Journal of Technical Science and Technologies; ISSN 2298-0032; Volume 3, Issue 2, 2014

# Designing Optimal Integral Aggregate Structure

**Alexander MILNIKOV***
**Daniyar SATYBALDIEV****

## Abstract

In the absence of aprioristic information about the arrangement of the Support points in Integral Aggregate of nonlinear regression analysis there is a need for elimination of statistically non significant Reference point. It is shown that the task is referring to the solution of a well-known task in the multidimensional linear regression analysis on selecting the best regression equation. The efficiency of the developed method is shown on the example of the seven point Integrated Aggregates, approximating a parabolic configuration of observed points.

**Keywords:** Integral Aggregate, Support Points, multiple nonlinear regressions, F-criterion, partial F-criterion

## Introduction

It is (Milnikov A., Sayffulin S., 2012) shown that, nonlinear function, defined in the interval (0, P max) can be approximated with the help of the sum of the linear dependencies of the form

$$D_{A\Sigma}(P) = \sum_{i=i_0}^{r}(d_i - a_i P) = \sum_{i=i_0}^{r} d_i (1 - \frac{z}{x_i})$$
$$(a_i \geq 0; d_i > 0) \quad (1)$$

where $a_i = \frac{d_i}{P_i}$ ; Pi – known values of support points; $(d_i - a_i P)$ (i=i0,…,r) – support functions; i0 – index of Pi, for which the condition is $\min_{P_i}((P_i - P) \geq 0)$; $\min_{P_i}((P_i - P) \geq 0)$ ; $x = \frac{P}{P_{\max}}$ ; z - independent variable (0≤z≤1)иdi–parameters (determinants of support functions), to be estimated.

## Methodology

To evaluate the determinants of support functions the approximating linear regression model was developed, comprising the system of nonlinear functions $(1 - \frac{1}{x_i})$ of independent variables $x_i$ (i=i0,…,r)[1].

Several remarks on terminology; name the model (1) Integral Aggregate. Then, it is clear that it consists of support functions $(d_i - a_i P)$ (i=i0,…,r), which in turn depend on their determinants di. We emphasize that we are dealing with function of one D=f(P) variable, but the model (1) concerning Determinants di represents this dependence as function ofr zi variables (i=1,…, r).In (1) as independent variables are used not zi but functions of $-(1 - \frac{z_i}{x_j})$ . The latter for j = (1… n) generate, observation matrix A

$$A = \begin{vmatrix} (1-\frac{z_1}{x_1}) & \dots & (1-\frac{z_1}{x_i}) & \dots & (1-\frac{z_1}{x_r}) \\ \dots & \dots & \dots & \dots & \dots \\ (1-\frac{z_{k_1}}{x_1}) & \dots & (1-\frac{z_{k_1}}{x_i}) & \dots & (1-\frac{z_{k_1}}{x_r}) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & (1-\frac{z_{k_{i-1}+1}}{x_i}) & \dots & (1-\frac{z_{k_{i-1}+1}}{x_r}) \\ 0 & \dots & (1-\frac{z_{k_i}}{x_i}) & \dots & (1-\frac{z_{k_i}}{x_i}) \end{vmatrix}$$

*Prof.Dr.Department of Computer Technologies and Engineering, International Black Sea University, Tbilisi, Georgia
E-mail: amilnikov@ibsu.edu.ge
** Department Electronics and Nanoelectronics, International Ataturk Alatoo University, Bishkek, Kyrgyzstan
E-mail: daniyar.satybaldiev@iaau.edu.kg

Journal of Technical Science and Technologies; ISSN 2298-0032; Volume 3, Issue 2, 2014

$$\begin{vmatrix} \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & (1-\frac{z_{k_{r-1}+1}}{x_r}) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & (1-\frac{z_{k_r}}{x_r}) \end{vmatrix}$$

of dimension n × r. Values of Determinants d *i* act as the estimated parameters (i=1,…, r). Thus, the problem of evaluating the Integral Aggregate is presented as a linear regression problem for r variables. Obviously, it can be solved by using standard procedures of the multidimensional linear regression analysis, however, previously it is necessary to create a matrix A. Observe that variables are equal to zero for all values of zi> xj (i = 1,2, ..., n ; j = 1,2, ... r).

## Results

In figure 1 Integral Aggregate of parabolic configurations approximated by means of usage of seven support points 2, 5, 8, 11, 14, 17 and 20, is presented. In table 1 results of ANOVA are presented which show quite high statistical adequacy of Integral Aggregate.

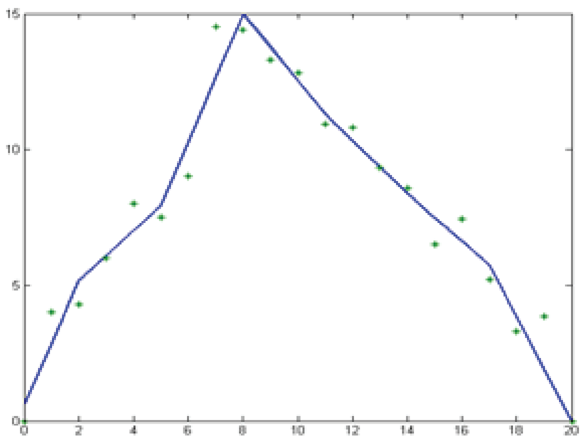

**Figure 1.** *Smoothing using seven support points: 2, 5, 8, 11, 14, 17 u 20.*

**Table 1.** *ANOVA of approximation with seven support points*

| Source | Degree of freedom | SS | MS | F |
|---|---|---|---|---|
| Regression | 7.0000 | 347.4576 | 49.6368 | 43.1435 |
| Residual | 13.0000 | 14.9566 | 1.1505 | |
| Total | 20.0000 | 362.4141 | | |

From fig.1, it is easy to notice that practically there is no change in Support Points 11 and 14 of a configuration of clouds of points under consideration, in contrast, for example, to the point 8, where there is a sharp change.

Concerning the changes of trends in Support Points 2, 5 and 17, it seems that structural changes at the points take place. However, such arguments are qualitative in nature and cannot be considered sufficiently objective. Estimation of the statistical significance of structural changes demands the development of a decisive rule which would allow eliminating statistically excessive Support points. The purpose of this rule is to test the following hypotheses for each of Support points, (i=i0,…, r).

H0: di=0;

H1: di≠0.

The present work is devoted to the development of the decisive rule that would allow carrying out elimination of statically non significant Support points.

As each Support point creates one variable in linear model (1), the task of eliminating statistically non significant Support points is reduced to the solution of the multidimensional linear regression analysis on a choice of the best regression equation. There are many various approaches to its solution, however regarding to our task, the most economical way seems to be the method of elimination (Norman R. Draper, Harry Smith, 1998, David A. Freedman, 2005).

The fact is that in this case the purpose is consecutive test of hypotheses about the appropriateness of inclusion in the linear multidimensional regression model of one variable (one Support point). This allows being limited with single criterion use - partial F-criterion (Norman R. Draper, Harry Smith, 1998, Hardle W. 1990).

Main steps of this method, slightly modified for this task, are as follows.

1. Calculate the regression equation including all variables.

2. Calculate the regression equation with the variable (Support point) excluded by i, that allows to estimate the value of partial F-criterion for this predictor variable.

$$F_c(z_i) = \frac{R^2(z_1,\dots,z_i,\dots,z_r) - R^2(z_1,\dots,z_{i-1},z_{i+1}\dots,z_r)}{1-R^2(z_1,\dots,z_i,\dots,z_r)}$$

where $F_c(z_i)$ - estimate of partial F-criteria for i variable;

$R^2(z_1,\dots,z_i,\dots,z_r)$ - determination coefficient for all predictors;

$R^2(z_1,\dots,z_{i-1},z_{i+1}\dots,z_r)$ – determination coefficient after an exception of i variable.

Number of degrees of freedom is obviously equal: $f_1$=1 and $f_2$=n-r-1, where n – number of observations and r – number of support points.

Journal of Technical Science and Technologies; ISSN 2298-0032; Volume 3, Issue 2, 2014

3. The obtained value of the partial F-test, denoted as Fc, is compared with the tabulated value Ft (1, nr-1) which is taken for the 95% confidence level. If $F_c(z_i) < Ft(1, nr-1)$, the variable i, is excluded from consideration, that is hypothesis H0: di = 0 is rejected. Otherwise, this support point remains in the linear model.

The described procedure is repeated for each of support points (variables) r, therefore excessive variables will be eliminated. Returning to the analysis, task of Integral demands as complex objects composed of elementary linear dependences such as Price - Demand (Support functions). It is possible to claim that the offered procedure allows testing hypotheses of these dependences because the choice of Support points is no other than the statement of hypotheses about Elementary demands. Then H0 hypothesis: di=0 is a hypothesis of the statistical importance of of the Elementary demand i. determined by the value di in the general structure of Integral demand (the Integral Aggregate).

Above we have already noted that at Support Points 11 and 14 of these Aggregates practically there is no change of a trend (structure). In contrast, for example, to a point 8, where there is a sharp change. As for change of trends in Support points 2, 5 and 17, it seems that at them structural changes are happening although it is impossible to tell it with a confidence. We held testing of five internal points (obviously that it is impossible to exclude outer points from consideration). The Integrated Aggregate with the first excluded Support point of z=2 is presented in fig. 2.

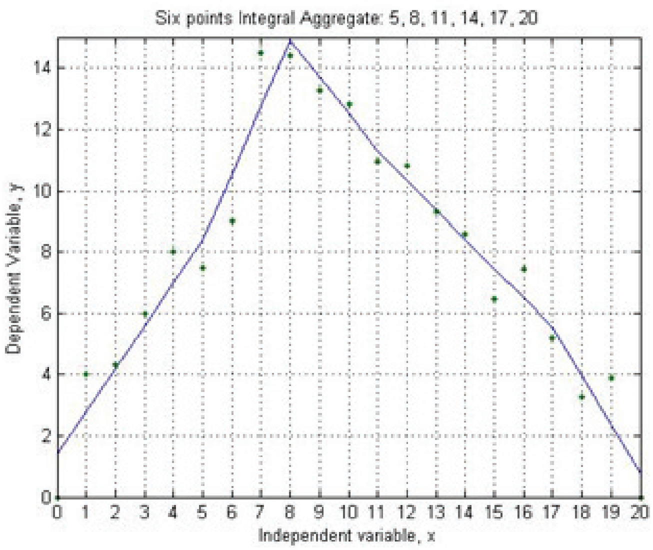


**Figure 2.** *Seven point Integrated Aggregate with excluded support point z=2*

Visually, there is practically no difference in the initial seven point Integrated Aggregate. We don't demonstrate, for obvious reasons, graphical representation of other six point aggregates with consistently excluded internal points.

In Tab. 2 calculated values of partial F-criteria for all six analyzed cases are presented, where it is clearly shown that the only statistically significant point is point z=8, as the settlement $F_c(z_i)$ corresponding to it – criterion

**Table 2.** *Calculated values of partial F-criteria*

| Excluded inflection points | z=2 | z=5 | z=8 | z=11 | z=14 | z=17 |
|---|---|---|---|---|---|---|
| F-criteria | .88 | 3.78 | 24.63 | 0.14 | 0.00066 | 0.74 |

Much greater than tabular F8 (1, 14) =4.60.

It follows that all intermediate points except point z = 8, should be excluded from consideration, as a result, we receive integral aggregate, which should be considered as an optimal one (Figure. 3)

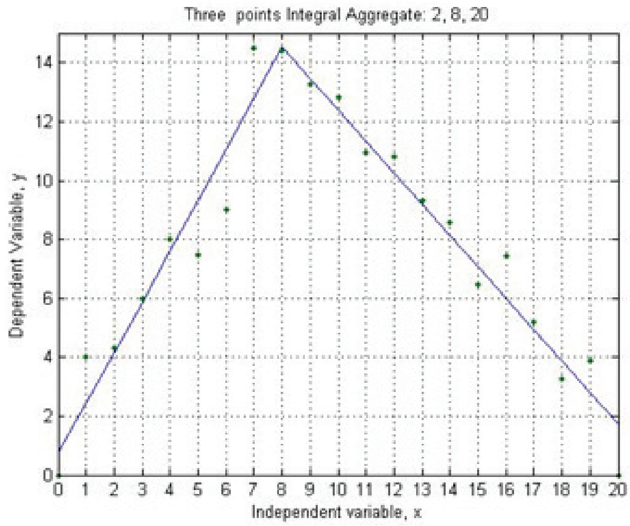


**Figure 3.** *The optimum Integral Aggregate for parabolic configuration*

Note that evaluation of the determination coefficient R2 = 0.94 and Fisher criterion $F_c(z_i) = 142.07$ of this Aggregate suggests that, despite the highly simplified model the loss of adequacy has not occurred. This is confirmed visually in Figure 3.

## Conclusion

In the absence of aprioristic information about the arrangement of the Support points in Integral Aggregate of nonlinear regression analysis there is a need for elimination of statistically non significant Reference point. It demands the existence of the appropriate decisive rule. The purpose of this rule is to test the following hypotheses

H0: di=0;

H1: di≠0.

For each of Support points, (i=i0,…, r). It is shown that the task refers to the solution of a well-known task in the multidimensional linear regression analysis on selecting the best regression equation. For this purpose, the elimination method was used, which consists of a consecutive elimination of internal support points and calculation of the corre-

Journal of Technical Science and Technologies; ISSN 2298-0032; Volume 3, Issue 2, 2014

sponding partial F-criteria. It allows judging their statistical significance.

The efficiency of the developed method is shown on the example of the seven point Integrated Aggregates, approximating a parabolic configuration of observed points. It is shown that out of arbitrarily chosen seven support points, only three are statistically significant: two outer and one internal. The comparative ANOVA of seven - and three point aggregates showed the optimizing character of elimination process: adequacy of the three points aggregate appears not less than above mentioned seven points.

## References

David A. Freedman, (2005). Statistical Models: Theory and Practice, Cambridge University Press.

Hardle W.(1990). Applied Nonparametric Regression.

Milnikov A., Sayffulin. (2012)Principles of Analysis of Internal Structures of Aggregate Demands. IBSU Journal of Business, 1(1) pp.13-17.

Norman R. Draper, Harry Smith. (1998). Applied Regression Analysis (Wiley Series in Probability and Statistics).