SMPL

SMPL *beginning obs. id.* *ending obs. id.*
[*beginning obs. id.* *ending obs. id.* .....] ;
or
SMPL *SMPL_vector_name* ;

Function:

SMPL is used to define the observations of the data which will be used in the following TSP procedures. The SMPL vector is a set of pairs of observation identifiers which define the range(s) of observations which are in the current sample.

Usage:

The sample of observations may be specified in four ways:

1. A SMPL statement listing the beginning and ending pairs of observations to be used.

2. A SMPL statement containing the name of a variable that contains a SMPL vector of pairs of observations ids.

3. A SMPLIF statement with an expression which is true for the observations to include in the sample (see the SMPLIF section).

4. A SELECT statement (same as SMPLIF except it selects observations from the previous SMPL statement instead of the current sample).

The first of these is by far the most common: the simplest form of the SMPL statement just specifies one continuous group of observations. For example, to request that observations 1 through 10 of the data be used, use the command

```
SMPL 1 10 ;
```

If you want to use more than one group of observations, specify the groups in any order on the SMPL statement. For example, the following SMPL skips observation 11:

```
SMPL 1,10 12,20;
```

The observation identifiers on the SMPL statement can be any legal observation identifier:

Simple integers if the frequency is none.

Years if the frequency is annual. If the year is less than 201 and greater than 0, 1900 will automatically be added; this can be reset with the BASEYEAR= option. (See the OPTIONS command entry for details.)

Years followed by a colon and the period if the frequency is monthly or quarterly.

SMPL can be changed as often as you like during a TSP program. While a SMPL is in force, no observations on series outside that SMPL will be stored or can be retrieved, unless they are specified with lags (or leads) and the lagged (led) value is within the sample. For example if the sample runs from 48 to 72 and the variable GNP(-1) is specified, the 1947 value of GNP will be used for the 1948 observation of GNP(-1).

Examples:

SMPL

```
FREQ A ;
SMPL 56,80 ;
SMPL 21,40 46,82 ;
FREQ Q ; SMPL 72:1,82:4 ;
FREQ M ; SMPL 78:5,81:9 ;
FREQ N ; SMPL 2,7 9,14 16,21 23,28 30,35 ;
```

The last example specifies groups of six observations at a time, skipping every seventh observation beginning with the first. This is a common arrangement for panel data with a single lagged endogenous variable.

Suppose we have a vector called SAMPLE loaded with 2,7,9,14,16,21,23,28,
30,35. Then the last example could also be done with

```
SMPL SAMPLE ;
```

It is also possible to specify the SMPL pairs in any order, as long as each pair is a positive range. This can be used for entering data in a different order than it is stored. An example of this would be when you have state data which has the states alphabetical in one data source and grouped in regions in another source. For the alphabetical source, you could use SMPL 1,50; , and for the regional source, you could use
SMPL 33,33 21,21 14,14 17,17 49,49 20,20 ... etc. ;.

Output:

Every time the SMPL is changed, TSP prints out the current sample unless SUPRES SMPL; has been specified earlier in the program. The sample vector is also stored in data storage under the name @SMPL. The number of observations in the current sample is stored as a scalar, under the name @NOB. This can be quite convenient if you do not know exactly how many observations a SMPLIF or SELECT command will yield, for example.