one tariff, and always several tariffs to be evaluated. RAGS enables Chemicals Distribution to represent this large amount of data in a more easily understood medium: graphs. Because the graphics convey the large amount of information so readily, the time spent in negotiations is reduced. By additionally using regression analysis, they can also understand (or at least make better predictions about) the underlying strategy the carrier is pursuing. When plotted, the tariffs are rarely perfectly straight lines; often the underlying process is linear, but at certain mileages the rates seem acceptable; however, a look at the plot of residuals in Figure 2 reveals a pattern in the form of a peak which indicates that there are actually two line segments in the data and the breakpoint is at \( x = 13 \). In a good regression model, the residuals should be randomly scattered with no pattern.

Figure 3 shows the final piecewise linear model. Note in Figure 4 that the pattern has been removed from the residuals indicating an improvement in the goodness of fit.

RAGS’ COMPONENTS

RAGS is composed of eight major pieces of software, TSO\textsuperscript{TM}, ISPF\textsuperscript{TM}, DFS/HIP, SAS/GRAPH\textsuperscript{TM}, SAS\textsuperscript{TM}, IDMS\textsuperscript{TM}, TELLA, FOCUS\textsuperscript{TM}, and CRAY. Each piece was chosen for its relative strengths, Figure 5 is an illustration of how each of these pieces of software logically fits together.

The following paragraphs discuss the functions that each numbered piece of software in Figure 5 performs and how all are tied together to form one package.

TSO\textsuperscript{TM} (1)

TSO\textsuperscript{TM} is the interactive software without which RAGS would not have been possible. All of the data files which RAGS uses are created with TSO\textsuperscript{TM} commands. All of the other software runs under TSO\textsuperscript{TM}.

ISPF\textsuperscript{TM} (2)

ISPF\textsuperscript{TM} was chosen for its panel generating capability. RAGS also makes heavy use of ISPF\textsuperscript{TM} utilities such as ECF and BROWSE. ISPF\textsuperscript{TM}'s ability to execute TSO\textsuperscript{TM} commands makes the transition from one piece of software to another virtually transparent to the user.

ANYWHERE within RAGS, the user can press the "HELP" key and see documentation relevant to his current situation.
VS FORTRAN (3)
Since there is no direct interface from SAS™ to an IDMS™ database, a bridge in the form of a physical sequential flat file had to be built. ISPF™ and VS FORTRAN allow the user to select the data to be regressed. VS FORTRAN retrieves the data via IDMS™ and writes the data to a flat file. Once the flat file is built, SAS™ reads the flat file, performs the calculations, and writes the results to another flat file. A VS FORTRAN program reads the results and stores them in the database via IDMS™. Programming purists will note that input and output have been doubled, but with the relatively small amount of data being moved, there is no noticeable impact on RAGS' performance.

SAS/GRAPH™ (4)
Proc GLPLOT is used to generate two plots. The first plots the predicted and actual values vs. the independent variable values. The second is a residual plot, an invaluable tool in judging the goodness of fit of the model. It is important to note at this point that RAGS does require an understanding of regression and residual analysis. The user is the sole judge of whether the fit is good or not; there was no attempt made to automate this decision process.

SAS™ (5)
RAGS needed to perform regression analysis and graph the resulting model and associated residual plots in an interactive mode. SAS™ and SAS/GRAPH™ were logical choices. Proc GLM is used to perform the regression calculations, generate the statistics, and output slopes, intercepts, residuals, and predicted values. The piecewise linear SAS™ code uses SAS™ macros to generate the proper GLM code for a varying number of independent variables. Slopes and intercepts are passed back to the RAGS database via Proc PRINTTO, a DATA step, and a FORTRAN program. The residuals, predicted, and actual values are passed to SAS/GRAPH™.

IDMS™ (6)
An efficient means of interactively storing the raw data and model statistics was needed; IDMS™ provided a powerful database functionality. The data management functionality is provided by a relatively simple IDMS™ database. A data diagram of the RAGS database is shown in the Figure 6.

The following is a description of the data stored on each record:
Model record
Contains summary information about the regression model, its raw data, and piecewise linear coefficients. For example:
- a unique key for each model
- summary statistics for the model raw data
- regression statistics (r-squared, f-stat, etc.)
- model slope and intercept for simple linear model
- summary information for the piecewise linear model.

The model key is the concatenation of following seven variables:
- origin code
- carrier code
- tariff number
- shipment mode
- item number
- column number
- version number

This key completely describes a tariff to the Chemicals Distribution group. The key is also indexed in IDMS™ by origin and carrier codes to provide an efficient means of comparing similar tariffs.

Raw Data Record
A variable length record which stores all observations of the independent and dependent variables.

Piecewise Coefficient Record
Another variable length record containing the associated breakpoint, slope, and intercept for each line segment in the model.

Model Description Record
This record is used to provide space for the user to document each regression model. In an attempt not to force the user to fit his description into a preset number of lines, each record represents one line of documentation. The user can therefore type as many lines as desired. In order to document graphs and reports, RAGS does require a minimum of one line of documentation.

TELL-A-GRAF™ (7)
A natural means of comparing several different tariffs is to overlay plots of each tariff on one graph. As the number of plots being compared increases, it becomes increasingly difficult to identify the individual lines. Changes of color and texture can help, but a need to further document the plots was needed. TELL-A-GRAF™ and the message block facility was the solution. Figure 7 is an actual example generated by RAGS making full use of the message blocks.

FOCUS™ (8)
Finally, FOCUS™ was chosen for its reporting capability against IDMS™ databases and other files. FOCUS™ is a fourth generation language that enables the user to generate reports using statements resembling English. Thus the responsibility for generating reports has been transferred from MIS to the user. This results in an increase in the timeliness of reports and a reduction in MIS costs. Another useful feature is that the user can use FOCUS™ to extract data from any file which is defined to FOCUS™ and place it in a flat file. A VS FORTRAN program is then used to load the data via IDMS™ into the RAGS database for further regression analysis. This greatly relieves the user from the burden of data entry.

CONCLUSION
RAGS is a solid application of simple techniques and proven software that successfully provide a useful function. It is part of a growing family of decision support tools where the emphasis is not on computer syntax but on user friendliness, online documentation, and the linking together of different functions. (LOTUS 1-2-3™ has demonstrated the popularity of this kind of software.)

RAGS has been very successful. Most importantly, RAGS has paid for itself in real dollar savings to APCI within the first three months of use. The savings were realized by the re-negotiations of tariffs being used by APCI. There has also been an increase in the user's acceptance of regression analysis and its use in tariff negotiations.
REFERENCES


TSO™ and ISPF™ are registered trademarks of International Business Machines Corporation, Cary, N.C., U.S.A.
SAS™ and SAS/GRAPH™ are registered trademarks of SAS Institute Inc., Cary, N.C., U.S.A.
IDMS™ is a registered trademark of Cullinet Software, Inc., Westwood, Mass., U.S.A.

TELL-A-GRAPH™ is a registered trademark of Integrated Software Systems Corporation, San Diego, Ca., U.S.A.
FOCUS™ is a registered trademark of Information Builders, Inc., New York, N.Y., U.S.A.
1-2-3™ is a registered trademark of Lotus Development Corp., Cambridge, Ma., U.S.A.

©Copyright 1985 by Air Products and Chemicals, Inc.

SIMPLE LINEAR REGRESSION MODEL

FIGURE 1

PIECEWISE LINEAR REGRESSION MODEL

FIGURE 3

PLOT OF RESIDUALS VS INDEPENDENT

FIGURE 2

FIGURE 4
**FIGURE 5**

**RAGS DATABASE**

- **INDEX UNION**
- **INDEX CARRIER**
- **MODEL RECORD**
  - **RAW DATA RECORD**
  - **PIECEWISE COEFFICIENT RECORD**
  - **MODEL DESCRIPTION RECORD**

**FIGURE 6**
FIGURE 7