Computational Methods for Statistic with Applications Project no 3: Applying the Total Least Squares on a Forestry problem

NGSSC, LU, UU

September-October, 2008

In connection to forests growth monitoring and planning, the following problem occurs, namely, to estimate the amount of tree massa and its quality in terms of width and height of the trees in a given area.
To this end, we compare two types of data. On one side, we have data from satellite pictures, available for the whole area in question, which gives some estimation of the width of the stems based on the tree crowns. On the other hand, we have data for some parts of the area, due to direct measurements, done by people. The first data is considered less accurate compared to the second data, although the second data contains errors as well due to the measurement process.
For one particular part of the area, where we have the two types of data available, we want to predict the second from the first.
The satellite data (the explanatory variables) is in the file `Independent.txt` and the measured data (dependent variables) is in the file `Dependent.txt`.
Both files contain more columns than should be used, which should be neglected during the computations.

There are 12 dependent variables, referred to as $Y(m, 12)$, namely
`ST_D25, ST_D50, ST_D75, ST_D90, ST_D100`
`ST_H25, ST_H50, ST_H75, ST_H90, ST_H100`
`ST_n_A, ST_VOL`
There are 14 independent variables, referred to as $X(m, 14)$, namely `perc10 perc20 perc30 perc40`
`perc50 perc60 perc70 perc80`
`perc90 perc95 perc100`
`testavg teststdh vegkvot`
What is known is that the dependent variables Y are correlated. Same holds for X.

There are 180 observations in total, i.e., $m = 180$.
The major aim is to construct a model, which will enable us to explain Y via X.

Task 1: *Blind TLS test:* We assume, that there is no prior knowledge about those dependencies. Consider the linear model

$$X\alpha + Y\beta = 0$$

where $X(m, 14), Y(m, 12), \alpha(14, 1), \beta(12, 1)$.

Compute the TLS solution to the above model (the TLS fit of the parameters α and β) and the corresponding residual. Compute the corresponding correction to the data.

Task 2: Test how good the TLS prediction is. Compute the TLS solution based on observations 1 to 170 and test how good it is to predict dependent variables from the last 10 observations. Since there are 12 dependent variables, the straightforward prediction test is not applicable. Therefore, test the dependent variables one by one, treating the rest of them as independent variables.

Task 3: In certain cases, the matrix which we construct when we use the TLS metnod, may have zero singular values. Then, the literature suggests to use the so-called *Truncated TLS method*. Give a short description of the latter method. A possible reference is
Sima D., Van Huffel S., "Level choice in truncated total least squares", *Computational Statistics and Data Analysis*, vol. 52, no. 2, Oct. 2007, pp. 1103-1118 (or `ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/dsima/abstracts/06-208.html`)

Please write a report following the instructions given at `http://user.it.uu.se/~maya/Courses/NGSSC/index_Stat.html` and submit it to Maya Neytcheva (`maya@it.uu.se`) not later than by November 12, 2008.

Success!