INTERFACING SAS SOFTWARE WITH THE SCA SYSTEM: THE SCALINK PROCEDURE

Houston H. Stokes and Lon-Mu Liu
University of Illinois at Chicago

ABSTRACT
The SCALINK procedure provides a link whereby SAS jobs can utilize the extensive time series capability of the SCA statistical system.

1. INTRODUCTION
The SCA Statistical System, developed by Liu, Hudak, Box, Muller, and Tiao (1983) implements the time series methods proposed by Box and Jenkins (1976) and Tiao and Box (1981). The ability to access the SCA system from SAS enhances the time series capabilities of SAS presently in the ETS library. The SCA system allows the user to tentatively identify, estimate, diagnostic check and forecast using ARIMA, transfer function and vector ARIMA models. The system provides an easy to use control language and is a powerful research tool for time series analysis.

PROC SCALINK has been designed to facilitate the use of the SCA system both in a batch mode and interactively. In the first method of operation, the user would utilize PROC SCALINK to build an SCA input procedure file and execute any SCA statements found after the SAS PARMCARDS command. Such a setup is illustrated in Figure 1 below where the SAS variables GASRATE and CO2 are placed in an SCA input procedure named SASDATA. The SCA statements, listed in lines 210-290, are executed in the next step by SCA.

In the batch mode, SCA system can be run under SAS and use the SAS generated SCA input procedure immediately. In the interactive mode the SCA input procedure file generated by SCALINK (on unit 8 in this example) would be made permanent and SCA system could invoke this file interactively. PROC SCALINK will allow the user to add data to an existing SCA input file under another SCA program name on the same file. In this mode the SAS user can build variables and make repeated calls to PROC SCALINK to build different SCA input procedures.

Advantages of the SCA Statistical System include:

1. Easy specification for multi-parameter univariate and multivariate models,
2. Adaptive capability facilitating iterative model identification, estimation, diagnostic checking and forecasting for univariate and multivariate autoregressive-moving average time series,
3. Flexibility for setting constraints on parameters,
4. Identification procedures for mixed ARIMA models using extended sample autocorrelation function,
5. Comprehensive methods for identification of multiple-input transfer function models,
6. Conditional least squares and maximum likelihood estimation algorithms,
7. Ability to retain multiple model information.

2. SPECIFICATIONS
The statements used with PROC SCALINK are:

PROC SCALINK options;
VAR variables;
PARMCARDS;
SCA statements ;

PROC SCALINK Statement
PROC SCALINK options;

The options below may appear in the PROC SCALINK statement. Note that only the FORMAT option is used in the example in Figure 1, other options are omitted and the default conditions are used.

DATAUNIT=m specifies the file unit where the SCA data input procedure is to be placed. If DATAUNIT is not specified, m defaults to 8.
PROCUNIT=n specifies the file unit where the SCA statements after the PARMCARDS are to be copied. If PROCUNIT is not specified, n defaults to 3.
MISSING=r specifies that SAS missing values are to be recoded to r. If MISSING is not specified, r defaults to .10D75.
FORMAT='xxx' specifies that the SCA input procedure is to use format xxx. If FORMAT is not specified, it defaults to (25, 16).
PROCNAME=label specifies that the SCA input procedure will have the name "label". If PROCNAME is not specified, "label" defaults to SASDATA.
DATA=SASdataset specifies the SAS data set that the SCALINK procedure is to process. If DATA is omitted, PROC SCALINK uses the most recently created SAS data set.

NODATA specifies that no data is to be placed on the SCA input procedure file. This option is used in batch mode to run SCA system under SAS.

NOPRINT specifies that SAS output will not indicate what variables were passed to the SCA system.

PRINT specifies that SAS output will indicate what variables were passed to the SCA system. PRINT is the default.

VAR Statement

VAR variables;

The VAR statement specifies the variables to be passed to the SCA system. If the VAR statement is not present, all SAS numeric variables in the current dataset will be passed to the SCA system unless NODATA has been specified as an option.

PARMCARDS Statement

PARMCARDS;
SCA control statement

The PARMCARDS statement signals that the SCA statements follow.

3. DETAILS

The example listed in Figure 1 illustrates a SAS job which loads the gas furnace data to a file (see SAS ETS manual page 103) and passes it to the SCA system. The default format has been overridden and a more compressed format has been supplied. SCA statements are given in lines 210-290. Sample output from this job is also given in the Example section. If the command

PROC SCALINK; PARMCARDS;

had been given, the same result would have occurred except that a more accurate format for data transmission would be used (since the default format has larger field).

If the user changes the format from the default, it is important that care be taken so that variables too large for the format do not become accidentally missing.

PROC SCALINK allows multiple calls in one job. For more than one call per SAS JOB the user must specify distinct SCA procedure names (with the PROGNAME option). Failure to follow this warning will result in the PROC SCALINK building multiple SCA procedures with the name SASDATA on unit DATAUNIT. The SCA system will only read the first procedure. Assuming that the SAS user has three SAS datasets (LL1, LL2 and LL3) and wishes to make an SCA input procedure file on unit 23, the commands listed below will transfer all numeric data from the three SAS data sets to the SCA input procedure file on unit 23 and name the SCA input procedures SS1, SS2 and SS3.

PROC SCALINK DATA=LL1 DATAUNIT=23 PROGNAME=SS1;
PROC SCALINK DATA=LL2 DATAUNIT=23 PROGNAME=SS2;
PROC SCALINK DATA=LL3 DATAUNIT=23 PROGNAME=SS3;

The user can then use this SCA input procedure file in an interactive SCA session.

4. EXAMPLE

The example below shows that the control statements related to the use of the SCALINK procedure (Figure 1) and the SAS log and output (Figure 2) that results from running the control cards. The statements in Figure 1 can be divided in four categories: lines 10 to 110 are IBM JCL necessary to execute SAS, lines 120 to 170 are usual SAS statements, lines 180 to 300 contain the instructions of the PROC SCALINK, and lines 310 to 340 are JCL to execute the SCA procedure generated by SCALINK.

The PARMCARDS in this example contains six SCA statements. The first statement (at line 210) invokes the data input procedure created by SCALINK. The time series named SASRATE and CO2 are stored in the SCA workspace after the execution of this statement. Upon examination of the results of the MIDDEN statement (for identification of vector ARMA models), a vector AR(6) model is considered to be appropriate for the series. The model is specified using the MTSMODEL statement (for specification of an vector ARMA model), and then estimated using the MESTIM statement (for estimation of vector ARMA models). The residual series of the estimated model is held under the variables labeled RGASRATE and Re02. The sample cross correlation matrices of these residual series are computed (using another MIDDEN statement) for the purpose of diagnostic checks. The estimated model is appropriate since almost all sample cross correlations of the residuals are insignificant. The last statement (MFORECAST) is used to forecast the future values of the series using the estimated model. In practice, it is important to constrain insignificant parameter to zero and re-estimate the model. This step is not illustrated for the sake of brevity.
The document contains code examples and data sets for SAS programming. It includes procedures for reading data, performing statistical analyses, and creating visualizations. The code is written in SAS, and there are notes indicating the use of specific procedures and data sets. The document also includes a note on the version of SAS used and contact information for the author. The layout is typical of a technical report or manual, with sections, subsections, and numbered lines for the code examples.
SIZE OF WORKSPACE IS 50,000 SINGLE PRECISION WORDS
DATE: 12/12/83  TIME: 9:24:11

CALL PROCEDURE IS SASDATA.  FILE IS 10.

INPUT VARIABLE IS GASRATE.  REDEFINE .1000D+75
IS REDEFINED AS A MISSING VALUE.

INPUT VARIABLE IS CO2.  REDEFINE .1000D+75.

RETURN.

WIDTH VARIABLES ARE GASRATE, CO2.  ARFITS ARE 1 TO 7.  CCCM, RCM 6, 7.
MAXLAG IS 12.

TIME PERIOD ANALYZED:  1 TO 296
EFFECTIVE NUMBER OF OBSERVATIONS (NOBE):  296

SERIES NAME  MEAN  STD. ERROR
 1 GASRATE  -0.0568  1.0710
 2 CO2  59.5091  3.1967

NOTE: THE APPROX. STD. ERROR FOR THE ESTIMATED CORRELATIONS BELOW IS (1/NOBE**.5) = 0.05812

SAMPLE CORRELATION MATRIX OF THE SERIES

1  0.00
   -0.48  1.00

SUMMARY OF CROSS CORRELATION MATRICES USING +,-... WHERE
+ DENOTES A VALUE GREATER THAN 2/SQRT(NOBE)
- DENOTES A VALUE LESS THAN -2/SQRT(NOBE)
= DENOTES A NON-SIGNIFICANT VALUE BASED ON THE ABOVE CRITERION

BEHAVIOR OF VALUES IN (I,J)TH POSITION OF CROSS CORRELATION MATRIX OVER ALL OUTPUTTED LAGS WHEN SERIES J LEADS SERIES I

1 2
1  ****************
2  ********************

CROSS CORRELATION MATRICES IN TERMS OF +,-

LAGS 1 THROUGH 6
   +  +  +  +  +  +  +  +  +  +  +  +
   +  +  +  +  +  +  +  +  +  +  +  +

LAGS 7 THROUGH 12
   +  +  +  +  +  +  +  +  +  +  +  +
   +  +  +  +  +  +  +  +  +  +  +  +

NOTE: INFORMATION REGARDING LAGS 1 - 5 AUTOREGRESSION IS SUPPRESSED FOR BREVITY.

AUTOREGRESSIVE FITTING ON LAG(S) 1 2 3 4 5 6

BEHAVIOR OF VALUES IN (I,J)TH POSITION OF RESIDUAL CROSS CORRELATION MATRIX OVER ALL OUTPUTTED LAGS WHEN SERIES J LEADS SERIES I

1 2
1  ********************
2  ********************

CROSS CORRELATION MATRICES IN TERMS OF +,-

LAGS 1 THROUGH 6
   .  .  .  .  .  .  .  .  .  .  .  .
   .  .  .  .  .  .  .  .  .  .  .  .

LAGS 7 THROUGH 12
   .  .  .  .  .  .  .  .  .  .  .  .
   .  .  .  .  .  .  .  .  .  .  .  .

73
AUTOREGRESSIVE FITTING ON LAG(S) 1 2 3 4 5 6 7

Note: Information regarding lag 7 autoregression is also suppressed.

STEPWISE AUTOREGRESSION SUMMARY

LAG RESIDUAL EIGENVAL CHI-SQ SIGNIFICANCE
VARIANCES SIGMA TEST AIC OF PARTIAL AR COEFF.

1 1.02E+00 723E-01 1677.26 -3.592 \( \ast \times \)
   1.349E+00 2.72E+00 \( \ast \times \)

2 1.366E-01 1.362E-01 779.14 -5.945 \( \ast \times \)
   1.684E-01 1.88E-01 \( \ast \times \)

3 1.355E-01 1.353E-01 32.28 -6.034 \( \ast \times \)
   1.626E-01 1.623E-01 \( \ast \times \)

4 1.353E-01 1.348E-01 1.2283 -0.089
   1.848E-01 1.834E-01 \( \ast \times \)

5 1.351E-01 1.346E-01 5.73 -5.082
   1.571E-01 1.566E-01 \( \ast \times \)

6 1.342E-01 1.340E-01 12.80 -6.101
   1.598E-01 1.594E-01 \( \ast \times \)

7 1.341E-01 1.339E-01 1.76 -6.081
   1.556E-01 1.552E-01 \( \ast \times \)

NOTE: CHI-SQUARED CRITICAL VALUES WITH 4 DEGREES OF FREEDOM ARE

%: 9.5 1%: 13.3

MODEL NAME IS GASMODEL. SERIES ARE GASRATE, CO2. MODEL IS
(1-PH1*B-PH2*B**2-PH3*B**3-PH4*B**4-PH5*B**5-PH6*B**6)SERIES
CONSTANT-NOISE.

ESTIM MODEL GASMODEL. HOLD RESID(GASRATE,C02).

SUMMARY FOR THE MULTIVARIATE ARMA MODEL

SERIES NAME MEAN STD DEV DIFFERENCE ORDER(S)
1 GASRATE -0.0568 1.0710
2 C02 0.5350 3.1960

NUMBER OF OBSERVATIONS = 296 (EFFECTIVE NUMBER = 290)

MODEL SPECIFICATION WITH PARAMETER VALUES

Note: The initial display of parameter values is suppressed.

ERROR COVARIANCE MATRIX

\[
\begin{bmatrix}
3.750037 \\
-2.136201 \\
6.956253
\end{bmatrix}
\]

FINAL MODEL SUMMARY WITH CONDITIONAL LIKELIHOOD PARAMETER ESTIMATES

---- CONSTANT VECTOR (STD ERROR) ----
0.770 (0.064)
0.024 (0.006)

---- PHI MATRICES ----

ESTIMATES OF PHI(1) MATRIX AND SIGNIFICANCE
1.931 -0.916
0.063 1.345

STANDARD ERRORS
0.050
0.74

ESTIMATES OF PHI(2) MATRIX AND SIGNIFICANCE
-1.204 1.109
-1.93 -0.95

STANDARD ERRORS
0.15
0.108

74
### Estimates of \( \phi(1) \) Matrix and Significance

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.70</td>
<td>0.08</td>
</tr>
<tr>
<td>-1.44</td>
<td>0.17</td>
</tr>
</tbody>
</table>

### Estimates of \( \phi(2) \) Matrix and Significance

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.81</td>
<td>0.16</td>
</tr>
<tr>
<td>0.52</td>
<td>0.13</td>
</tr>
</tbody>
</table>

### Estimates of \( \phi(3) \) Matrix and Significance

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.60</td>
<td>0.06</td>
</tr>
<tr>
<td>-0.27</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### Estimates of \( \phi(4) \) Matrix and Significance

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.30</td>
<td>0.11</td>
</tr>
<tr>
<td>-0.30</td>
<td>0.21</td>
</tr>
</tbody>
</table>

### Estimates of \( \phi(5) \) Matrix and Significance

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.14</td>
<td>0.10</td>
</tr>
<tr>
<td>0.19</td>
<td>0.12</td>
</tr>
</tbody>
</table>

### Estimates of \( \phi(6) \) Matrix and Significance

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.19</td>
<td>0.14</td>
</tr>
<tr>
<td>0.40</td>
<td>0.12</td>
</tr>
</tbody>
</table>

### Error Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0348</td>
<td>0.0368</td>
</tr>
<tr>
<td>2</td>
<td>0.0924</td>
<td>0.0950</td>
</tr>
</tbody>
</table>

#### Variables
- GASRATE
- CO2

### Time Period Analyzed
- 7 to 206

### Effective Number of Observations
- 200

### Cross-Correlation Matrices in Terms of...

#### Lags 1 through 6

<table>
<thead>
<tr>
<th></th>
<th>GASRATE</th>
<th>CO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Incremental Model: GAS/MODEL

### 24 Forecasts, Beginning at Origin = 206

<table>
<thead>
<tr>
<th>Series</th>
<th>Gasrate</th>
<th>CO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast</td>
<td>Std. Err</td>
<td>Forecast</td>
</tr>
<tr>
<td>297</td>
<td>0.334</td>
<td>0.165</td>
</tr>
<tr>
<td>298</td>
<td>0.446</td>
<td>0.402</td>
</tr>
<tr>
<td>299</td>
<td>0.517</td>
<td>0.616</td>
</tr>
<tr>
<td>300</td>
<td>0.590</td>
<td>0.796</td>
</tr>
<tr>
<td>301</td>
<td>0.639</td>
<td>0.912</td>
</tr>
<tr>
<td>302</td>
<td>0.657</td>
<td>0.978</td>
</tr>
<tr>
<td>303</td>
<td>0.648</td>
<td>1.014</td>
</tr>
<tr>
<td>304</td>
<td>0.615</td>
<td>1.033</td>
</tr>
<tr>
<td>305</td>
<td>0.570</td>
<td>1.044</td>
</tr>
<tr>
<td>306</td>
<td>0.521</td>
<td>1.051</td>
</tr>
<tr>
<td>307</td>
<td>0.473</td>
<td>1.057</td>
</tr>
<tr>
<td>308</td>
<td>0.430</td>
<td>1.061</td>
</tr>
<tr>
<td>309</td>
<td>0.293</td>
<td>1.065</td>
</tr>
<tr>
<td>310</td>
<td>0.262</td>
<td>1.068</td>
</tr>
<tr>
<td>311</td>
<td>0.196</td>
<td>1.072</td>
</tr>
<tr>
<td>312</td>
<td>0.313</td>
<td>1.076</td>
</tr>
<tr>
<td>313</td>
<td>0.292</td>
<td>1.079</td>
</tr>
<tr>
<td>314</td>
<td>0.272</td>
<td>1.081</td>
</tr>
<tr>
<td>315</td>
<td>0.252</td>
<td>1.082</td>
</tr>
<tr>
<td>316</td>
<td>0.232</td>
<td>1.082</td>
</tr>
<tr>
<td>317</td>
<td>0.212</td>
<td>1.083</td>
</tr>
<tr>
<td>318</td>
<td>0.193</td>
<td>1.083</td>
</tr>
<tr>
<td>319</td>
<td>0.179</td>
<td>1.083</td>
</tr>
<tr>
<td>320</td>
<td>0.167</td>
<td>1.083</td>
</tr>
</tbody>
</table>

### References


Houston H. Stokes
Department of Economics
University of Illinois
Box 4348
Chicago, Illinois 60638

Len-Mu Liu
Department of Quantitative Methods
University of Illinois
Box 4348
Chicago, Illinois 60638