

Statistical Software and Regression Diagnostic Reporting with Fuzzy-AHP Intelligent Zax (FAIZ)

Shahid Anjum

Abstract—Most statistical Software do efficient regression reporting. This reporting is based on multiple criteria and uses multiple fragmented diagnostic tables. In this study Fuzzy AHP Intelligent Zax (FAIZ) has been proposed which can be incorporated in statistical software as a scoring technique for ranking the simple linear regressions (SLR) based on fuzzy Analytical Hierarchy Process (AHP) logic to push ‘Further Evolution’ in Software Engineering Evolution process for Statistical Software. This approach will help to prevent regressioneering (ignorant sub-optimal uses, or even misuses, of regression analysis) and assist researchers in choosing best regression based on linearized score with FAIZ approach.

Index Terms—Statistical software engineering, artificial intelligence, linear regression analysis, MCDM.

I. INTRODUCTION

Variety of statistical software are there in the market, some are specialized in highly sophisticated statistical analysis like SAS, SPSS, Statistica, S-Plus and Matlab’s statistical and curve fitting tools etc. Others focus on disciplines that use statistical analysis like econometrics, bio-informatics, time series analysis in economics and various other disciplines of social sciences and engineering sciences. Statistical analysis belongs to conventional data modeling culture of twin modeling cultures in statistics and the other one is newly developed algorithmic modeling culture [1]. Both cultures start with data and perform data analysis for prediction or information extraction where nature functions (black box) associate the predictor (input) variables with the response (output) variables. The algorithmic modeling culture uses knowledge discovery and data mining (KDDM) algorithms where black box can contain decision tree or neural nets or other KDDM algorithmic models which treats the data mechanism as unknown and considers the inside of black box as complex. This approach is focused to find a function $f(x)$ – an algorithm that operates on x to predict the responses y . Model validation is measured by predictive accuracy. As study [1] has noted that statistical community (98% of all statisticians) has been committed to the almost exclusive use of data models where black box may contain linear regression whose parameters are estimated from data and are used for prediction or information extraction. This commitment to data modeling approach, in itself, has led to irrelevant theory, questionable conclusions, and has kept statisticians from

working on a large range of interesting current problems. In short, in an age of low cost and timeless data processing, over-reliance on data modeling is not efficient. Regression analysis is widely used (or misused) statistical tool in various disciplines. There are limitations of regression technique itself or its estimation method or there is lack of its proper use (or has even misuse). Ordinary Least Square (OLS) continues to be by far the most frequently used method even when it is obviously inappropriate. As a result, hundreds if not thousands of regression lines with too-small slopes are being published annually [2]. There is also issue of Regressioneering, a phenomenon which is present when researchers try to use regression without knowing the strengths or weaknesses of regression for particular problem at hand. Moreover, researchers don’t perform enough assumption tests or which criteria to focus on for judging the validity of regression or even when to stop the process. Sometimes simply mathematical ignorance on part of social scientists leads to wrong applications or unexplainable or even wrong results. So far none of the statistical software has developed a mechanism to turn multidimensional diagnostics into a single dimensional score for easy comparisons of different regressions which can aid users to make right choice. This is because no ranking or scoring method is available to report ranking score for comparison of regressions.

Software evolution is part of software engineering, to make the software to continuously capture new requirement from people, including function and non-function requirements. Function requirement is to make the system can do more for people whether non-function requirement is mainly related to quality of software, software architecture optimization, system evolvability, system stability and system robust [3]. The software evolution approach has six major steps in the evolution life cycle which are evolution requirement specification, evolution design definition, evolution implementation, evolution execution, evolution testing, and further evolution. Current statistical software are mature enough in many aspects, however, still ‘further evolution’ is ever needed and this paper attempts to extend it for statistical software by introducing a regression ranking method in order to assist the researcher to make prudent choice regarding regression analysis.

II. PROBLEM ANALYSIS

Linear regression can generally be used for estimation or descriptive purposes (i.e. for cause/effect relationships provided that there is only one independent variable or the data have come from a designed experiment), control (esp. when it is necessary to try to control the value of the dependent variable e.g. cost) and prediction of outcomes

Manuscript received June 14, 2013; revised August 28, 2013. Views expressed in this paper and all errors and emissions are that of author and in no case reflect those of LTU, its staff or management. Author reserves the right of filing patent(s) against new ideas introduced in his articles.

S. Anjum is with the College of Management, Lawrence Technological University (LTU), Southfield, MI 48075 USA (e-mail: anjumsw@hotmail.com).

purpose. Specific uses of it can be calibration (i.e. using inverse regression, classical theory of calibration, or Bayesian calibration) and process monitoring (using regression control chart or a cause-selecting chart) [4]. Simple linear regression (SLR) needs two sets of n observations for dependent (or response) variable (DV) and independent variable (IV) represented by Y and X respectively. SLR model equation is $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (where $i = 1 \dots n$ and $\varepsilon_i =$ random errors) and have three parameters i.e. β_0 (i.e. intercept), β_1 (i.e. slope) and ε_i (random errors). Before deciding what straight line to use, you must decide what you want it for. Do you wish to estimate (predict) one quantity from another, or do you want a descriptive trend line relating two sets of observations [5]. Important feature of SLR is that ε_i are normally distributed and $Y_i \sim N(\alpha + \beta X_i, \delta^2)$. Fitted SLR equation is $\hat{Y}_i = b_0 + b_1 X_i$ where parameters b_0 , b_1 and \hat{Y}_i are estimated intercept, slope and mean for Y when the predictor is X_i (i.e., the point on the fitted line) respectively. And residual (e_i) is the residual for the i th case which is the vertical distance from the data point to the fitted regression line i.e. $e_i = Y_i - \hat{Y}_i$ [6]. Ordinary least square (OLS) is usually chosen as estimation technique that minimizes the sum of the squared residuals and is an effort to find best estimators i.e. b_0 and b_1 . Only if the assumptions are true, these estimators are best, linear in errors and unbiased (its expected value is equal to the true value) with minimum variance (i.e. are BLUE). These are also maximum likelihood estimators (MLE).

Logical steps for simple linear regression (SLR) are to check the level of measurement, compute linear regression including descriptive statistics, test significance, statistical inference and power (PWR) of SLR, check direction of the relationship, proportional reduction in error, to test assumptions (linearity, homoscedasticity, independence and normality) and test for the robustness based on validation data in a sequential way [7]. In order to get satisfactory, practicable, and interpretable results from OLS, we need to check the assumptions of regression analysis. When these assumptions are not met, results may not be trustworthy, resulting in a Type I or Type II error, or over- or under-estimation of significance or effect size(s) [8]. There are various assumptions (ASS) for data, variables, functional form and errors. The assumptions for SLR are: (1) Linearity assumption (LN) is that relationship between the dependent and independent variable or functional form is linear. (2) Assumption of independence (IN) of the errors i.e. no autocorrelation or randomness of residuals or $E(\varepsilon_i \varepsilon_j) = 0$ ($i \neq j$). (3) Assumption of normality (NR) of error distribution i.e. $\varepsilon_i \sim N(0, \delta^2)$. This also implies that there are no outlier distortions or influential effects [9]. (4) Assumption of constant variance (δ^2) of errors or homoscedasticity (CV). However, as noted by study [10], only few articles (especially ones using regression analysis) report the statistical testing of these assumptions to draw their research conclusions. This is especially true for literature of social sciences and thus the validity of many of results of these studies, conclusions, and assertions can be called into question. Financial professional and economists also seldom try to test the assumptions especially about whether data is deterministic or stochastic in nature, belongs to a class of elliptical distribution or whether there are sufficient data points to start with or the application

of central limit theorem and assuming that assumption will be true are applicable.

Various measures are in fashion in literature to test the assumptions. Measures to estimate LN include a plot of observed values versus predicted values (LN-T1) or a plot of residuals versus predicted values (LN-T2) where points in the former should be symmetrically distributed around 45 degrees (diagonal) line and around horizontal line respectively. We can visually observe the cautious patterns (i.e. bowed pattern) in the plots to judge the outcome and outcome is ranked by scholar (RBS) on five point scale i.e. from excellent, good, fair, poor, and bad. IN must be checked [4] as autocorrelation inflates the t , r , and f statistics and warps the significance tests. For strong autocorrelation case, IN can be assessed from the plots of residuals vs. time (i.e. look at autocorrelation (AC) plot of the residuals (IN-T1) or cross correlation of residual plots (IN-T2). In former measure, most residual autocorrelation (RAC) should fall within the 95% confidence bands around zero, (located at roughly $[\pm 2 \cdot \text{over-square-root-of-}n]$, where n is the sample size and $L1RAC = RAC$ of lag 1. In IN-T2, some cross correlation for lag values of residual (CC4LVR) are out of ± 1 , out of ± 2 or out of ± 3 standard error lines (SEL). Alternately for all the plots, we can calculate cross correlation coefficient points (CCCP) and then we can check that how many of these points lie within ± 1 standard error lines (SEL) or standard errors of mean (SEM) [11]. And then the percentage number of points (PNP), which lie outside of \pm of considered level of SEL, will be considered to calculate the utility for that ASS measure. If all points lie within ± 1 SEL, then utility is equal to one. The larger PNP outside of ± 1 SEL means less linearity and deviation from linearity and thus disutility can be calculated from minmax formula where min is zero and max is PNP value [12]. And if more than 1% of PNP outside of ± 1 SEL, then utility from this result is zero. For not so strong case of autocorrelation, Durbin-Watson statistic can be calculated and $DW = 2(1 - L1RAC)$ and as a rule of thumb if $DW = 2$ (or to some extent between the range of 1.4 and 2.6), it is considered acceptable [13]. To test NR, we can observe the properties of first four moments of the normal distribution. For the same mean and variance (NR-T1a), normal probability plot can be obtained for the actual percentiles of the residuals vs. the theoretical percentiles of a normal distribution (CCRTN). Ideally, this plot should be a diagonal straight line. Tests for Skewness of normal distribution (NR-T1b) and Kurtosis (NR-T1c) can be done by observing the coefficients of these two. Test for outlier can be done by converting data to z-scores (NR-T2) or getting of studentized deleted residuals (SDRs) (NR-T3) [14]. In former, if the standardized residuals (Z_{e_i}) have values in excess of 3.5 and -3.5, they are outliers. In the later if $[t_i = e_i / \text{MSE}_{(i)}(1-h_{i,i})]$ (using a Bonferroni correction for n tests) $> [t_{n-p-1}(1-\alpha/2n)]$, then outlier [6]. Homoscedasticity or CV should be tested as though slight heteroscedasticity has little effect on significance tests, however, when heteroscedasticity is marked, it can lead to serious distortion of findings and seriously weaken the analysis thus increasing the possibility of a Type I error [15] and [16]. CV can be assessed by getting either plots of residuals vs. time (CV-T1) or plots of residuals vs. predicted values (CV-T2). In CV-T1, cautious sign is that there is a tendency for the errors to get larger in magnitude

over time and we can assess it by calculating cross correlation of residuals with time (CCRWT). In CV-T2, we check for tendency for the errors to change in magnitude vis-à-vis different predicted values. It can be assessed through cross correlation of residuals with Predicted values (CCRWP). Third test can be a plot of standardized residuals (Z_{e_i}) against \hat{Y} (CV-T3) and we can assess it through cross correlation of Z_{e_i} with \hat{Y} (CCZWP). The rules for relative weights of assumptions are summarized in Table I.

TABLE I: ASSUMPTIONS BASED RULES FOR UTILITY FUNCTION FOR SLR

Test & [TS]	Rules to apply Kramer style Min-Max function for scoring	E-1 [E-2]
LN-T1 [\sum LN=1]	If RBS, then ranking & If no RBS or TNP then = 0 i.e. Bad	0 [0]
LN-T2	If RBS, then ranking & If no RBS then = 0 i.e. Bad	0 [0]
IN- T1 [\sum IN=1]	If [L1RAC $\leq \pm 0.2$] then =1 & If [L1RAC is between ± 0.2 and ± 0.4] then use CBUF MinMax & If [L1RAC > ± 0.4] or if TNP then =0	0 [0]
IN- T2 [\sum IN=1]	If [CC4LVR $\leq \pm 1$ SEL] then =1 & If [PNP of CC4LVR are between ± 1 and ± 3 SEL] then = CBUF MinMax & If all [$\geq 1\%$ of PNP of CC4LVR $\geq \pm 3$ SEL] or if TNP then = 0	0 [0]
IN- T3 [\sum IN=1]	If [DW ≥ 1.9 to ≤ 2.1] then = 1 & If [DW is from < 1.9 & > 2.1 to ≤ 1 and ≥ 3] then = CBUF Minmax & If DW is either < 1 or > 3 or if TNP then = 0	0 [0]
NR-T1a [\sum NR=1]	If [CCRTN $\leq \pm 1$] then =1 & If [PNP of CCRTN between > ± 1 to $\leq \pm 3$] then use CBUF MinMax & If [$\geq 1\%$ of PNP of CCRTN > ± 3] or if TNP then = 0	0 [0]
NR-T1 b [\sum NR=1]	If [s = 0] then =1 & If [3 $\geq s > 0$] then = CBUF Minmax & If [s > 3] or if TNP then =0	0 [0]
NR-T1c [\sum NR=1]	If [k = 3] then = 1 & If [6 $\geq k > 3$] then = CBUF Minmax & If [k > 6] or if TNP then = 0	0 [0]
NR-T2 [\sum NR=1]	If [$Z_{e_i} \leq \pm 1$] then = 1 & If [$\pm 1 < Z_{e_i} \leq \pm 3.5$] then = CBUF Minmax & If [$Z_{e_i} \geq \pm 3.5$] or if TNP then = 0	0 [0]
NR-T3 [\sum NR=1]	Minmax with min limit of zero is when [$t_i = e_i / \text{MSE}_{(1-h_{i,i})}$] divided by [$tn-p-1(1-\alpha/2n)$] and max limit is the max value of [$t_i = e_i / \text{MSE}_{(1-h_{i,i})} / [tn-p-1(1-\alpha/2n)]$]; & If TNP then = 0	0 [0]
CV-T1 [\sum CV=1]	If [CCRWT $\leq \pm 1$ SEL] then =1 & If [PNP of CCRWT change from $\leq \pm 1$ SEL to $\geq \pm 3$ SEL] then = CBUF MinMax & If [$\geq 1\%$ PNP of CCRWT > ± 3] or TNP then = 0	0 [0]
CV-T2 [\sum CV=1]	If [CCRWP $\leq \pm 1$] then =1 & If [PNP of CCRWP change from $\leq \pm 1$ to $\geq \pm 3$] then = CBUF MinMax & If [$\geq 1\%$ PNP of CCRWP > ± 3] or TNP then = 0	0 [0]
CV-T3 [\sum CV=1]	If [CCZWP $\leq \pm 1$] =1 & If [CCZWP change from $\leq \pm 1$ to $\geq \pm 3$] then = CBUF MinMax & If [$\geq 1\%$ PNP of CCZWP > ± 3] or TNP then = 0	0 [0]
Notes: (1) Scores of examples E-1 & [E-2] are in last column; (2) Relative Score for each ASS category should sum to 1 (3) relative Total Score (TS) of all ASSs should also sum to 1 (4) Continuous band utility function (CBUF) (of study [21]) (5) Test not Provide (TNP)		

In trying to ascertain the reliability of error measures in the estimation period, we need to look at the issue of over-fitting. If the model has only one or two parameters (e.g. random walk, exponential smoothing, or simple linear regression) and it was fitted to a moderate to large sample of data (i.e. ≥ 30 observations), then model is probably unlikely to have over-fitted the data. Besides assumptions, various other diagnostic tests are conducted to assess the effectiveness of the regression analysis. In order to test the effectiveness of

distribution tests (e.g. t-test or F-test), we use either p-value or power analysis. We use t-test to test that the coefficients (b_0 and b_1) are true representative of the population; we need to construct confidence intervals and do hypothesis significance tests for the mean of a normal distribution, based on the statistics of a random normal sample.

For the significance tests, p -value is the probability of observing a test statistic at least as extreme as what was actually observed, when the null hypothesis is really true. We reject H_0 , if $p \leq \alpha$ (where $p = \text{Prob}(|t_{n-2}| > |t|)$). This is same as $|t| \geq t_c$ where $t_c = t_{n-2}(1-\alpha/2)$ and $n-2$ is degrees of freedom (df) and in this case we accept H_A which means that there is sufficient evidence in the data to conclude that H_0 is false, and so we assume that H_A is true and vice versa [6]. The rules for utility function (i.e. for CBUF) or RUF for p -value test (PVT) can be that if [PVT $\geq 99\%$] then =1 and If [PVT is between 90% and 99%] then = MinMax & If [PVT < 90%] or if test not provided (TNP) then it scores zero. Note that a failure to reject H_0 does not mean that there was any evidence in favor of H_0 . The power of a significance test is the probability that the null hypothesis will be rejected when, in fact, it is false (Type I error). As statistical power increases, the chances of a Type II error decrease. Power analysis is appropriate when the concern is with the correct rejection, or not, of a null hypothesis. How can we avoid making Type I error is either by applying a higher threshold of stringency to reject a hypothesis in order to compensate for the multiple comparisons being made (e.g. as in the Bonferroni method) or by ignoring the statistical power (p -value) altogether and basing our interpretation on the tiny effect size (or substantive power or significance) as suggested by [17]. In statistics, an effect size is a measure of the strength of the relationship between two variables in a statistical population, or a sample-based estimate of that quantity. The effect size does not determine the significance level (or p -value), or vice-versa. Effect sizes complement inferential statistics such as p -values and larger the “size effect,” the smaller the p -value will be, other things being equal [18]. The calculation of power of β_1 is provided in [6]. The rule for power of a test can be as: use continuous band utility function (CBUF) (of study [21]) based MinMax if test provided (TP) & If TNP then = 0. To test the overall effectiveness (OE) of the model, typically analysis of variance (ANOVA) reports for model, errors and of corrected total list degrees of freedom (df), sum of squares (total, model and error), root mean square error (RMSE), coefficient of variance (r), R^2 and F-test. However, when different sets of observations are used as estimation period, R-squared is not a reliable guide to model quality [19]. In order to test robustness, the estimation of $E(Y_h) [= \beta_0 + \beta_1 X_h]$ i.e. mean value of Y for the new subpopulation with $X = X_h$, we will calculate $Y_h = b_0 + b_1 X_h$. And then the parameters from the original regression estimates are compared with the test data results or simulated data results. Thus we define robustness (RO) score for test data (RSTD) as (Coefficient of validation data regression (VDR)) minus (Coefficient of estimated regression (ER)). In SLR, there are two coefficients α and β and thus two ROs i.e. α -RO and β -RO. As the smaller RO is better, in order to make it bigger is better so that we can use in final score summation equation, we will use maxmin (MM) function. Maxmin is equal to difference between realized value and maximum value divided by the difference between minimum value and maximum value. Finally compute the “average RO” (ARO) of MM- α -RO and

MM- β -RO for each alternative regression. Robustness score for simulated data (RSSD) replaces validation data with simulated data. Besides, model should pass various quantifiable diagnostic tests or eyeball tests. Typically, these QCs may include keep it simple...(K.I.S.S), sign (SOC) and/or magnitude (MOC) of coefficients based on theoretical considerations, intuitive reasonableness of model appearance of forecast plots, confidence interval for coefficients (CIC) and even Working-Hotelling confidence bands for entire regression line [4] and are based on researchers' judgments. Besides, there can be other tests for efficiency, consistency, other assumptions, parameters, and data characteristics which can be performed and tested but has been excluded for simplicity.

TABLE II: QUALITATIVE & DIAGNOSTIC (Q&D) BASED RULES FOR CBUF

Q&D Tests	Rules for Utility Function (RUF)	E-1 [E-2]
Tests for Parameters Effectiveness (i.e. SI)		1 [1]
t-value	If $ TV = 2 $ then = 1 & If $ TV $ is between $ 2 $ & $ 1.5 $, then = MinMax & If $ TV < 1.5 $ or if TNP then = 0	α : 0 [0] β : 1 [1]
Tests for Overall Effectiveness (i.e. OE)		0 [1]
F-test	Use the rule for p-values (described in text)	0 [1]
RMSE	CBUF MinMax if TP; If TNP then = 0	TNP
R ²	CBUF MinMax if TP; If TNP then = 0	TNP
Test for robustness (i.e. RO)		0 [0]
RSTD	Maxmin if TP; If TNP then = 0	TNP
RSSD	Maxmin if TP; If TNP then = 0	TNP
Test for Qualitative Considerations (i.e. QC)		1 [0.95]
SOC	If RBS, then ranking & If no RBS then = 0	α : 5[5] β : 5[5]
MOC	If RBS, then ranking & If no RBS then = 0	α : 5[4] β : 5[5]
CIC	If RBS, then ranking & If no RBS then = 0	TNP
WHCB	If RBS, then ranking & If no RBS then = 0	TNP
Notes: (1) Scores of examples E-1 & [E-2] are in last column; (2) Relative Score for each Q&D Test category should sum to 1. (3) Grand Total Score (GTS) of all ASSs plus Q&D Test categories should be 5 i.e. $\sum [\sum ASSs + \sum SI + \sum OE + \sum RO + \sum QC] = 5 = GTS$; (6) In QC case, we have assumed equal weights for all tests within QC category. SOC and MOC for α and β sum to one. Here SOC for α and β is 10/10 [10/10] i.e. 1 [1] & MOC for α & β is 10/10 [9/10] i.e. 1 [0.9] & total score for QC = $\frac{1}{2}(1) + \frac{1}{2}(1) [\frac{1}{2}(1) + \frac{1}{2}(0.9)] = 1 [0.95]$		

III. METHODOLOGY

Previous section completes our discussion of sub-criteria (SC) within five global criteria (GC). In statistical software, an option can be provided to get input RBS inputs and can be considered as further evolution [3] in software engineering evolution stages. The values for either all or some of the SC can be obtained, the more SCs are used, the better final evaluation will be. At least one SC must be used for each GC category and if more are used within each GC, then either they are can be weighted equally or be given weights based on Analytical Hierarchy Process (AHP) or some other weighting scheme can be used. And finally these SC will be normalized based on a utility function from study [20] and continuous band utility function (CBUF) of study [21] in order to avoid the problems of differences in measurement units. CBUF takes the value of either 1 (for ideal range of values or over) or zero (for values below the least ideal (or lowest) boundary i.e. has 100% disutility) or in between these two ranges with the help of minmax function for all sub-criteria considered. Finally if any of the tests is not provided (i.e. TNP), then the utility in this case will be zero.

Before we can calculate final scores for regressions, we need to perform fuzzy AHP pair wise comparisons for five GC which are assumptions, parameters' effectiveness, robustness, error measures, and quality considerations. Normal or crisp pair-wise comparison are based on Saaty's 1-9 scale are provide in table II below as Crisp pair-wise comparison matrix values (CPCM). AHP is a method of converting a multi-dimensional problem to single dimension by building hierarchy by providing pair wise weights to criteria. However, AHP has been criticized because of its imprecision and subjectivity [22]. This can be improved with the help of fuzzy AHP as it incorporates decision makers (DM) uncertainty. Conversion of CPCM values to Fuzzy PCM values [23] has been given in table III. From a range of fuzzy AHP, DM can choose the values that reflect her confidence for the GC (i.e. ASS, SI, OE, RO, and QC).

TABLE III: CONVERSION OF CRISP (OR NORMAL) PCW MATRIX TO FUZZY PCW MATRIX

CPCM	Fuzzy PCMV	CPCM	Fuzzy PCMV
1	(1,1,1) if diagonal; (1,1,3) otherwise	1/1	(1/1,1/1,1/1) if diagonal; (1/3,1/1,1/1) otherwise
2	(1, 2, 4)	1/2	(1/4, 1/2,1/1)
3	(1,3, 5)	1/3	(1/5, 1/3,1/1)
5	(3, 5, 7)	1/5	(1/7, 1/5,1/3)
7	(5,7,9)	1/7	(1/9, 1/7,1/5)
9	(7,9,11)	1/9	(1/11, 1/9,1/7)

Two parameters in fuzzy reasoning method are fuzzy membership curve (FMC) showing the confidence level of the DM (cab be triangular, Gaussian and trapezoidal) and slope of FMC (which represents three different attitudes of the DM i.e. liberal, moderate and strict). Triangular FMC (used in this study) is simple, commonly used, can approximate non-triangular as well and is especially used when DM has least confidence (i.e. is very uncertain) about degree of membership of values to particular class. Liberal attitude has gentle slope and gives wider range of input values as a member of certain class while strict attitude has a steep slope with narrow range of values. Moderate one is in between the two and we will consider this in this study. The Fuzzified PCM is shown in Table IV.

Fuzzy AHP weights for different criteria has been calculated from table IV by use of fuzzy extent analysis [23] as: Left Fuzzy PCW (b1) = $(1+1+3+5+7+0.2+1+1+3+5+0.1428+0.25+1+1+3+0.1111+0.1428+0.2+1+1+0.0909+0.1111+0.1428+0.2+1) = 36.5915$. Likewise Middle Fuzzy PWC (b2) and Right Fuzzy PWC (b3) can be calculated and values are 56.4968 and 79.5428 respectively. Thus total sum for whole fuzzy PWC matrix is (b1, b2, b3) & sum of row of satisfactory assumptions = First row Left (FRL) + First row Middle (FRM) + First row Right (FRR) is calculated as: FRL = $a_1 = (1+1+3+5+7) = 17$; FRM = $a_2 = (1+3+5+7+9) = 25$; FRR = $a_3 = (1+5+7+9+11) = 33$. Fuzzy performance of Satisfactory ASS (FPSA) = [FPSA Left (=a1/b3), FPSA Middle (= a2/b2), FPSA Right (= a3/b1)]. Normalized FPSA for Statistical Inference (NFPSI) i.e. for five GC has been provided in Table V.

TABLE IV: FUZZY AHP CRITERIA AND FUZZIFIED PAIR-WISE COMPARISON MATRIX FOR SLR

	ASS	SI	OE	RO	QC
ASS	(1,1,1)	(1,3,5)	(3,5,7)	(5,7,9)	(7,9,11)
SI	(0.2,0.333,1)	(1,1,1)	(1,2,4)	(3,5,7)	(5,7,9)
OE	(0.1428,0.2,0.3333)	(0.25,0.5,1)	(1,1,1)	(1,3,5)	(3,5,7)
RO	(0.1111,0.1429,0.2)	(0.1429,0.2,0.3333)	(0.2,0.333,1)	(1,1,1)	(1,3,5)
QC	(0.0909,0.1111,0.1429)	(0.1111,0.1429,0.2)	(0.1428,0.2,0.3333)	(0.2,0.333,1)	(1,1,1)

Notes: 1. ASS = Assumptions; SI = Statistical Inference; OE = Goodness of Fit and Error Measures (F, RMSE, R2); RO = Robustness (Validation – estimated); QC = Quality Considerations (i.e. sign of coefficients, model etc)

Finally, CBUF values for all sub-criteria based on tables I and II will be multiplied with NFPSI for global criteria and we get a final score to rank regressions (FSRR). This method of comparison of various regressions or the quality of single regression based on the highest FSRR is called Fuzzy-AHP Intelligent Zax (FAIZ) approach. Remember that there is no “right answer” when analyzing regression data - there are only good and bad analyses [4] and the purpose of this approach is to separate good from bad model based on a linearized scores derived from multi-dimensional (non-linear) diagnostics. “All models are wrong, but some are useful” ([4] has quoted George Box) and we simply want a good proxy for the true, unknown model with the help of FAIZ approach. The best regression identified, out of various regression alternatives obtained from different sets of random samples of data of the same problem, with the help of this fuzzy ranking process is named as FAIZ Regression.

TABLE V: NORMALIZED FUZZY PERFORMANCES FOR GLOBAL CRITERIA

	Left	Middle	Right
ASS	0.4646	0.4425	0.4149
SI	0.2787	0.2714	0.2766
OE	0.1474	0.1717	0.1802
RO	0.0672	0.0828	0.0947
QC	0.0422	0.0316	0.0336

Note: Decision maker can choose either left, middle or right from Normalized (Norm) depending on her/his choice based on fuzzification assumptions

TABLE VI: TWO MARKET RISK (B) REGRESSIONS FOR CHILE [24]

Data Period	α (t-value)	β (t-value)	F-test (p-value)	SOC for (α & β)	MOC for (α & β)
1995-2008	0.0001 (-0.04)	0.4599 (-11.38)	132.26 (<90%)	10/10	10/10
2000-2008	0.00392 (-1.19)	0.32677 (-6.63)	43.97 (>99%)	10/10	9/10

IV. SIMPLE EXAMPLE AND CONCLUSION

The results of Beta estimation case using OLS for Chile

from study [24] for two periods (Total as E-1 and Second sub-period as E-2) has been provided as an example in Table I and Table II as well. In order to calculate market risk (i.e. β), monthly index returns adjusted for dividend of Chile has been regressed on FTSE’s emerging market price index provided in Data Stream. The information provided in [24] about these two regressions is given in Table VI.

We can get the scores for these two regressions (table VI) against the total maximum score of 5 which is obtained from addition of maximum score of 1 for each criteria category. From Table I and Table II, the category scores (CS), are as: ASS {0[0]}, SI {1 [1]}, OE {0 [1]}, RO {0 [0]} and QC {1 [0.95]}. As we may not be fully certain or completely uncertain about the criteria that we are using for each category of criteria, thus we choose the middle value of normalized fuzzy performance for global criteria (NFPGC) from Table V. Multiplying NFPGC with CS for ASS, SI, OE, RO, and QC gives us the GS or FAIZ scores, which are 0.303 and 0.47312 for E-1 and E-2 respectively. Thus E-2 is a superior regression and market risk β from E-2 can be more reliable measure.

REFERENCES

- [1] L. Breiman, “Statistical modeling: two cultures,” *Statistical Science*, Institute of Mathematical Statistics, vol. 16, no. 3, pp. 199-215, 2001.
- [2] D. S. Riggs, J. A. Guarnieri, and S. Addelman, “Fitting straight lines when both variables are subject to error,” *Life Sciences*, vol. 22, 1978.
- [3] Z. Wang, X. Liu, K. Chalmers, G. Cheng, and H. Wei, “Evolution feature oriented model driven product line engineering approach for synergistic and dynamic service evolution in clouds,” *Journal of Industrial and Intelligent Information*, vol. 1, no. 1, March 2013.
- [4] T. P. Ryan, “Modern regression analysis for scientists and engineers,” presented at National Institute of Standards of Technology (NIST) as external training course, Gaithersburg, MD, 2003.
- [5] W. E. Ricker, “Computation and uses of central trend lines,” *Canadian Journal of Zoology*, vol. 62, pp. 1897-1905, 1984.
- [6] K. Jennings, “Applied regression analysis,” presented at Department of Statistics as Class notes for Statistics 512, Purdue University. IN. 2013.
- [7] A. J. Schwab, “Course material: Data Analysis and Computers I,” presented at School of Social Work, University of Texas at Austin, fall 2006.
- [8] J. Osborne and E. Waters, “Four assumptions of multiple regression that Researchers should always Test, *Practical Assessment, Research, & Evaluation*, vol. 8, no. 2, 2002.
- [9] R. A. Yaffee, “Regression analysis with SPSS,” presented at NYU, 2004.
- [10] J. W. Osborne, W. R. Christensen, and J. Gunter, “Educational psychology from a statistician’s perspective: A review of the power & goodness of educational psychology research,” presented at national meeting of American Education Research Association, Seattle, 2001.
- [11] J. H. McDonald. (2009). Handbook of Biological statistics. [Online], Available: <http://udel.edu/~mcdonald/statintro.html>
- [12] M. J. Campbell and A. Tobias, “Causality and temporality in the study of short-term effects of air pollution on health,” *International Journal of Epidemiology*, vol. 29, pp. 271-273, 2000.
- [13] R. Nau. (2013). Decision models – forecasting. *A course presented at the Fuqua School of Business*. Duke University, Durham. NC. [Online]. Available: <http://people.duke.edu/~rnau/testing.htm>
- [14] R. M. O'Brien, “A caution regarding rules of thumb for variance inflation factors,” *Quality and Quantity*, vol. 41, pp. 673-690, 2007.
- [15] W. D. Berry and S. Feldman, “Multiple regression in practice,” *Sage University Paper Series on Quantitative Applications in the Social Sciences*, series no. 07-050, Newbury Park, CA, 1985.
- [16] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 3rd ed., Harper Collins College Publishers, NY, 1996.
- [17] P.D. Ellis. (July 2010). The essential guide to effect sizes: statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge University Press. [Online], Available: <http://effectsizefaq.com/refs/>
- [18] K. O. McGraw and S. P. Wong, “A common language effect-size statistic,” *Psychological Bulletin*, vol. 111, pp. 361-365, 1992.
- [19] Karen. (2008). Assessing the Fit of Regression Models. *Analysis Factor: Making Statistics Make Sense*. [Online]. Available: www.analysisfactor.com

- [20] B. Kramer, D. Kronbichler, T. van Welie, and J. Putman, "comparing social objectives for decision-making in housing corporations," *Applied Working Paper no. 2008-02*, Ortec Finance Research Center (OFRC) working paper, Rotterdam, May, 2008, pp. 1-22
- [21] S. Anjum. (July 2013). Algorithms for predictive classification in data mining: A comparison of evaluation methodologies. *Journal of Industrial and Intelligent Information (JIII)*, [Online]. 1(2). pp. 116-120. Available: <http://www.jiii.org/>
- [22] T. L. Saaty, *Theory and Applications of the Analytic Network Process*, RWS Publications, Pittsburgh, 2005, pp. 5-49
- [23] R. Kusawandari, "Assessment of Different Methods for Measuring the Sustainability of Forest Management," M. Sc. Dissertation, International Institute of Geo-Information Science and Earth Observation, Enschede, The Netherlands, ITC, March 2004
- [24] S. Nathaphan and P. Chunhachinda, "Estimation risk modeling in optimal portfolio selection: An empirical study from emerging markets," *Research Article, Economics Research International*, Hindawi Publishing Co., vol. 1, pp. 1-10, 2010
- Shahid Anjum** is an adjunct faculty member (Finance and MIT) and has been teaching Financial Valuation course to Doctor of Business Administration (DBA) program at College of Management, LTU, (USA), besides working for one of the leading financial institutions in Canada. He has earned his Ph. D. for his research on application of data mining for the early warning system of financial crisis from GSID, Nagoya University (NU), Japan. He has studied in the Master of Computer Science at Dept. of Mathematics and Computer Science, School of Engineering, University of Detroit Mercy (UDM), MI., USA. He is M. Phil. Economics, a senior member of International Association of Computer Science and Information Technology (IACSIT), an associate member of Global Association of Risk Professionals (GARP) and Canadian Securities Institute (CSI). He has worked in Japan, USA, Canada, and as an intern at Asian Development Bank (ADB), Manila, Philippines. His papers has been accepted in International Conferences of Information and Financial Engineering (ICIFE-2013, Indonesia), Databases and Data Mining (ICDDM-2013, Korea) and Information and Social Sciences (ISS-2013, Nagoya, Japan). His publications has appeared in Journal of Industrial and Intelligent Information (JIII), Pakistan Development Review (PDR) and Journal of Economics, Business and Management (JOEBM).