**Extreme bound analysis based on correlation coefficient for optimal regression model**

**Abstract**

**1. Introduction**

**2. Methodology**

I propose two concepts of correlation such as local correlation and global correlation. Local correlation is also called model correlation, which implies fitness of a target regressive parameter with subject to a given regression model. Note, regressive parameter ***α*** = (*α*0, *α*1, *α*2,…, *αn*) is the set of regressive coefficients corresponding to regressors ***x*** = (*x*1, *x*2,…, *xn*) and let *z* and be the responsor and its estimate, respectively. Given regression model *k*, let *Rk*(*xj*, ) and *Rk*(, *z*) be the correlation between *xj* and and the correlation between and *z* within model *k*, respectively.

Suppose the estimate of the *j*th coefficient *αj* with regard to regressor *xj* is , let *Rk*(*xj*, z) be the local correlation of *xj* and *z* within model *k*. Obviously, *Rk*(*xj*, *z*) reflects fitness or appropriateness of the regressive coefficient estimate regarding model *k*. The local correlation *Rk*(*xj*, *z*) is defined as product of *Rk*(*xj*, ) and *Rk*(, *z*) as follows:

For *K* models which are estimated, the averaged local correlation is calculated as follows:

Global correlation implies fitness of the target regressive parameter without concerning any regression models. Let *R*(*xj*, *z*) denote the global correlation between regressor *xj* and responsor *z*, which is defined as usual correlation coefficient as follows:

A regressor *xj* along with its implicit regressive coefficient *αj* are good if they can give sufficient explanatories to possible models and they can be more independent to reflect the responsor *z*. In other words, the first condition of sufficient explanatories to possible models is represented by the local correlation and the second condition of independent reflection is represented by the global correlation. Therefore, the fitness of *xj* and *αj* are defined as product of the averaged local correlation and the global correlation *R*(*xj*, *z*) follows:

The larger the fitness *φj* is, the better the implicit estimate is, and the better the regressor *xj* is. Good regressors *xj* (also *αj* or ) which have large enough fitness *φj* are called robust regressors. Consequently, Regressive Expectation Maximization with RObust regressors (REMRO) algorithm searches for robust regressors and sorts them according to descending ordering with their fitness as searching criterion. Another problem is how to produce *K* models to calculate the averaged local correlation . Fortunately, Martin-i-Salsa generated a set of *K* combinations of doubtful regressors which need to be checked their fitness. Each model in *K* models is estimated with each combination of doubtful ones and estimation method can be least squares method as usual. Moreover, REMRO can resist incomplete data because it applies REM algorithm into filling missing values for both regressors and responsor by estimated values based on ideology of expectation maximization (EM) algorithm. Following is the flow chart of REMRO algorithm.

**Figure 2.1.** Flow chart of REMRO

The final regression model estimated by REMRO with only robust regressors is called optimal regression model.

**3. Experimental results and discussions**

**4. Conclusions**

**References**