PRIN

PRIN (FRAC=*fraction of variance*, NAME=*name of components*,
NCOM=*number of components*, PRINT, SILENT) *list of series* ;

Function:

PRIN obtains the principal components of a group of series. The number of such components obtained may be a fixed number, or it may be determined by the amount of variance in the original series explained by the principal components. Principal components are a set of orthogonal vectors with the same number of observations as the original set of series which explain as much variance as possible of the original series. Users of this procedure should be familiar with the method and uses of principal components, which are described in many standard texts such as Harman or Theil (see the references).

Usage:

To obtain principal components in TSP give the word PRIN followed by a list of series whose principal components you want. The options determine how many principal components will be found. The resulting principal components are also series and are stored in data storage under the names created from the NAME= option.

Options:

NAME= the *prefix* to be given to the names of the principal components: the components will be called *prefix1*, *prefix2*, and so forth. You may use any legal TSP name as the name for the principal components, but the names generated by adding the numbers must also be legal TSP names (i.e., of the appropriate length).

NCOM= the maximum number of components to be determined. The actual number will be the minimum of the number requested, the number of variables, and the number needed to explain FRAC of the variance.

FRAC= the fraction of the variance of the input variables which you wish to explain with the principal components.

The default values of the PRIN options are NAME=P, NCOM=number of variables, FRAC=1.00. This set of options is non-limiting, that is, the maximum number of components possible will always be constructed.

PRINT/NOPRINT tells whether the results of PRIN are to be printed or just stored in data storage.

SILENT/NOSILENT turns off all printed output (equivalent to NOPRINT).

Example:

```
PRIN (NAME=PC,NCOM=3,FRAC=.95) I TIME CONS GOVEXP EXPORTS ;
```

specifies that three principal components are to be found for five variables I, TIME, GOVEXP, and EXPORTS. If 95% of the variance of the five variables can be explained by fewer than three components, the program will stop there. The principal components found will be stored under the names PC1, PC2, and PC3, for further use in the program.

PRIN

Output:

If PRINT is on, the output of the principal components procedure begins with a title, the list of input series, the number of observations, and the correlation matrix of the input series. This is followed by a table for the components, showing the corresponding characteristic root, and the fraction of the variance of the original series which was explained by all the components up to and including this one.

Finally a table of factor loadings is printed: this table shows the weights applied to each component in expressing each input series as a function of the components.

PRIN stores the correlation matrix of the input variables under the name @CORR and stores the components themselves under the names P1,P2,P3, etc. as time series. If you supply a different prefix for the names from P, PRIN will use that when making the names.

Method:

TSP standardizes the variables (subtracts their means and divides by their standard deviations) before computing the principal components. The resulting components have the following properties:

1. They have mean zero, standard deviation unity, and are orthogonal.

2. The correlation coefficients between a principal component vector and the set of original variables are identical to that component's loading factor.

3. The sum of squared loading factors equals the characteristic root. (In some other principal component packages, the sum of squared factor loadings equals unity; this is a matter of arbitrary scaling.) In calculating the principal components, the factor loadings are divided by the characteristic root to obtain a principal component with standard deviation of unity. Other programs treat the scaling differently.

4. The fraction of the variance of the original variables explained by a principal component is its characteristic root divided by the number of variables.

To be more explicit, you could obtain the same results as the PRIN X Y Z; command with the following TSP commands:

```
CORR X Y Z;    MAT EVEC = EIGVEC(@CORR);
MAT FACTLD = EVEC*DIAG(SQRT(@EIGVAL));   ? PRIN may change  signs to make  top row positive
MFORM(TYPE=TRI,NROW=3) ONE=1;
MAT FRACVAR = ONE'@EIGVAL/3;                    ? fraction of variance explained = sum/3
PRINT @EIGVAL FRACVAR FACTLD;   MMAKE XM X Y Z;
MAT PCOM = XM*(FACTLD*(DIAG(@EIGVAL))");  ? assumes X Y Z are already standardized
```

References:

Harman, Harry H., **Modern Factor Analysis,** University of Chicago Press, First Edition (1960), Sec. 9.3 or Third Edition (1976), Sec. 8.3.

Judge et al, **The Theory and Practice of Econometrics,** John Wiley & Sons, New York, 1980, Section 12.5.

Mundlak, Yair, "On the Concept of Non-Significant Functions and its Implications for Regression Analysis," **Journal of Econometrics** 16 (1981), pp. 139-149.

Theil, Henri, **Principles of Econometrics,** John Wiley & Sons, Inc., 1971, pp. 46-56.