A Study on the Development of Fuzzy Linear Regression I

Hakyun Kim
(Department of MIS, National Fisheries University of Pusan)

ABSTRACT

This study tests the fuzzy linear regression model to see if there is a performance difference between it and the classical linear regression model. These results show that FLR was better as a forecasting technique when compared with CLR.

Another important find in the test of the two different regression methods is that they generate two different predicted P/E ratios from expected value test, variance test and error test of two different regressions, though we can not see a significant difference between two regression models doing test in error measurements (GMRAE, MAPE, MSE, MAD).

So, in this financial setting we can conclude that FLR is not superior to CLR, comparing and testing between the two different regression models. However, FLR is better than CLR in the error measurements.

1. INTRODUCTION

Many researchers have been interested in the improvement of the decision-making process and numerous fuzzy research stud2ies have been done.

Tanaka, Uejima and Asai (1982) developed the idea of fuzzy linear regression (FLR) as a fuzzy linear function where the parameters are given by fuzzy sets and the fuzzy linear functions are defined by Zadeh's extension principle.

This study tests the fuzzy linear regression model to see if there is a performance difference between it and the classical linear regression model. The results will describe the difference between the two regression models and also show which regression model produces better estimates. The methods based on error measures are important tools to test the difference between fuzzy linear regression and classical linear regression.

This study has the P/E ratio model as the dependent variable with three independent variables, which are dividend payout ratio, earnings growth rate, and beta (risk). The model of this study is based on Litzenberger and Rao (1971).

2. FUZZY THEORY

2.1 Fuzzy set theory

In abstract set theory, an element either does or does not belong to a set. In fuzzy set theory, which is a generalization of abstract set theory introduced by zadeh(1965). A fuzzy set is a class that allows the elements to have partial membership, and therefore, a fuzzy set A is a set of ordered pairs $(x, X_A(x))$, where

$X_A(x)$ is a set of membership of element x in set A. After Zadeh published his new set theory(fuzzy set theory) in 1962, Fuzzy set theory was developed Gupta, Kandel, Kaufmanm Zimmermann[1] and especially Tanaka[2] in Japen. They point out that much of the decision making in the real world takes place in an environment in which the goals, constraints, and consequences of possible actions are not precisely known. A fuzzy member in the fuzzy linear regression considered to be an extention of the concept of confidence interval, which is familiar to anyone who has made computations using imprecise data.

2.2 Fuzzy Linear Regression with Actual Data

Classical linear regression can be represented as follows;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n = \sum_{k=0}^{n} \beta_k x_k$$

Here y depends are a weighted combinations of n independent variables where the β weights are unknown parameters.

The dependent variable, measured in the presence of random and in terms of observed values of the independent variables, is selected as

---

1) Kandel(1982), Fuzzy Techniques in Pattern Recognition, New York: Wiley, Kaufman and Gupta(1985), Introduction to Fuzzy Arithmetic, Newyork: Van Nostrand Reinhold, Zimmermann, Zadeh and Gaine(1984), Fuzzy Set and Decision Analysis, Amsterdam:Northholland

2) Tanaka developed the fuzzy linear regression in 1980

$$y_i = \sum_{k=0}^{n} b_k x_{ik} + e_i, \qquad i = 1, \ldots, m$$

There are m sets of observations on n+1 variables and b_k is the sample estimate of $\beta_k + e_i$ estimates the random distance.

In Fuzzy Linear Regression (FLR), b_k is fuzzy coefficients with a central value and width and so

$$b_k = (c_k, w_k), \quad w_k > 0$$

where c_k = fuzzy center values and w_k = width or wideness

Fuzzy Linear Regression is

$$y_i = \sum_k (c_k, w_k) x_{ik}$$

$$y_i = \left(\sum_k c_k x_{ik} \, , \, \sum_k w_k x_{ik} \right)$$

The term $\sum_k w_k x_{ik}$ is found as non negative values, one way to make $\sum_k w_k x_{ik} \geq 0$ is to take the absolute values of the x_{ik}.

From now, ($\sum_k = \sum$)

$$y_i = \left(\sum c_k x_{ik} \, , \, \sum w_k | x_{ik} | \right)$$

We want a minimum degree of fitness for each observed y_i value. This minimum degree of fitness is called H and we need to solve for c_k and w_k to achieve this. the problem can be stated as linear programming problem.

$$\text{Min} \quad \sum w_k$$

$$\text{S.T.} \quad u(Y_i) \geq H, \text{ for each observed value of the dependent variables}$$

$$u(y_i) = 1 - \frac{|y_i - \sum c_k x_{ik}|}{\sum w_k |x_{ik}|} \geq H$$

LP determines c_k and w_k. Give the minimum degrees of the fitness for the observed values and minimize the sum of the width.

$$\text{Min} \quad \sum w_k$$

$$\text{S.T.} \quad (1-H)\sum w_k |x_{ik}| + \sum c_k x_{ik} \geq y_i$$

$$(1-H)\sum w_k |x_{ik}| - \sum c_k x_{ik} \geq -y_i$$

3. Setting P/E Model

The P/E model of this study is as follows:

P/E = a + bDIV + cGROWTH + dRISK

where,

P/E = Price-Earnings Ratio

DIV = Dividend Payout Ratio

GR2OWTH = Earnings-Growth Rate

RISK = Risk(Beta)

The above P/E model is based on the well-known Litzenberger and Rao model (1971). In their article, earnings-growth rate and risk are the most important variables for estimating the P/E ratio. Beaver and Morse (1978) explored the ability of earnings growth rate and risk to explain P/E ratio. They reaffirmed the Litzenberger and Rao's model. Many research papers, Pari et al(1989), Craig et al (1987), Visser et al (1989), and Mukherjee et al (1989) found the dividend payout ratio as a significant variable for explaining the P/E ratio. Based on the above P/E ratio research, it is clear that growth rate, Risk (Beta) and dividend payout ratio are important variables to determine the P/E ratio. Therefore, the model for the P/E ratio will be estimated with three independent variables. They are dividend payout ratio, earnings-growth rate and risk.

This study compares FLR and CLR to see the differences between the two regression models when actual data is used. It examines the difference between FLR and CLR to see if one method is empirically the more effective regression method for an actual financial regression situation. The data is used 56 American electric companies of 1989. Those companies are retrieved from Compustat. First test is to see if there is a significant difference in the expected P/E ratio and second test is to see if there is a significant difference in variance of the P/E ratio between two different regression models. Third test is to see if there is a significant difference in the errors of P/E ratio between the two different regression models. A paired t-test is used for first and third test. A F-test is used for second test. They use a significance level of 0.05. Four error measurements are used to estimate errors between predicted and observed values for this study - Geometric Mean of the Relative Absolute Error (GMRAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) Mean Absolute Error. Relative Absolute Error (RAE) is calculated for a given error by dividing the absolute forecast error ($|Y_{Fi} - Y_i|$) for a proposed model by the corresponding error for the random walk($|Y_{Ri} - Y_i|$). Forecasted value is defined as Y_{Fi}, random walk value is defined as Y_{Ri}, actual value is Y_i.

$$GMRAE = \left[\prod_{s=1}^{S} RAE \right]^{1/S}, RAE = \frac{|Y_{Fi} - Y_i|}{|Y_{Ri} - Y_i|}$$

MAPE considers the relative error of each forecast. The relative error at value i is $|Y_{Fi} - Y_i| / Y_i$. MAPE is defined to be

$$MAPE = \frac{\sum \frac{|Y_{Fi} - Y_i|}{|Y_i|}}{n}$$

The Mean Absolute Deviation(MAD) is the average of the absolute values of each residual

$$\text{MAD} = \frac{\sum | Y_{Fi} - Y_i |}{n}$$

The MSE is similar to the MAD, except we find the average of the square residuals

$$\text{MSE} = \frac{\sum(| Y_{Fi} - Y_i |)^2}{n}$$

The above four error measurements are used for the accuracy of fuzzy and classical linear regression models. This study uses actual data (complete data) to compare fuzzy and classical linear regression models. Classical linear regression (CLR) models cannot deal with incomplete or vague data. Incomplete data means the data, has missing values. It is theoretically possible to deal with vague data in FLR. However, there is a problem in deciding the error range of the data. This is because the judgment of the decision maker is relied on for deciding this error range.

4. RESULTS

Table 1 shows the P/E ratio model using the fuzzy linear regression and the classical linear regression. Dividend Payout ratio and Beta (Risk) are highly significant variable to estimate the P/E ratio.

Table 1. Fuzzy and Classical Regression Models

Type of Regression	Intercept	DIV	GROWTH	BETA
Fuzzy Linear Regression	9.24**	4.21**	0.54*	-1.06
Classical Linear Regression	9.81**	4.25**	1.09**	-1.92

** P-value is less than 0.05
* P-value id less than 0.1

Table 2 shows the results of three tests. First test is to see that there is a difference between two predicted P/E ratios, the P value is less than 0.05. Thus, predicted P/E ratios of FLR significantly differ from those of CLR. In the second test, variances of predicted P/E ratios are significantly different. The "ANOVA approach for one factor" tests are used. The third test is to see that there is a difference between the error of two regression models. The errors of FLR is significantly differ from those of CLR. Thus, those test concluded that the predicated values, the error of CLR and FLR and the variance of CLR and FLR had difference. In other words, they are different models.

Table 2. Test Results

	First Test	Second Test	Third Test
Statistics	5.590**	4.085**	4.686**

** P-value is less than 0.05

Table 3 shows the error measurement of two different regression models. Error measurements of fuzzy linear regression model are better than those of classical linear regression model. It means fuzzy linear regression presents the better performance than classical linear regression.

Table 3. Four error measurements

Type of Regression	GMRAE	MAPE	MSE	MAD
Classical Linear Regression	2.152	0.238	5.167	2.216
Fuzzy Linear Regression	1.789	0.172	2.758	1.569

5. CONCLUSION

An important purpose of this research is to apply FLR to a financial setting and evaluate its performance. The results of the FLR model can be explained below;

First, important find in the test of the two different regression methods is that they generate two different predicted P/E ratios. Second, test results show that fuzzy linear regression was better as a forecasting technique when compared with classical linear regression. So, in this financial setting we can conclude that FLR is better than CLR, comparing and testing between the two different regression models. This suggests that FLR may be used along with CLR as a methodology to help management forecasting decisions.

If we have data, it is possible to forecast using FLR and CLR. CLR is a very popular and powerful technique, which many researchers are using. FLR is better than CLR in terms of performance in this research paper. There are other new forecasting tools, which have different characteristics for dealing with data.

The estimated fuzzy linear coefficients are given as the solution to an LP problem. If we have knowledge about the coefficient (independent variable) of FLR, it is possible to use such knowledge as a constraint. This means that FLR accept the decision m8aker's knowledge, experience, and opinion from the constraints. Fuzzy theory transfers the qualitative (vague) data such as knowledge, experience and opinion to quantitative data and then uses these quantitative data as constraints in FLR. However, error range of fuzzy data transferred from vague data is a problem for fuzzy theory, be2cause it is decided by the viewpoint of the decision maker. Much fuzzy research tries to reduce error arising from the decision maker. This is another difference between FLR and CLR.

The unique merit of FLR is its ability to control and incorporate incomplete data. FLR has another uniqueness in forecasting data width and fitness. Width and fitness, as well as incomplete data are important topics for further applied FLR studies.

References Cited

1. Armstrong, J. and Collopy, F.(1992), Error Measures for Generalizing about Forecasting Methods: Empirical Comparison, International Journal of Forecasting, 8, 69-80.
2. Box, G.E.P. and Tiao, G.C.(1973), Bayesian Inference in Statistical Analysis, MA:

Addison-Wesley.

3. Craig, D., Johnson, G. and Joy, M.(Mar/April 1987), Accounting Methods and P/E Ratios, Financial Analysts Journal, 41-45.

4. Dubois, D and Prade, H. M.(1980), Fuzzy Sets and Systems : Theory and Application, New York : Academic Press.

5. Heshmaty, B and Kandel, A (1985) Fuzzy Linear Regression and Its Application to Forecasting in Uncertain Environment, Fuzzy Sets and Systems, 15, 159-191.

6. Kandel, A. and Byatt, W.J.(1978), Fuzzy Sets, Fuzzy Algebra, and Fuzzy Statistics, Proceedings of the IEEE, 66, 1619-1639

7. Litzenberger, R. and Rao, C.(1971), Estimates of the Marginal Rate of Time Preference and Average Risk Aversion Shares: 1960-1966, Bell Journal of Economics and M8anagerial Science, Spring, 265-277.

8. Mukherjee, T., and Dukes, W.(1989),A Multi-variate Analysis of Small Bank Stock Valuation, Quarterly Journal of Business and Economics, 1, 3-18.

9. Natvig, B.(1983), Possibility Versus Probability, Fuzzy Sets and System, 10, 31-36.

10. Neter, J., Wasserman W. and Kutner, M. H.(1985), Applied Linear Statistical Models (Second edition), Homewood, IL:Irwin.

11. Savage J. A. (1990), The Future is Uncertain for U.S. Fuzzy Logic, Computer World, 24, 114-119.

12. Schrage, L., (1991), LINDO an Optimization Modeling

System, 4th Edition, CA: The Scientific Press

13. Tanaka, H., Uejima, S., and Asai, K.(1980) Fuzzy Linear Regression Model, IEEE Trans. Systems Man Cybernetics, 10, 2933-2938.

14. Terano, T., Asai, K., and Sugeno, M. (1992), Fuzzy Systems Theory and Its Application, CA:Academic Press.

15. Zadeh, L.A.(1965), Fuzzy Set, Information and Control, Vol 8, 338-865.