# The *F*-test by testing of hypotheses about structure variations in the fuzzy regression models

O.N. Vanjukevich, A.A. Popov

NSTU
630092, Novosibirsk, K.Marks, 20, Russia

***Abstract* –We consider influence on the power *F*-test of outliers distribution, number of observations and method of estimation model parameters.**

## I. INTRODUCTION

By regression model construction on the base fuzzy rules appears problem finding of optimal complexity model. We can use different model selection criteria and *F*-test for solving this problem. In this paper we consider *F*-test by testing of hypotheses about structure variations in the fuzzy regression models.

Consider the Takagi-Sygeno system linear regression models [2,3,4]:

$$if\ \ x_1 \in A_{1j}\ \&\ x_2 \in A_{2j}\ \&\ldots\&\ x_k \in A_{kj}\ \ THEN\ \ y = \eta^j(x),\quad j = \overline{1,l}, \tag{1}$$

where $l$- number of fuzzy rules, $\eta^j(x) = \theta^j x$.

If we parting, for example, $A_{1j}$ -fuzzy partition as a result we have received the following:

$$if\ \ x_1 \in A^1_{1j}\ \&\ x_2 \in A_{2j}\ \&\ldots\&\ x_k \in A_{kj}\ \ THEN\ \ y = \eta^{j,1}(x),$$
$$if\ \ x_1 \in A^2_{1j}\ \&\ x_2 \in A_{2j}\ \&\ldots\&\ x_k \in A_{kj}\ \ THEN\ \ y = \eta^{j,2}(x), \tag{2}$$

where $\eta^{j,1}(x) = \theta^{j,1}x$, $\eta^{j,2}(x) = \theta^{j,2}x$.

Consider two alternative hypotheses:

*H0*: $\theta^{j,1} = \theta^{j,2}$ (not structures variation)

*H1*: $\theta^{j,1} \neq \theta^{j,2}$ (alternatively).

For testing *H0*- hypothesis we can use the following *F*-test [1]:

$$\frac{e^T e - e_1^T e_1}{e_1^T e_1}\,\frac{N-m}{m_1}, \tag{3}$$

where $e^T e$- residual sum of squares of the fuzzy regression model where $\theta^{j,1} = \theta^{j,2}$, $e_1^T e_1$ - residual sum of squares of the fuzzy regression model unrestricted on the model parameters, *N*-number of observations, *m*-number of parameters full fuzzy regression model, $m_1$-vectors length $\theta^{j,1}$ or $\theta^{j,2}$ ( $\dim(\theta^{j,1}) = \dim(\theta^{j,2})$ ).

The theoretical distribution of F-test (3) is the *F*-distribution with degrees of freedom $m_1$ and *N-m*.

In this paper, we used least squares method (LSM) and robust method namely iteratively reweighted least squares method (ILSM) for estimated of model parameters.

## II. THE *F*-TEST BY TESTING OF HYPOTHESES ABOUT STRUCTURE VARIATIONS

In order to research influencing on the *F*-test of distribution of outliers, number of observations and method of estimation model parameters we constructed empirical distributions for *F*-test (3) when *H0* hypothesis were true and when *H1*- hypothesis were true.

In fig. 1 we give plots of empirical distributions for *F*-test (3) when *H0*- hypotheses were true and *N*=82.

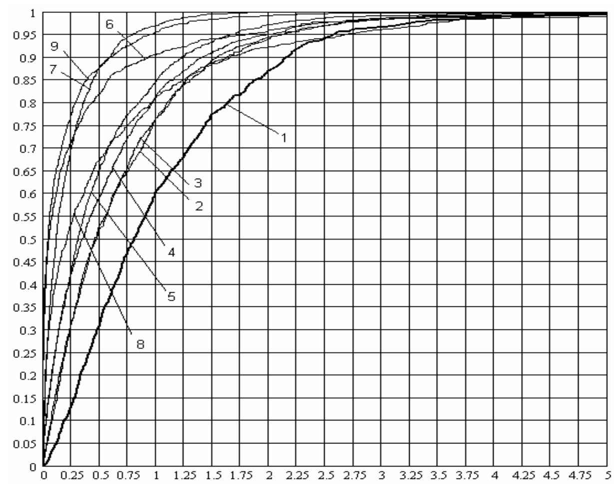


Fig 1. Empirical distributions for *F*-test (3) when *H0*- hypothesis is true and *N=82*:

1- The *F*-distribution with degrees of freedom $m_1$=3 and *N-m*=70. 2,3 - outliers distribution conformed to the normal distribution *N(0,1)*, estimated parameters found LSM and ILSM corresponding. 4,5-outliers distribution conformed to the contaminated *0.05Cauchy(0,1)+ + 0.95N(0,1)*, estimated parameters found LSM and ILSM corresponding. 6,7-outliers distribution conformed to the contaminated *0.5Cauchy(0,1)+0.5N(0,1)*, estimated parameters found LSM and ILSM corresponding. 8,9-outliers distribution conformed to the contaminated *Cauchy(0,1)*, estimated parameters found LSM and ILSM corresponding

In tables 1 and 2 we can see empirical error of first kind. For testing *H0*-hypothesis we selected error of first kind equal 0.05.

It is evident from these figures and tables empirical error of first kind smaller when estimated parameters found ILSM for all distributions of outliers.

Consider influencing on the power *F*-test of

0-7803-8943-3/05/$20.00  

distribution of outliers, number of observation (*N*=50, 82, 289) and method of estimation of model parameters.

TABLE 1

THE ERROR OF FIRST KIND WHEN ESTIMATED PARAMETERS FOUND LSM

| Outliers distribution | Value error | | |
|---|---|---|---|
| | *N*=50 | *N*=82 | *N*=289 |
| *N(0,1)* | 0.004 | 0.028 | 0.048 |
| *N(0,10)* | 0.01 | 0.024 | 0.06 |
| *Log(0,1)* | 0.005 | 0.023 | 0.059 |
| *L(0,1)* | 0.018 | 0.029 | 0.047 |
| *Cauchy(0,1)* | 0.03 | 0.053 | 0.124 |
| *v Cauchy(0,1)+(1-v)N(0,1)* | | | |
| *v*=0.01 | 0.007 | 0.024 | 0.068 |
| *v*=0.05 | 0.007 | 0.022 | 0.084 |
| *v*=0.1 | 0.022 | 0.027 | 0.098 |
| *v*=0.2 | 0.023 | 0.021 | 0.116 |

TABLE 2

THE ERROR OF FIRST KIND WHEN ESTIMATED PARAMETERS FOUND ILSM

| Outliers distribution | Value error | | |
|---|---|---|---|
| | *N*=50 | *N*=50 | *N*=50 |
| *N(0,1)* | 0.007 | 0.026 | 0.049 |
| *N(0,10)* | 0.009 | 0.031 | 0.059 |
| *Log(0,1)* | 0.002 | 0.018 | 0.054 |
| *L(0,1)* | 0.008 | 0.019 | 0.033 |
| *Cauchy(0,1)* | 0.011 | 0.003 | 0 |
| *v Cauchy(0,1)+(1-v)N(0,1)* | | | |
| *v*=0.01 | 0.007 | 0.011 | 0.024 |
| *v*=0.05 | 0.007 | 0.005 | 0.011 |
| *v*=0.1 | 0.022 | 0.007 | 0.011 |
| *v*=0.2 | 0.023 | 0.004 | 0 |

In table 3 we give values of power *F*-test (3) when estimated parameters found LSM. We selected error of first kind equal 0.05.

TABLE 3

THE POWER *F*-TEST (3) WHEN ESTIMATED PARAMETERS FOUND LSM

| Outliers distribution | The power *F*-test | | |
|---|---|---|---|
| | *N*=50 | *N*=82 | *N*=289 |
| *N(0,1)* | 1 | 1 | 1 |
| *N(0,10)* | 0.983 | 0.999 | 1 |
| *Log(0,1)* | 1 | 1 | 1 |
| *L(0,1)* | 1 | 1 | 1 |
| *Cauchy(0,1)* | 0.743 | 0.768 | 0.751 |
| *v Cauchy(0,1)+(1-v)N(0,1)* | | | |
| *v*=0.01 | 0.998 | 0.999 | 0.998 |
| *v*=0.05 | 0.994 | 0.993 | 0.998 |
| *v*=0.1 | 0.984 | 0.989 | 0.964 |
| *v*=0.2 | 0.973 | 0.977 | 0.941 |
| *v*=0.5 | 0.923 | 0.937 | 0.862 |

TABLE 4

THE POWER *F*-TEST (3) WHEN ESTIMATED PARAMETERS FOUND ILSM

| Outliers distribution | The power *F*-test | | |
|---|---|---|---|
| | *N*=50 | *N*=82 | *N*=289 |
| *N(0,1)* | 1 | 1 | 1 |
| *N(0,10)* | 0.979 | 0.999 | 1 |
| *Log(0,1)* | 1 | 1 | 1 |
| *L(0,1)* | 1 | 1 | 1 |
| *Cauchy(0,1)* | 0.677 | 0.683 | 0.685 |
| *v Cauchy (0,1)+(1-v)N(0,1* | | | |
| *v*=0.01 | 1 | 1 | 0.998 |
| *v*=0.05 | 0.997 | 0.999 | 0.994 |
| *v*=0.1 | 0.995 | 0.997 | 0.965 |
| *v*=0.2 | 0.989 | 0.998 | 0.923 |
| *v*=0.5 | 0.972 | 0.984 | 0.809 |

In table 4 we give values of power *F*-test (3) when estimated parameters found ILSM.

It is evident from these tables the power *F*-test (3) depended on distribution of outliers, method of estimation and number of observations.

Consider influencing on empirical distribution *F*-test (3) when H1-hipothesis is true.

In fig. 2 and 3 we give plots of empirical distributions of *F*-test (3) when *H1*- hypothesis were true and *N*=82.

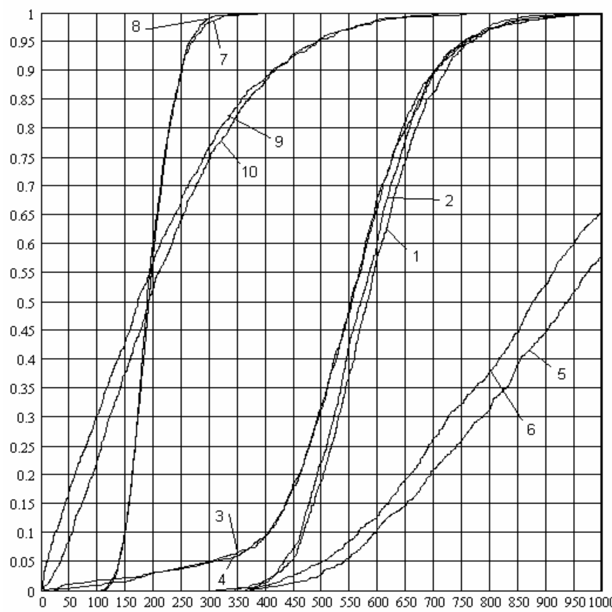


Fig. 2. Empirical distributions for *F*-test (3) when *H1*- hypothesis is true:

1,2- distributions of outliers conformed to the normal distribution *N(0,1),* estimated parameters found LSM and ILSM corresponding. 3,4- distributions of outliers conformed to the logist distribution *Log(0,1)*, estimated parameters found LSM and ILSM corresponding. 5,6- distributions of outliers conformed to the Laplace distribution *L(0,1)*, estimated parameters found LSM and ILSM corresponding. 7,8- distributions of outliers conformed to the contaminated distribution *0.01Cauchy(0,1)+0.99N(0,1)*, estimated parameters found LSM and ILSM corresponding. 9,10 - distributions of outliers conformed to the contaminated distribution *0.2Cauchy(0,1)+0.8N(0,1)*, estimated parameters found LSM and ILSM corresponding

It is evident from these figures and tables the power *F*-test (3) depended on distribution of outliers, method of estimation and number of observations. The power *F*-test (3) larger by estimated parameters found LSM when distribution of outliers conformed to *Cauchy(0,1)* or *N(0,10)* or *vK(0,1)+(1-v)N(0,1)(* if *v*> 0.2 and *N*=289).

We see that the power *F*-test depending on number of observations as would be expected. The power *F*-test decreased where reduced number of observations. On the other hand, we see that power *F*-test decreased when we have large number of observations too. In these cases, we can increase the error of first kind in order to increase the power test.

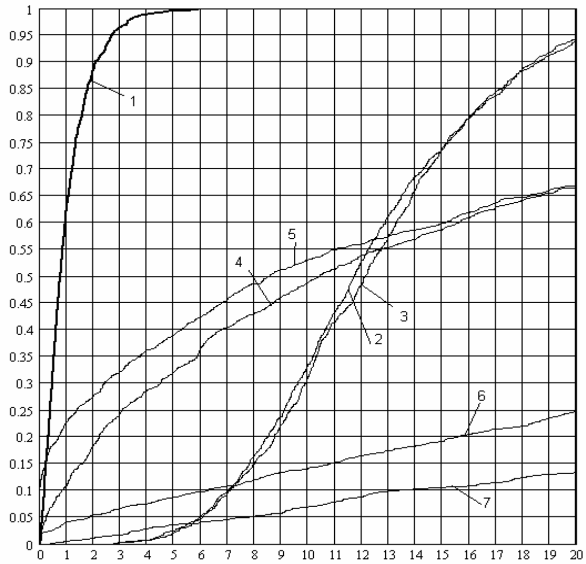

Fig. 3. Empirical distributions of *F*-test (3) when *H1*- hypothesis is true:

1- the *F*-distribution with degrees of freedom $m_1$=3 and *N-m*=70. 2,3- distributions of outliers conformed to the normal distribution *N(0,10),* estimated parameters found LSM and ILSM corresponding. 4,5- distributions of outliers conformed to the contaminated *Cauchy*(0,1), estimated parameters found LSM and ILSM corresponding. 6,7- distributions of outliers conformed to the contaminated distribution *0.5 Cauchy((0,1)+0.5N(0,1)*, estimated parameters found LSM and ILSM corresponding

## III. CONCLUSION

It is evident from these figures and tables the power *F*-test when estimated parameters found ILMS greater than the power *F*-test when estimated parameter found LMS. However, the power *F*-test when estimated parameters found ILMS increased less than on selected error of first kind. On the other hand, empirical error of first kind was much less when estimated parameter found ILMS for all distribution of outliers. But empirical error of first kind was smaller than selected error of first kind even if estimated parameters found LMS. Thus, LMS is rather used for estimating of parameters model by structural identification as ILMS is more laborious method.

### REFERENCES

[1] Aivazian S.A., Eniukov I.S., Meshalkin L.D. Applied statistics. Published by «Finances and Statistics», Moscow, 1985. (in Russian)

[2] Vanujkevich O.N. Popov A.A. Construction of regression dependence with used fuzzy rules. //Collected scientific papers of NSTU - Novosibirsk: NSTU, 2002-№2(28).-pp.3-8. (in Russian)

[3] Kishore Bubna, Charles V. Stewart. Model selection Techniques and Merging Rules for Range Data Segmentation Algorithms// CVIU(80), No. 2, November 2000, pp. 215-245

[4] Popov A.A. Regression modeling on the base of fuzzy rules. // Collected scientific papers of NSTU - Novisibirsk: NSTU, 2000-№2(19).-pp. 49-57. (in Russian)