An Example of an Integrated Programming and Data Analysis System

Eric M. Klusman
Federal Reserve Bank of Chicago

Abstract.
CERES, an integrated SAS programming and data analysis system, allows researchers to analyze production databases in a large number of ways without needing to write the SAS code themselves. Users can choose from a variety of panels; each panel generates SAS code based on information from the user. Users can also write SAS programs directly on a SAS programming panel. CERES builds a structured Browse Index of all submitted SAS code and the resulting SAS log and output. With the Browse Index users can review or print SAS input, log, and output, or recapture for resubmission any previously submitted SAS code. CERES was built by enduser programmers using IBM's ISPF dialog management facility and Tangram Systems Corporation's IIF, an interface between ISPF and SAS.

Disclaimer.
The opinions and findings discussed below are those of the author, and do not necessarily reflect a position of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Acknowledgements.
Steve Strongin, Research Officer and research department MIS manager, developed the original design for CERES. The author gratefully acknowledges assistance in preparing the SUGI presentation, as in so many other things, from Steven J. Langford, manager of the departmental technical support group, of which the author is a member.

Previous tools for data analysis.
Economists in the Economic Research Department of the Federal Reserve Bank of Chicago analyze economic and financial time series to better understand the economy and to support recommendations for monetary policy. These economists previously faced two choices: either learn the SAS system in a considerable level of detail, which many found impractical, or work with a research assistant, which typically meant delays between developing an idea and getting the printouts to see whether the data support the idea.

For the last several years the department has sought ways to improve that choice. The goal has been to increase the quality, quantity, and timeliness of economic research by reducing the users' computer effort. Moreover, as the department greatly expanded its production databases of economic and financial time series, new ways were needed to make accessing the databases easier and clearer.

Previous systems for data analysis and SAS programming were built using, variously, CLISTs to invoke SAS for each submission from inside the ISPF/PDF editor, the SAS macro language, SAS/FSP, SAS/AF and SAS Display Manager System. The current system represents a great improvement for our OS/MVS/ZA shop.

User's view: Avoiding SAS programming.
The user gets to CERES by selecting an option on the primary ISPF panel displayed after logon. Once in CERES (figure 1), the user specifies the beginning and ending dates of a sample period and a level of periodicity (i.e., whether to use daily, weekly, monthly, quarterly, or yearly data). After pulling in data from production or personal data libraries, the user moves from panel to panel, specifying statistics or transformations to be calculated, graphs to be drawn, models to be estimated, or even SAS source code to be directly edited and submitted. The path through CERES depends solely on the work to be done and the ideas to be tested. The researcher is able to focus on the economics and able to largely ignore the programming details.

One good example of how programming details are handled for the user is the Data panel (figure 2), the interface to the production databases. Currently production data is maintained in SAS libraries; however, the interface could be changed to accommodate any form of database without changing the look of the panel. The panel is independent from the code generated after displaying the panel.

To use the Data panel, the user can specify exactly which data categories (i.e., SAS datasets) and variables to bring into the current session, or instead just leave those fields blank and press ENTER. At this point selection lists are displayed (figures 3 and 4). The user checks a desired categories and desired variables from the lists. Selection lists are displayed until a list is turned down; then the panel information is turned into a SAS program to pull the requested variables from the production database. The user only had to check off items on lists organized by topic; the computer details were handled by CERES.

Once the user has data, the next step might be to get a preliminary look at it. The user could jump to the Explore panel (figure 5) and check off desired statistics and graphs. If the variables list is left blank, a selection list of variables in the current level of periodicity (figure 6) is displayed. The user selects the variables to be ana-
lyzed, and appropriate SAS output and graphs are presented. Note that the user can with a single key-stroke select a graph involving an annotation data set, axis and symbol statements, and whatever other SAS minutiae (from a user's point of view) may be required, and never need to know about those details.

Perhaps the user now wishes to build a model, perhaps using the AM.R (Advanced Modeling Regression) panel (figure 7). In a panel like this, several of the most common modeling features used by the economists can be specified, merely by filling in variable names and checking off services. When ENTER is pressed the requested SAS/ETS or other modeling procedure output, other requested output, and graphs are displayed.

The user can jump between different panels to try the model under various statistical techniques (but not need to know how to program the techniques in SAS), jump to the Calculate or Explore panels to construct or analyze variables, jump to utility panels to index data or conveniently construct BY variables for subsample partitioning, and so on, all in any order as the need arises. The analyst is in control, and the SAS programming details are in the background. In this way ideas develop more fluidly and more quickly, and economists' and research assistants' productivity increases.

**User's view: The Browse Index.**

Of course, in a lengthy session a user may wish to review or print previous work. The INDEX command, available anywhere in CERES, brings up the Browse Index panel (figure 6). Here submissions are organized by title line, and each SAS data or procedure step is shown. Users can browse the input code, SAS log, or output for any step or the entire submission by entering the appropriate selection code of I, L, or O on the appropriate line. Anything users can browse they can print by using the selection codes PI, PL, and PO.

**User's view: Recapturing code.**

The third use of the Browse Index is to recapture previously submitted SAS code. This function is available on the SAS programming panel (figure 9), which was discussed in some detail in a previous SUGI paper ("Subsystems for the production of Version 5 Graphics", Eric M. Klusman and Steven J. Langford, SUGI 11 Proceedings). This panel invokes the standard ISPF/PDF editor ("Option 2 Edit") if editing is requested, then submits the specified source code to SAS. By checking "Previous Submissions" under the "Alternate Source" heading the user sees the Previous Submissions Index, which has a Browse Index format. The user checks desired submissions, and the input code from those is copied from the Browse Index into a scratchpad dataset. The user edits and resubmits the code, browsing the SAS log and output if selected.

**User's view: Summary.**

From the user's point of view, there are four particular advantages to CERES, in addition to the general advantage of not having to write SAS code directly in many cases. First, under CERES's equidistant topology any function is just a jump away. With no preordained path through CERES, the user moves as desired, answering analytic questions as they arise. Second, via the Browse Index the user is able any point to review, print, or recapture previous work. Third, because panels range from the simple to the complex in function, with the SAS programming panel at the extreme, CERES accommodates a wide variety of skill levels, thereby fitting the needs of a wide range of users. This is not a simple system that users outgrow. Fourth, all of this adds up to make CERES easy to use as well as easy to learn: researchers start with what they know and learn about other panels in turn, all the while able to get things done.

**Programmer's view: Components.**

CERES is composed of CLIST and other code combining three software products: ISPF, IBM's dialog management facility; the SAS system; and IIF, Tangram Systems Corporation's interface between ISPF and the SAS system.

ISPF manages the dialog application and supports the segmentation of dialog components. It transfers control between parts of the dialog, controls the scope of dialog variables, and handles displays of information to the user. Specifically, ISPF displays panels, messages, and help information, provides for the tailoring of general skeleton files into particular SAS programs based on information supplied by user via a panel display, and supports dialog development through test, log, and list facilities.

IIF supports the simultaneous execution of ISPF and SAS. SAS is there when it is needed. In this way, multiple invocations of SAS are eliminated and overhead is reduced. IIF enhances the services and verifications available in ISPF in the following ways. IIF enables one-line verifications of SAS libraries, datasets, variables, and informats from panels or programs. It submits code to SAS, and maintains the resulting Browse Index described above (the application of CLISTs, or in PL/I, FORTRAN, or PASCAL programs. The SAS system is used to analyze, model, and display data. CERES skeleton file tailoring currently writes code to access the SAS base product, SAS/GRAPH, SAS/ETS, SAS/FSP, and SAS/IML. Of course, all installed SAS products are available to the user writing SAS code directly on the SAS programming panel.
How does all the software fit together? The typical CERES menu choice is an ISPF function, coded as a CLIST, that displays a panel (ISPF), verifies information entered by the user on that panel (ISPF and IIF), displaying error messages as necessary (ISPF), then tailors a skeleton file to a particular SAS program (ISPF), submits the program to interactive SAS (IIF), and browses the results (ISPF and IIF). ISPF handles the dialog, SAS handles the data, and IIF fits the two together.

**Programmer's view: Structural advantages.**

Organizing CERES in this way gives many advantages, but they can be grouped under two headings that apply to both the user's and programmer's view.

**Programmer's view: Modularity.**

ISPF provides a well-developed structure for arranging dialog functions into an application. Panels, messages, skeletons, and CLISTs are each stored in separate libraries. This forces a beneficial modular structure on the application. From the user's point of view the application will feel organized, cohesive but not muddled. From the programmer's point of view, the modular structure will pay off at four stages.

First, in development, the task can be conceptualized in its parts and tested a module at a time. The panels can be quickly prototyped, the messages tested for clarity, the skeletons checked for tailoring problems, all independently of each other using ISPF's various test facilities. Additionally, each dialog function's set of CLIST-panel-skeletons can be tested individually.

Second, in debugging, when something breaks it is clear where it broke, since the transfer of control between segments is clear, and each segment has a different kind of error message. By working with three kinds of software, moreover, the scope of derivative errors is contained; for example, SAS data step problems do not in general contaminate the process of displaying panels.

Third, enhancements are easier when only the affected modules need be modified and other modules can be left alone. For example, the SAS code of a particular skeleton can be improved without needing to modify other parts of the function.

Fourth, maintenance is easier with a clear structure. New staff members can more readily learn how the application is programmed, and the original programmers can more accurately remember how the application hangs together. Problems that surface after a long period of hard production use are more rapidly resolved.

**Programmer's view: Playing to Strengths.**

More generally, by using each kind of software to its best advantage, the application is improved overall, in both structural clarity and execution speed.

Apart from the specific advantages of the ISPF-induced modularity mentioned above, organizing the application by software specialization will in general clean up the structure of the application. Tools give better results when they are used more naturally than when they are fought. Playing the piano is easier if you don’t needlessly cross your hands.

Lastly, other things equal, execution speed will be better when each piece is used at its strength, especially when one of the piece’s strength is transferring control between the other pieces. To substitute a lesser tool for any particular task is to lose an advantage.

**Summary.**

CERES, an ISPF dialog using the SAS system and IIF, an interface between ISPF and SAS, provides an integrated environment for SAS users of different skill levels working on a variety of tasks to analyze, model, and graphically display economic time series.

This is accomplished through a variety of panels that, by generating SAS code based on verified information from the user, automate the commonest SAS programming tasks, and a SAS programming panel that allows users to quickly and flexibly develop SAS programs either from scratch or from previously submitted code, all while using the superior ISPF/PDF editor and IIF SUBMIT-to-SAS service, bypassing SAS Display Manager System. All SAS submissions during a session are available in a structured Browse Index for reviewing, printing, and recapturing as detailed above.

All of this is implemented in a clear and modular structure that uses each software to its best advantage, improving application development, debugging, enhancement, and maintenance, and increasing execution speed.

**Trademark notices.**

SAS, SAS/FSP, and SAS/GRAPH are registered trademarks of SAS Institute Inc., Cary, NC, USA.

SAS/ETS and SAS/IML are trademarks of SAS Institute Inc.

IIF is a trademark of Tangram Systems Corporation, Cary, NC, USA.

ISPF is a trademark of International Business Machines Corporation, Armonk, NY, USA.

**Communicating with the authors.**

Address questions on this paper to Eric Klusman, Economic Research Department, Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, Illinois 60690, (312) 322-5601.
Select option ===>  
D DATA        - Production data library interface  
C CALCULATE - Data transformations  
E EXPLORE    - Data explorations - statistics and quick graphs  
R REGRESS    - Data analysis - single-equation regression  
S SAS        - SAS programming facility  
AC CALCULATE - Advanced calculation functions  
AD DATA      - Advanced data functions  
AG GRAPHICS  - Advanced graphics  
AM MODELING  - Advanced modeling  
I ISPF/POF   - Formerly the regular programming environment  
O OPTIONS    - Set options - initialization, hardcopy  
X EXIT       - Exit from CERES

Level: From: To: Device: GODMPCC Obs:

--- Production Data Panel ---

Command ===>  
Gives selection lists for blank fields  
Library ===> MAC (required)  
Category ===>  
Variables ===> _ALL_ merges entire category )

Libraries: MAC Macro  
INT International  
AGR Agricultural (not yet available)  
REG Regional (not yet available)

Level: QUARTER From: 01/01/64 To: 01/01/87 Device: GODMPCC Obs:

--- Figure 1. CERES Primary Panel. ---

--- Figure 2. Data Panel. ---

703
Select category -- PF3=Execute ------------------------------------ ROW 1 OF 1
Command ===) Sera II ===) PAGE

Library: MACQ

? Name          Updated Label

Commercial 87/01/27 16:55 CENSUS, COMMERCE, SCB, AND MISC DATA
EMPLOY 87/01/27 16:55 EMPLOYMENT, LABOR FORCE, AND POPULATION
FLOWS 87/01/27 16:55 FLOW OF FUNDS DATA
FREQ 87/01/27 16:55 FREQUENTLY USED QUARTERLY VARS
INTQ 87/01/27 13:09 INTERNATIONAL QUARTERLY DATA
IP 87/01/27 16:55 IP INDEX AND CAPACITY UTILIZATION
NOMEN 87/01/27 16:55 MONEY, RESERVES, AND BORROWINGS
NIPA 87/01/27 16:55 NIPA, NIA, AND BEA DATA
NIPA07 87/01/07 13:09 NIPA DATA, PRE BENCHMARK REVISION
PRICES 87/01/27 16:55 CPI AND PPI
RATES 87/01/27 16:55 SHORT AND LONG TERM INTEREST RATES
************************************************************************ Bottom of Data ************************************************************************

Figure 3. Data Panel - Dataset Selection.

Select variables -- PF3=Execute ---------------------------------- ROW 51 OF 83
Command ===) Scral I ===) PAGE

Data Set: MACQ.COMMERCE

? Name          Label

MDSF MFG SHIPMENTS - DEFENSE (MIL, SA)
MDSDF MFG SHIPMENTS - NONDEFENSE (MIL, SA)
PSI PERSONAL CONTRIBUTIONS FOR SOCIAL INSURANCE (SAAR)
PSYM PERSONAL SAVINGS (SAAR, BIL)
PSVRE PERS SAV AS PERCENT, DISP PERS INC (SAAR)
PITAX PERSONAL TAX AND NON-TAX PAYMENTS (SAAR)
PTSPM PER TRANSF PAYMENTS TO FOREIGNERS (NET, SAAR)
QTR QUARTER
SASDATE
SMG EST MLY SALES - MANUFACTURERS (SA)
SMB2 WSC AND TRADE SALES (SA, 82$)
SMW EST MLY SALES - MERCH WHOLESALE (SA)
SMW2 MANUFACTURING SALES (SA, 82$)
SPPIR STANDARD AND POOR'S DIVIDEND PRICE RATIO
SFS500 STANDARD & POOR'S 500 STOCKS INDEX (NSA)
SR TOTAL RETAIL SALES (SA)
SRAUT RETAIL SALES - AUTOS (SA, MIL)

Figure 4. Data Panel - Variables Selection.
Command ==> Data Explorations Panel

Variables ==> Check to select a service:
- Display of variables
- Indexed display of variables
- Univariate statistics
- Detailed univariate statistics
- Correlations
- Correlations based on ranks
- Time-plot of each variable
- Graph of each variable
- Trend-line on graph of each variable
- Overlayed graph of up to 6 variables
- Indexed overlayed graph of up to 6 variables
- Cumulative-sum graph of each variable

BY variable ==> Level: QUARTER From: 01/01/64 To: 01/01/87 Device: GDDMPCG Obs: 77 to 169

Figure 5. Explore Panel.

Select variables -- PF3=Execute

Command ==> Data Set: WORK.QUARTER

? Name Label

DATE
M1 M1 - SEASONALLY ADJUSTED
M12P RATE ON COMM. BANK PASSBOOK DEPOSITS
MC21 3 MO CD RATE (SECONDARY MKT)
C20 20 YR CONSTANT MATURITY BOND RATE
C30 30 YR CONSTANT MATURITY BOND RATE
P30 PRIME COMMERCIAL PAPER (3 MONTH)
C30S COPR AAA - SEASONED BOND RATE
SASDATE
SPS00 STANDARD & POOR'S 500 STOCKS INDEX

Figure 6. Explore Panel - Variables Selection.
\begin{verbatim}
--- Model: Regression Panel ---
Command ===>
Dependent = independent variables (2 screen lines)
SP500 = RC3 Y RC3AS

----- 00 ------- 00 ------- 00 ------- 00 <-> Polynomials

Restrict ===> Restrict ===> Test ===> Options ===>
Name for residuals ===> R
Name for predicted ===> P
Test sum of polynomial coeffs:
P1 = P2 = P3 = P4 =

Check to select a service:
- Timeplot of residuals
- White noise tests of residuals
- Full autoregressive tests of dep var
- Timeplot of predicted, actual
- No intercept

Level: QUARTER From: 01/01/64 To: 01/01/87 Device: GDDPCG Obs: 77 to 169

--- Figure 7. Advanced Modeling Regression Panel. ---

--- Browse Index ---
--- ROW 3 OF 27 ---
Command ===>

<table>
<thead>
<tr>
<th>Sub</th>
<th>Proc</th>
<th>Submission/Procedure Name</th>
<th>LOG</th>
<th>OUTPUT</th>
<th>INPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>? ID</td>
<td>ID</td>
<td></td>
<td>Pages</td>
<td>Pages</td>
<td>Lines</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>CERES PRODUCTION DATA</td>
<td>1</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>DATASET</td>
<td>1</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>CERES DATA EXPLORATION</td>
<td>2</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>PRINT</td>
<td>1</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>MEANS</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>CERES DATA EXPLORATION</td>
<td>3</td>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>DATASET</td>
<td>2</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>GPLOT</td>
<td>2</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>CERES ADV MODEL: SINGLE-ION OLS</td>
<td>3</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>DATASET</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>SYSLIN</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>DATASET</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>DATASET</td>
<td>3</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>GPLOT</td>
<td>1</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>CERES DATA EXPLORATION</td>
<td>2</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>GPLOT</td>
<td>1</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>CERES DATA EXPLORATION</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

--- Figure 8. Browse Index Panel. ---
\end{verbatim}
SAS Programming Panel

Command ===>

Dataset name
   Library ===> ER.PRO.SRC.GIEMK00
   Member ===> HILDRETH

Alternative sources
   Previous submissions ===> S
   Fresh scratchpad ===> S
   Previous scratchpad ===> S

Options
   Edit ===> S
   Submit job to SAS ===> S
   Display SAS Log ===> S
   Display procedure output ===> S
   Display browse index ===> S
   Cycle on Edit/Submit options ===> S

Type INDEX on the Command line to browse or print previous submissions

Level: QUARTER From: 01/01/64 To: 01/01/87 Device: GDDM/CGG Obs: 77 to 169

Figure 9. SAS Programming Panel.