USING SAS'S PROC CALIS TO ESTIMATE EQS MODELS[1]

/* FOLLOWING IS A SIMPLE EXAMPLE OF HOW TO USE PROC CALIS IN SAS TO PERFORM EQS ANALYSIS. SEE PAGES 7-146 IN "SAS TECHNICAL REPORT P-229, SAS/STAT SOFTWARE: CHANGES AND ENHANCEMENTS, RELEASE 6.07" FOR COMPLETE DOCUMENTATION OF THE PROCEDURE. THIS EXAMPLE ASSUMES BASIC FAMILIARITY WITH SAS, AND STRUCTURAL EQUATIONS MODELING (EITHER LISREL OR EQS). PAGE 3 OF THIS DOCUMENT PROVIDES A DIAGRAM OF THE MODEL ESTIMATED BY THIS PROGRAM.

SAS CAN USE A VARIETY OF TYPES OF DATA, INCLUDING RAW DATA, A SAS FILE, CORRELATION MATRIX WITH STANDARD DEVIATIONS, OR A COVARIANCE MATRIX FOR PROC CALIS. THIS EXAMPLE USES A SAS FILE.

AN SPSSX FILE CAN BE CONVERTED TO A SAS FILE BY USING PROC CONVERT. A DOCUMENT WHICH PROVIDES CAN BE FOUND ON THIS WEB SITE.

THERE ARE SOME OPTIONS AVAILABLE FOR ANALYSIS IN VERSION 3 OF EQS FOR THE PC WHICH ARE NOT AVAILABLE IN SAS. SPECIFICALLY, MULTI-SAMPLE ANALYSIS DOES NOT SEEM TO BE AVAILABLE WITH SAS. ON THE OTHER HAND, SAS HAS SOME OPTIONS NOT AVAILABLE ON THE PC VERSION.

IN THE FOLLOWING EXAMPLE, SAS CODE IS IN BOLD, AND COMMENTS ARE NOT. */

```
OPTIONS COMPRESS=YES    CHARCODE LINESIZE=130 ;
LIBNAME SAST     '/FULL/PATH' ;
DATA TTMK; SET SAST.TTMKASC2;
```

/* THE FOLLOWING ARRAYS AND DO OVER LOOP ACCOMPLISH TWO THINGS. FIRST, THEY CREATE VARIABLES NAMED V1 THOUGH V15 SO THAT THE EQS CONVENTION OF USING THESE VARIABLE NAMES CAN BE FOLLOWED. (WHILE IT IS NOT NECESSARY WITH PROC CALIS TO USE A V PREFIX OR A NUMERICAL SUFFIX IN VARIABLE NAMES, IT WAS DONE HERE TO MAKE THE EXAMPLE MORE CLEAR). SECOND, THE STATEMENT "IF ORIG=. THEN MISSDATA=1;" IDENTIFIES ALL CASES THAT HAVE MISSING DATA FOR ANY OF THE VARIABLES IN THE ARRAY "ORIG". THE STATEMENT "IF MISSDATA NE 1;" FOLLOWING THE DO LOOP ELIMINATES CASES WITH MISSING DATA FOR ANY OF THE VARIABLES TO BE USED IN PROC CALIS. LISTWISE DELETION IS THE DEFAULT FOR PROC CALIS, BUT IT IS USEFUL TO ELIMINATE CASES NOT USED IN THE ANALYSIS IF YOU WANT UNIVARIATE OR BIVARIATE STATISTICS. */

```
ARRAY VS V1-V15 ;
ARRAY ORIG SP2WARM SP2STAND SP2TRACK SP2POSRE SP2HOST SP2INDSC
SP2HARSH P12HS S2BOTHPA P12UNDER BLACK HISP S45MINL S45MODL S45SERL;
DO OVER VS;
 VS=ORIG;
 IF ORIG=. THEN MISSDATA=1;
END;
IF MISSDATA NE 1;

LABEL
V1  ='SP2WARM'    V2  ='SP2STAND'
V3  ='SP2TRACK'   V4  ='SP2POSRE'
V5  ='P2HOST'     V6  ='P2INDSC'
V7  ='SP2HARSH'   V8  ='P12HS'
V9  ='S2BOTHPA'   V10 ='P12UNDER'
V11 ='BLACK'      V12 ='HISP'
V13 ='S45MINL'    V14 ='S45MODL'    V15 ='S45SERL';

TITLE1 'PROC CALIS EXAMPLE';

PROC CALIS COV METHOD=ML ALL ;
VAR  V1 V3-V13 ;
```

/* PUT ALL EQUATIONS IN THE LINEQS STATEMENT.

FOR EXAMPLE: V1=F1+E1, V3=BV3_F1(1) F1+E3, V4=BV4_F1(1) F1+E4,

IN THE ABOVE LINE, F1 IS DEFINED. F1 IS A FACTOR, OR LATENT CONCEPT. V1=F1+E1, MEANS V1=F1, PLUS ERROR. THE VALUE OF F1 IS FIXED TO THAT OF V1, SINCE NO NAME IS GIVEN FOR THE PARAMETER. THIS MEANS THAT F1 WILL HAVE THE SAME MEASUREMENT SCALE AS V1. V3=BV3_F1(1) F1+E3, GIVES THE NAME OF BV3_F1 TO THE COEFFICIENT AND AN INITIAL START VALUE OF 1. V4=BV4_F1(1) F1+E4, GIVES THE NAME OF BV4_F1 TO THE COEFFICIENT AND AN INITIAL START VALUE OF 1 .

---

[1]**Prepared by Patty Glynn, Center for Social and Demographic Analysis, University at Albany, State University of New York. 12/23/93, revised 2/25/99**

THE NAMES OF THE RESIDUALS MUST START WITH THE PREFIX LETTER E (FOR ERROR) OR D (FOR DISTURBANCE). THE CONVENTION IS TO USE E FOR MANIFEST VARIABLES, AND D FOR LATENT VARIABLES. THE TRAILING PART CAN CONTAIN LETTERS OR DIGITS.

VALUES IN PARENTHESES INDICATE USER-PROVIDED START VALUES. IF THEY ARE NOT GIVEN, SAS COMPUTES THEM AUTOMATICALLY. PROVIDING START VALUES MAY REDUCE THE NUMBER OF ITERATIONS NEEDED TO SOLVE THE EQUATIONS.

COMMAS MUST FOLLOW EACH EQUATION, AND THE STATEMENT MUST END WITH A SEMICOLON.

FOR NAMING COEFFICIENTS IN THE LINEQS STATEMENT, THIS EXAMPLE USES THE STRATEGY OF BEGINNING ALL NAMES WITH B (FOR BETA OR B FROM REGRESSION), FOLLOWED BY THE DEPENDENT VARIABLE, AN UNDERSCORE AND THEN THE INDEPENDENT VARIABLE. */

```
LINEQS
 V1=       F1+E1,
 V3=BV3_F1 F1+E3,
 V4=BV4_F1 F1+E4,
 V5=BV5_F1 F1+E5,
 V6=BV6_F1 F1+E6,
 V7=BV7_F1 F1+E7,

 F1=    BF1_V8   V8 +
        BF1_V9   V9 +
        BF1_V10 V10 +
        BF1_V11 V11 +
        BF1_V12 V12 + D1,

 V13=     BV13_F1    F1  +
          BV13_V8    V8  +
          BV13_V9    V9  +
          BV13_V10  V10  +
          BV13_V11  V11  +
          BV13_V12  V12  +
          E13  ;
```

/* ALL INDEPENDENT VARIABLES SHOULD BE LISTED IN THE STD STATEMENT. THE NAMING STRATEGY USED HERE IS THAT ALL STD COEFFICIENTS HAVE THE FIRST LETTER OF S, FOLLOWED BY THE NAME OF THE VARIABLE. D2=SD2(.2) MEANS THAT D2 IS AN INDEPENDENT VARIABLE FOR WHICH VARIANCE SHOULD BE ESTIMATED. THE NAME PROVIDED IS SD2, AND A STARTING VALUE OF .2 IS PROVIDED. */

```
STD   E1=SE1, E3=SE3, E4=SE4, E5=SE5, E6=SE6, E7=SE7, E13=SE13, D1=SD1,
      V8 =SV8, V9 =SV9, V10=SV10, V11=SV11, V12=SV12 ;
```

/* THE COV STATEMENT TELLS THE PROGRAM WHICH VARIABLES COVARIANCES SHOULD BE ESTIMATED BETWEEN. FOR EXAMPLE, F1 F5=CF1_F5(.1), TELLS PROC CALIS TO ESTIMATE A COVARIANCE BETWEEN F1 AND F5, TO NAME THE PARAMETER CF1_F5, AND TO START WITH THE VALUE OF .1. THE STRATEGY USED HERE FOR NAMING COVARIANCES IS C, FOLLOWED BY THE TWO VARIABLES INVOLVED, SEPARATED BY AN UNDERSCORE. IN MOST CASES, COVARIANCES FOR ALL EXOGENOUS VARIABLES SHOULD BE ESTIMATED. COVARIANCES FOR ERRORS CAN ALSO BE INCLUDED. */

```
COV   E1  E6   = CE1_E6,      E5  E6   = CE5_E6,      E3  E5   = CE3_E5,
      V8  V9   = CV8_V9,      V8  V10  = CV8_V10,
      V8  V11  = CV8_V11,     V8  V12  = CV8_V12,
      V9  V10  = CV9_V10,     V9  V11  = CV9_V11,     V9  V12  = CV9_V12,
      V10  V11 = CV10_V11,    V10  V12 = CV10_V12,    V11  V12 = CV11_V12 ;

WEIGHT WT75 ;  RUN;
```