

GUI to CRSP and COMPUSTAT Data

Peter C.C. Luan, Computer Support Services, College of Business, Arizona State University
Tempe, Arizona 85287-7106, Internet: Peter.Luan@asu.edu, 602-965-5465

ABSTRACT

CRSP (Center for Research in Security Prices) data and Standard & Poor's COMPUSTAT® data are probably the most often used data for instructional and research purposes in many universities and financial institutions. This paper shares the design, describes the functionality of two graphical user interfaces (GUIs), and details some of the experience of developing these two interfaces.

INTRODUCTION

Under the umbrella of developing an "Academic Data Warehouse" at Arizona State University (ASU), we are working to make large datasets available to students and faculty on line and accessible through user friendly interfaces. CRSP (Center for Research in Security Prices) data and Standard & Poor's (S & P) COMPUSTAT® data are first on the list of databanks which we like to work on, since they are probably the most often used data for instructional and research purposes in many universities and financial institutions. S & P does have a PC Windows product called PC-Plus®, and it does offer the major COMPUSTAT datasets in a CD ROM that may be used to make COMPUSTAT available on one PC or a local area network. But currently, at ASU, our applications and datasets are on a UNIX server. While dataset on a local area network's CD-ROM is only accessible on that network, SAS datasets on a Unix server can be accessed from several places: from users on the UNIX server by SAS programs, from users telnet to the UNIX server and run the SAS programs, from users on etherneted PC's using SAS ODBC drivers, or from users on etherneted PC's by SAS/CONNECT treating datasets as remote libraries.

The interfaces described later were developed under the SAS Application Builder which allows us to construct menus, list boxes, and push buttons easily. A strong advantage of using SAS to develop an application is that the same application can be run on many different platforms, since we need to run the

applications under, for example, UNIX Xwindow, Windows for PC, and VT100 terminal mode.

CRSP

The underlying CRSP data, the interface accessed, is in a structure based on a design proposed previously [Luan 1995]. In contrast to the approaches advocated by Ratnaraj and Katzman (1994), Cram (1996a), and Cram (1996b), we choose to break the huge CRSP file into 19,000+ files, one for each traded security. The single security files, each in the same format as the original CRSP database file, are organized in a tri-level directory tree. Luan (1995), therefore, demonstrated a method to store CRSP data which gets around a 2 gigabyte filesize storage limit present in some versions of UNIX. The method uses SAS's PROC DATASOURCE to access the CRSP data files under their directory tree. This storage system affords UNIX users ease-of-use equivalent to that possible on IBM/VSAM.

The CRSP interface starts out with the window in Figure 1 if you view it from a VT320 terminal mode. The "Beginning" and "Ending" lines allow user to specify the date range (the event

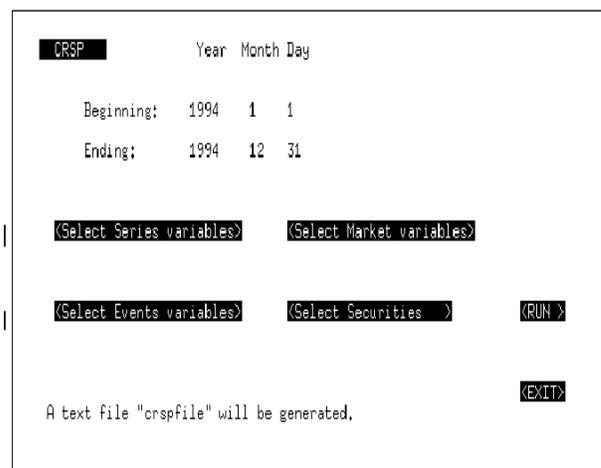


Figure 1 CRSP Main Window

window). There are four selection buttons associated with series, event, market indices, and securities. These four selection buttons will activate separate list boxes as shown in Figure 2, 3, and 4 (the screen for event variable is not shown).

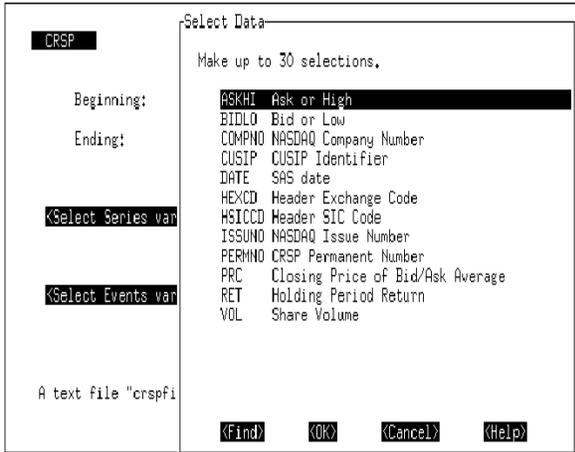


Figure 2 Select Series Variables

Within each list box, the user can use the cursor to move up and down the list, and hit "Enter" to select. The <Find> button allows the user to do string searches within that specific window. We defined the '8' key on the keypad to move a screen forward, and the '7' key to move a screen backward. User can use the "<OK>" button to exit the dialog box. After making selections in these four buttons, the <Run> button will activate a SAS job. At the end of the job submitted, a text file called "crspfile" is generated. Figure 5 shows the content of a sample output file.

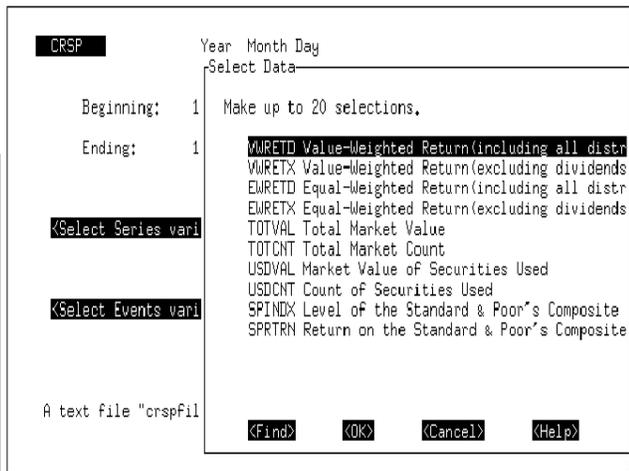


Figure 3 Select Market Variables

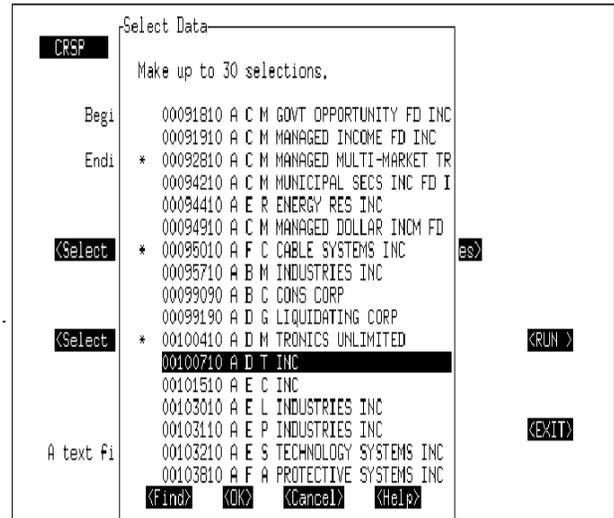


Figure 4 Select Securities

For the purpose of reading all or a subset of CRSP securities, we have developed a separate screen to read a list of "Cusips" from a file. In another similar application, we have also developed a screen to generate monthly returns from daily returns. Since these AF screens are modular, current screens can be moved to different level. We did run into some problems. The application was developed under SAS 6.11, and when running the application under VT100 mode, the screen becomes unclear, which indicates some incompatibility.

CUSIP	DATE	PRC	VOL	RET	USDCNT	SPINDX	SPRTRN	
00092810	03JAN1994		9	23000	0	7512	465.44	-0.002
1653								
00092810	04JAN1994		9	45700	0.00666667	7516	466.89	0.0031
1536								
00092810	05JAN1994		9,125	43700	0.0138889	7523	467.55	0.0014
1355								
00092810	06JAN1994		9	24200	-0.0136986	7526	467.12	-0.000
9197								
00092810	07JAN1994		9	6000	0	7532	469.9	0.0059
5136								
00092810	10JAN1994		9	9000	0	7534	475.27	0.0114
2795								
00092810	11JAN1994		9,125	10500	0.0138889	7536	474.13	-0.002
3986								
00092810	12JAN1994		9,125	28400	0	7538	474.17	0.0000
8438								
00092810	13JAN1994		9	9100	-0.0136986	7538	472.47	-0.003
5852								
00092810	14JAN1994		9.25	26900	0.0277778	7537	474.91	0.0051
6435								
00092810	17JAN1994		9,125	24800	-0.0135135	7540	473.3	-0.003
--More--(1%)								

Figure 5 Listing of the output file

COMPUSTAT

All COMPUSTAT data are converted to SAS data sets first. At ASU, students or faculty can retrieve COMPUSTAT data directly by typing in a single command “**compustat**” right after signing onto the

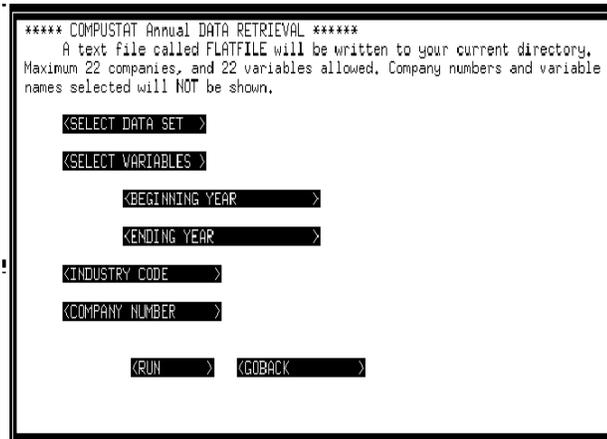


Figure 6 COMPUSTAT Main Window

UNIX Statistical Server (see Figure 6). Use the “Enter” key or the “Tab” key to move down to the “SELECT DATA SET” button. Hitting the “Enter” key again will bring up a dialog box (see Figure 7) which allows you to select one COMPUSTAT data set that you want to work with. Currently, we have five COMPUSTAT files: IND93 (1993 Industrial data), IND94(1994 Industrial data), FULL94(1994 Full-Coverage data), PDE94(1994 Price Dividend Earning data), and RESRCH94 (1994 Research data).

The list of data sets do not have to be the names of real SAS data sets. You can also use name of “View” created by SAS PROC SQL. For example, in order to save space, we dropped the CONAME (company name) and INAME (industry name) in the IND file, and create a view to merge the new IND file with a separate file that has CONAME and INAME, since both CONAME and INAME are long character variables.

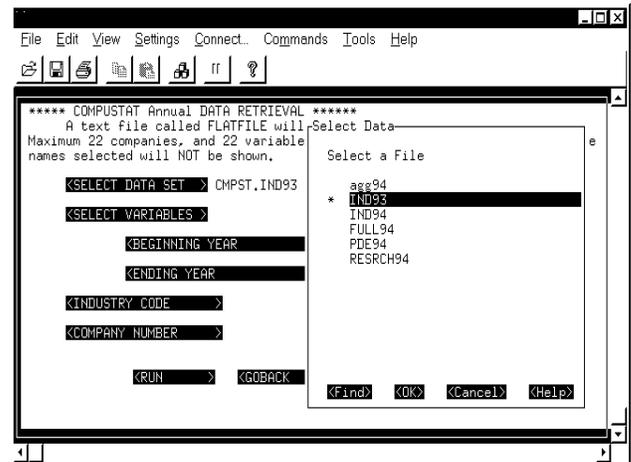


Figure 7 Select Data Set

In this example, “IND93” will be selected. “Tab” to the “SELECT VARIABLES” button, hit the “Enter” key will activate the dialog box that lists all the variables in the dataset selected (see Figure 8). In this dialog box, you can select up to twenty-two variables. This limitation was due to the length of a SCL character variable.

In similar fashion, the “BEGINNING YEAR” button and the “ENDING YEAR” button allow you to specify years of data to include. The “INDUSTRY CODE” button allows you to retrieve data by industry, up to 20 industries. The “COMPANY NUMBER” button allows you to select up to twenty-two companies.

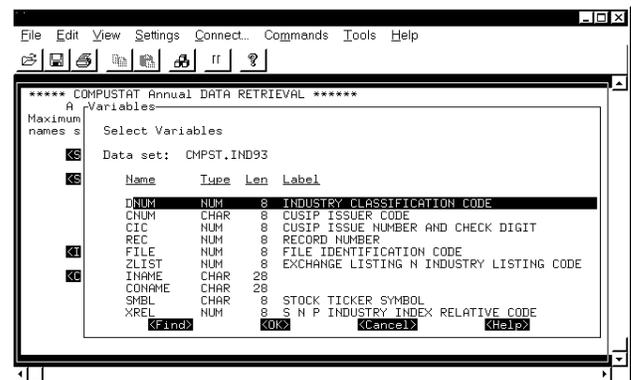


Figure 8 Select Variables

To submit your data request, move your cursor to the “<RUN>” button and hit the “Enter” key. A text file called “flatfile” will be written to the directory where

you started this Compustat program. The “<GOBACK>” button will take you back to UNIX.

FUTURE DEVELOPMENT

We would like to implement other concepts in our COMPUSTAT interface. For example, we would make some important financial ratios available for selection, either at the company level or at the industry level. We would like to use SAS views to save certain selection criteria, which could be subsets or merges of existing COMPUSTAT datasets.

It is desirable to build interfaces to port data from SAS to EXCEL ® through SAS/CONNECT, SAS/ODBC, and OLE. It will also be possible to access CRSP and COMPUSTAT data through the Internet, if proper HTML, CGI interfaces, using appropriate security measures, are built.

CONCLUSION

Building on the CRSP storage structure proposed previously, this paper demonstrates the feasibility of building a GUI to access CRSP data. Using a similar approach, we are also able to build GUI to COMPUSTAT data. The performance of both interfaces is reasonably fast. We built several indices for each COMPUSTAT data set to have better performance. Once the data is within the reign of SAS, many avenues of interfacing and manipulating the data are opened.

ACKNOWLEDGMENTS

The author would like to thank Donald P. Cram for his thorough review and very helpful comments.

REFERENCES

Cram, Donald P. "CRSP Data Retrieval and Analysis in SAS Software: Sample Programs and Programming Tips" in *Proceedings of the Twenty-First Annual SAS Users Group International Conference*, Cary, NC: SAS Institute, Inc., 1996a.

Cram, Donald P. "CRSP Data Retrieval and Analysis in SAS Software for Users and Administrators", Stanford University Graduate School of Business Technical Report #80, 1996b.

Luan, Peter C. C., "Accessing Restructured CRSP Data Using SAS PROC DATASOURCE: A Downsizing Experience", *Proceedings of the Third Annual Conference of the Western Users of SAS Software*, 1995.

Ratnaraj, Paul J. and Carol M. Katzman "Managing Large Financial Data with Ease: CRSP, COMPUSTAT, Etc., with SAS" in *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, Cary, NC: SAS Institute, Inc., 1994.

COMPUSTAT is a registered trademark of Standard and Poor's Compustat Services, Inc. IBM/MVS and IBM/VSAM are registered trademarks or trademarks of International Business Machines Corporation. SAS, SAS/CONNECT, SAS/SCL, SAS/FRAME, and SAS/DATASOURCE are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. EXCEL is a registered trademark or trademark of Microsoft Inc.

® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.