Stepwise regression

Regression analysis can be used as a part of a model building process. To determine an equation for *y* in which only the most significant determinants of $x_1, x_2, ..., x_q$ are included, some form of stepwise procedure for choosing the *x*-variables are included, some form of stepwise procedure for choosing the *x*-variables may be used. But there is no unique statistical method for making these choices; personal judgment may be necessary.

A way to do stepwise regression with LISREL is to start with an equation in which all the γ's are zero and add successively the one particular *x*-variable that improves fit maximally. The fit can be measured either by χ^2 or the improvement in fit by the modification index (MI). If a correlation matrix is analyzed and the GLS fit function is used, the MI's will be exact χ^2-differences.

Although it is often a good idea to run each step separately and inspect the results after each step, the procedure can be made fully automatic using the automatic model specification option (AM). Each MI is then used as a χ^2 with one degree of freedom and variables are added to the regression equation until the largest MI is no longer significant. The significance level in percent is specified with the SL keyword. The default value for SL is 1.

Table: Data for regression of GNP

Year	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
CHOLEST	1857.015							
AGE	154.514	97.978						
HEIGHT	1.220	2.192	6.161					
WEIGHT	128.106	51.804	24.093	420.242				
BIRTHPIL	1.965	0.279	0.204	0.823	0.251			
ALBUMIN	0.882	-0.280	-0.005	-1.725	-0.042	0.129		
CALCIUM	5.149	-0.040	0.168	0.627	-0.015	0.077	0.224	
URICACID	13.130	2.314	0.349	6.977	0.009	0.012	0.088	1.257

In the table above, the covariance matrix for the Werner Blood Chemistry Data (Dixon, 1981) is given. It is based on NO = 180.

The variables are:

y = Cholesterol
x_1 = Age
x_2 = Height
x_3 = Weight
x_4 = Birth pill
x_5 = Albumin
x_6 = Calcium
x_7 = Uric acid.

The following command file (**EX42.LIS**) will perform a stepwise regression analysis on the summary statistics of the table above.

```
Stepwise Regression for Werner Blood Chemistry Data
Using GLS on Correlation Matrix
DA NI=8 NO=180 MA=KM
LA
CHOLEST AGE HEIGHT WEIGHT BIRTHPIL ALBUMIN CALCIUM URICACID
CM FI=EX42.COV
MO NY=1 NX=7 GA=FI FI
OU ME = GLS TV SE MI AM
```

The results are shown below. The modification indices for *x*-variables 1 and 6 are equal to the drop in the GLS χ^2 statistics. After the addition of x_6, the largest remaining modification index, for x_7, is not significant.

Table: Stepwise regression for Werner blood chemistry data

Step	Parameter to free	MI	χ^2	*df*	*p*-value
0	γ_1	44.815	95.530	7	0.000
1	γ_6	17.840	31.297	6	0.000
2	γ_7	6.32	10.573	5	0.061