Testing for Structural Change in a Mixture of Linear Regressions

Susana Faria, DMCT, Officina Mathematica, Univ. Minho
Gilda Soromenho, LEAD, FPCE, Univ. Lisboa

Abstract

Over the last years, several structural change tests have been extended to monitoring of linear regression models where new data arrive over time.
In this work, we derive a new procedure for the problem testing for structural change in a mixture of linear regressions. In this procedure, the parameters of a mixture of linear regression model are estimated from all available data (initial data plus newly arrived data) and compared to the estimates based only on initial data.
The procedure is illustrated through a simulation study on simulated data sets.
Simulation results indicated that the proposed procedure is suitable for detecting if new observations are outliers of the estimated model of mixtures of linear regressions.

Keywords: Mixture of Linear Regressions, EM algorithm, Test for Structural Change, Simulation Study

Test for Structural Change

The mixture of linear regression model is given as:

$$Y = X\beta_j + \varepsilon_j \quad \text{with probability} \quad \pi_j \quad (j=1,\dots,g) \qquad (1)$$

where

Y - nx1 matrix of observations of dependent variable

X - nx(k+1) matrix of predictors

β_j - (k+1)x1 matrix of regression coefficients

ε_{ji} - random errors; under the assumption of normality, $\varepsilon_{ji} \sim N(0, \sigma_j^2)$

π_j - mixing probabilities with $0 \le \pi_j \le 1, \ \sum_{j=1}^{g} \pi_j = 1$

Given a set of n independent observations, the parameters of the mixture of linear regression model were estimated by maximizing the likelihood, via the EM algorithm.

Suppose we have L new observations and with n inicial observations, we have a new model:

$$Y = X\gamma_j + u_j \quad \text{with probability} \quad \pi_j \quad (j=1,\dots,g) \qquad (2)$$

where

Y - $(n+L)$x1 matrix of observations of dependent variable

X - $(n+L)$x(k+1) matrix of predictors

γ_j - (k+1)x1 matrix of regression coefficients

u_{ji} - random errors; under the assumption of normality, $u_{ji} \sim N(0, \sigma_j^2)$

π_j - mixing probabilities with $0 \le \pi_j \le 1, \ \sum_{j=1}^{g} \pi_j = 1$

Given a set of $(n+L)$ independent observations, the parameters of the new mixture of linear regression model were estimated by maximizing the likelihood, via the EM algorithm.

We wish to test the equality of the two regression coefficients:

$$H_0 : \beta_j = \gamma_j \ \forall j \in [1;g]$$

$$H_1 : \exists j \in [1;g] \ \beta_j \ne \gamma_j$$

and under the null hypothesis, we have:

$$F = \frac{\dfrac{(S^* - S)}{L}}{\dfrac{S}{n - g\times(k+1)}} \sim F(L, n - g\times(k+1))$$

with

$$S^* = \sum_{j=1}^{g} \frac{SQR_j^*}{\sigma_j^2} \quad \text{e} \quad S = \sum_{j=1}^{g} \frac{SQR_j}{\sigma_j^2}$$

where

S - residual sum of squares from the estimated linear regression model (1)

S^* - residual sum of squares from the estimated linear regression model (2)

Design of the Study

To illustrate the application of this test in a mixture of linear regressions, a simulation study was performed. The scope was limited to the study of two and three components.

-The parameters of mixture of linear regression models were estimated by maximizing the likelihood, using the EM algorithm. The true values were used as the starting values.

- Samples of three different sizes n (n=50, n=100 and n=500) were generated.

- The mixing proportion π_j lying from 0.1 e 0.9.

-The X-variates were generated from a uniform distribution in the interval $[-1;3]$ and in the interval $[0;2]$

-Three typical configurations of the true regression lines: parallel, perpendicular and concurrent.

- Different values of the parameters β_j were chosen in order to represent several practical situations.

The L new observations were introduced in three different situations:

Situation I: L new observations were *outliers* of the estimated model, with L=1, L=2 and L=5

Situation II: L new observations belonged to the estimated model, with L=1 e L=2

Situation III: one observation belonged to the estimated model, another observation was outlier of this model.

Some Examples

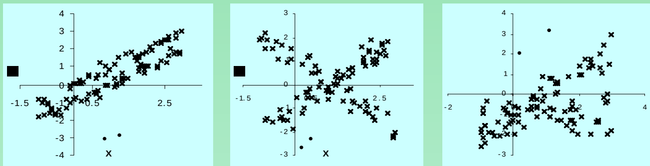

Fig. 1: Scatter plot of samples from situation I (n=100, L=2)

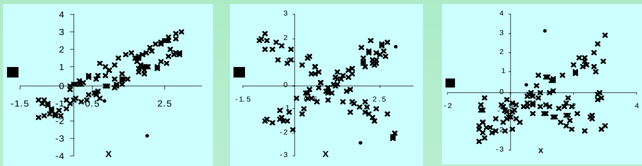

Fig. 2: Scatter plot of samples from situation III (n=100, L=1)

Conclusions

Our results indicated:

- Situation I: Reject the null hypothesis.
- Situation II: Not reject the null hypothesis at the 1% level.
- Situation III: Reject the null hypothesis.

-The results appear not to depend the mixing proportion value, the configuration of the true regression lines and the intervals of values of the X-variates.

The good performance of the test shows that it is suitable for detecting if new observations are outliers of the estimated model of mixtures of linear regressions

References

[1] Celeux, G. and Govaert, G. (1992) "A classification EM Algorithm and two stochastic versions" Computational Statistics and Data Analysis, 14, 315-332
[2] Johnston, J. (1991), *Econometric Methods*, McGraw-Hill International Editions
[3] McLachlan, G.J. and Peel, D. (1997) " Finite Mixture Models" Wiley, New York
[4] Murteira, B., Silva Ribeiro, C., Andrade e Silva, J., Pimenta, C., (2001) *Introdução à Estatística*, McGraw-Hill.