REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

APPLICATIONS OF MEASURES OF STATISTICAL ENTROPY TO HETEROSCEDASTICITY PROBLEMS IN LINEAR REGRESSION MODELS

Hatice Çiğdem ÇELİK

MASTER OF STATISTICS THESIS

Department of Statistics

Statistics Program

Supervisor

Assoc. Prof. Dr. Atıf Ahmet EVREN

December, 2022

REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

APPLICATIONS OF MEASURES OF STATISTICAL ENTROPY TO HETEROSCEDASTICITY PROBLEMS IN LINEAR REGRESSION MODELS

A thesis submitted by Hatice Çiğdem ÇELİK in partial fulfillment of the requirements for the degree of MASTER OF STATISTICS is approved by the committee on 29.12.2022 in Department of Statistics, Statistics Program.

Assoc. Prof. Dr. Atıf Ahmet EVREN
Yıldız Technical University
Supervisor

Approved By the Examining Committee

Assoc. Prof. Dr. Atıf Ahmet EVREN, Supervisor
Yildiz Technical University ____________________

Assoc. Prof. Dr. Gülder KEMALBAY, Member
Yildiz Technical University ____________________

Assist. Prof. Dr. Erhan USTAOĞLU, Member
Marmara University ____________________

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of Applıcatıons Of Measures Of Statıstıcal Entropy To Heteroscedastıcıty Problems In Lınear Regressıon Models supervised by my supervisor, Doç. Dr. Atıf Ahmet EVREN. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Hatice Çiğdem ÇELİK

Signature

Dedicated to my family

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Assoc. Prof. Dr. Atıf Ahmet EVREN for his guidance, effort and support. He has been a source of inspiration and motivation for me.

I would want to express my gratitude to all my teachers for their help.

Finally, I would like to express my sincere appretiation and thanks to my husband, brothers and parents for their understanding, patience and support.

Hatice Çiğdem ÇELİK

TABLE OF CONTENTS

LIST OF SYMBOLS **viii**
LIST OF ABBREVIATIONS **ix**
LIST OF FIGURES **x**
LIST OF TABLES **xi**
ABSTRACT **i**
ÖZET **xiv**
1 INTRODUCTION **1**
1.1 Literature Review 1
1.2 Objective of Thesis 1
1.3 Hypothesis 1
2 REGRESSION ANALYSIS **2**
2.1 Least Squares Estimation 2
2.2 Precision of Estimates 4
2.3 Properties of Least-Squares Estimators: The Gauss–Markov Theorem 6
2.4 The Coefficient Of Determination R: A Measure Of “Goodness Of Fit” 7
2.5 Hypothesis Testing 10
3 MATRIX NOTATION **11**
3.1 Fundamental Definitions 11
3.2 Specialized Matrix Types 11
3.3 Operations of Matrix 12
3.4 Multiple Regression in Matrix Notation 13
3.5 The Solution to the Normal Equations 14
3.6 The Y̑ & Residuals Vectors 15
3.7 Characteristics of Random Vectors’ Linear Functions 15
4 HETEROSCEDASTICITY TESTS **16**
4.1 White’s General Heteroscedasticity Test 16
4.2 Park Test 17
4.3 Weighted Least Squares 18
4.4 Box-Cox Transformation 20
5 ENTROPY **22**
5.1 Shannon Entropy 22

5.2 Rényi Entropy 22
5.3 Tsallis Entropy 23
6 APPLICATION **24**
7 RESULTS AND DISCUSSIONS **28**
REFERENCES **29**
APPENDIX **29**
PUBLICATIONS FROM THE THESIS **34**

LIST OF SYMBOLS

$Cov(U,W)$	Covariance between U and W
r^2	Coefficient of Determination for two-variable case
R^2	Coefficient of Determination for multiple regression
H_R	Rényi Entropy
H_T	Tsallis Entropy
$Var(\widehat{H}_R)$	Variance of Rényi Entropy
$Var(H_T)$	Variance of Tsallis Entropy

LIST OF ABBREVIATIONS

ESS	Explained Sum of Squares
OLS	Ordinary Least Squares
RSS	Residual Sum of Squares
TSS	Total Sum of Squares
WLS	Weighted Least Squares

LIST OF FIGURES

Figure 2.1 Correlation patterns 9
Figure 6.1 The scatter diagram of data-1... 25
Figure 6.2 The scatter diagram of data-2... 25
Figure Apx.1 Data.. 30
Figure Apx.2 Data (still) 31
Figure Apx.3 Data (still) 32
Figure Apx.4 Data (still) 33

LIST OF TABLES

Table 5.1 The entropy formulas for distributions 23
Table 6.1 Goodness of fit statistics model for data-1 26
Table 6.2 Goodness of fit statistics model for data-2 26
Table 6.3 Goodness of fit statistics model for data-3 26
Table 6.4 Goodness of fit statistics model for data-4 27

ABSTRACT

Applications of Measures of Statistical Entropy to Heteroscedasticity Problems in Linear Regression Models

Hatice Çiğdem ÇELİK

Department of Statistics

Master of Statistics Thesis

Supervisor: Assoc. Prof. Dr. Atıf Ahmet EVREN

In simple regression analysis, the dependent variable is assumed to have constant variance at various degrees of independent variable. Whenever the assumption of constant variance fails, some remedies like the weighted least squares, or Box-Cox transformation may be helpful. Weighted least squares technique rescales the independent variable by weights to preserve constant variance, unlike the Box-Cox approach, which is based on a transformation of the dependent variable nonlinearly. Although there is already a large literature on this issue, determining the weights seems a major problem. In this study, the weights are alternatively calculated by entropy approach, since statistical entropy, and variance conveys similar information about a probability distribution. In this study, by exploiting the normality assumption of linear models, the weights are determined by the reciprocals of Shannon, Tsallis and Renyi entropies of normal distribution. The weighting procedure has been applied on some simulated data having nonconstant variance. In some applications we have shown that weighting by Tsallis and/or Rényi entropies produced better goodness of fit results in terms of coefficient of determination, and the mean square error.

Keywords: Heteroscedasticity, entropy, weighted least squares, Shannon entropy, Tsallis entropy

ÖZET

İstatistiksel Entropiye Dayalı Ölçülerin Doğrusal Regresyon Modellerindeki Değişen Varyans Problemlerine Uygulanması

Hatice Çiğdem ÇELİK

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Assoc. Prof. Dr. Atıf Ahmet EVREN

Basit regresyon analizinde, bağımlı değişkenin farklı bağımsız değişken düzeylerinde sabit değişkenlik olduğu varsayılır. Sabit değişkenlik varsayımı başarısız olduğunda, ağırlıklı en küçük kareler veya Box-Cox dönüşümü gibi bazı çözümler yararlı olabilir. Box-Cox yaklaşımı bağımlı değişkene bağlı doğrusal olmayan bir dönüşümü temel alsa da ağırlıklı en küçük kareler yönteminde, bağımsız değişken sabit değişkenliği korumak için ağırlıklarla yeniden şekillendirilir. Bu konuda zaten büyük bir literatür olmasına rağmen, ağırlıkların belirlenmesi büyük bir sorun gibi görünüyor. Bu çalışmada, istatistiksel entropi ve değişkenlik olasılık dağılımı hakkında benzer bilgiler aktardığı için, ağırlıklar alternatif olarak entropi yaklaşımıyla hesaplanır. Bu çalışmada, doğrusal modellerin normallik varsayımından faydalanılarak ağırlıklar, normal dağılımın Shannon, Tsallis ve Renyi entropilerinin karşılıklarıyla belirlenir. Ağırlık prosedürü, sabit olmayan değişkenlik içeren bazı simüle edilmiş verilere uygulanmıştır. Bazı uygulamalarda Tsallis ve/veya Rényi entropilerinin ağırlığının, belirleme katsayısı ve ortalama kare hatası açısından daha iyi uyum sonuçları üretmiş olduğunu gösterdik.

Anahtar Kelimeler: Shannon Entropisi, Tsallis Entropisi, Renyi entropisi, değişen varyans problemi, Entropi

1
INTRODUCTION

1.1 Literature Review

One of the main assumptions of linear regression model is homoscedasticity. In other words, the variances of dependent variable at each level of independent variable/variables are the same. Heteroscedasticity, or nonconstancy of variances can be detected alternatively. The simplest way is to use visual representation of data, although relying only on visual means may be misleading. Therefore some statistical tools like White Test, or Park's Test are used.

1.2 Objective of Thesis

Once the heteroscedasticity is detected, to increase the goodness of fit, a kind of weighting procedure is followed. One way is to give more weights to the levels of independent variable/variables whose variabilities are lower. The Box-Cox transformation can also be used to avoid heteroscedasticity.

1.3 Hypothesis

Alternatively, although it's not in common usage, weighting with the reciprocal of entropy can be used. Shannon, Renyi and Tsallis are different types of entropy, and all of them could be used for weighing the data. Indeed, Rényi and Tsallis formulations of entropy bring more flexibility than Shannon entropy, due to their parametric nature.

2
REGRESSION ANALYSIS

A constant and one independent variable make up the simplest linear model.

$$\mathcal{E}(Y_i) = \beta_0 + \beta_1 X_i \tag{2.1}$$

β_0 represents the interception and β_1 represents the slope of the line. The real Y_i can be found by adding an error item to the actual mean of Y_i can be seen below.

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{2.2}$$

The observed X_i's product X which is a group of known constant numbers. The paired observations Y_i and X_i are both measured on each observational unit.

The random errors e_i's are zero mean and common variance σ^2 distributed and to be pairwise independent. It's shown as $e_i \sim N(0, \sigma^2)$, where N means "normally and independently distributed. The parenthesis, respectively, contain the normal distribution's mean and variance.

2.1 Least Squares Estimation

β_0 and β_1 are two parameters of the simple linear model. These are estimated from the data.

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be numerical estimates of the parameters β_0 and β_1, respectively, and let

$Y_i = \beta_0 + \beta_1 X_i$ be the estimated mean of Y for each X_i, $i = 1, \ldots, n$. The least squares principle chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the sum of squares of the residuals,

$$\text{SS(Res)} = \sum_{i=1}^{n}(\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i})^2 = \sum e_i^2 \tag{2.3}$$

We can write $e_i = (Y_i - \hat{Y}_i)$ since e_i is the difference between observed data and the real data. The sum shown by $\sum$ is on all observations of the dataset as illustrated by the summing indicator, i = 1 to n. The aim is to minimize SS(Res) and in order to reach the aim the estimators which are β_0 and β_1 are obtained by using calculus. By derivating the SS(Res) in respect to β_0 and β_1, we will obtain two normal equations. The equation has two unknowns and two equations is called as normal equations. Solving these equations for $\hat{\beta}_0$ and $\hat{\beta}_1$. We will find $\hat{\beta}_0$ and $\hat{\beta}_1$which are estimators of β_1 and β_0 as

$$\hat{B}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = x_i y_i \frac{\sum X_i Y_i}{\sum X_i^2} \quad (2.4)$$

$$\hat{B}_0 = \bar{Y} - \beta_1 \bar{X}. (1.7) \quad (2.5)$$

When $\hat{\beta}_0$ and $\hat{\beta}_1$ equations combined with x, and y, equations, we obtaion more favorable format as below:

$$\sum x_i^2 = \sum X_i^2 - \frac{(\sum Xi)^2}{n} \quad (2.6)$$

$$\sum x_i y_i = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} \quad (2.7)$$

Thus, we obtain the below equation for the slope by computing

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{\sum X_i^2}{n}} \quad (2.8)$$

The computed parameters provide the regression equation shown below.

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (2.9)$$

According to the equation above, the gaps between the predicted values produced by the regression equation and the observed values are known as residuals. The equation $Y_i = \hat{Y}_i + e_i$ is valid for each actual value.

$SS(Total_{uncorr}) = \sum Y_i^2$, can be similarly partitioned. In order to generalize the above formula for all observations, substitute $Y_i + e_i$ for each Y_i and expand the square. Thus,

$$\sum Y_i^2 = \sum(\hat{Y}_i + e_i)^2 \quad (2.10)$$

$$SS(Total_{uncorr}) = \sum \hat{Y}_i^2 + \sum e_i^2 \quad (2.11)$$

$$= SS(Model) + SS(Res) \quad (2.12)$$

$$SS(Model) = n\bar{Y}^2 + \hat{\beta}_1^2 \sum(X_i - \bar{X})^2$$

$$SS(Res) = SS(Total\ uncorr) - SS(Model) \tag{2.13}$$

$$\sum y_i^2 = \hat{\beta}_1^2 \sum(X_i - \bar{X})^2 + \sum e_i^2$$

$$= SS(Regr) + SS(Res) \tag{2.14}$$

Then $\sum y_i^2$ has found. The y 's adjusted sum of squares is now specified with $SS(Total)$.

The determinant's coefficient, represented by R^2, is one indicator of the contribution of the model's independent variables.

$$R^2 = \frac{SS(Regr)}{\sum y_i^2} \tag{2.15}$$

The square of the association between Y_i and $\hat{Y}_i$ constitutes the determinant's coefficient, which ranges from zero to one.

2.2 Precision of Estimates

We compute $\bar{Y}$, $\hat{Y}$, e, $\hat{\beta}_0$, and $\hat{\beta}_1$ from the Y_i. Since Y_i is a random variable, the computed ones also a random variable. When evaluating the accuracy of the estimations, variances or standard errors are utilized.

The description of the variance in a linear regression model may be used to identify the variances of estimated regression coefficients, $\hat{Y}_i$, and residuals.

Let $U = \sum a_i Y_i$ then

$$Var(U) = \sum a_i^2\, Var(Yi) + \sum \ \sum_{i \neq j} a_i a_j\, Cov(Y_i, Y_j) \tag{2.16}$$

Cov(·,·) reflects the covariance between the variables shown in the parentheses. When one of the variables changes, covariance demonstrates the impact of the other. When these variables are independent, the covariances must be zero. If each random variable's variance is the same.

$$Var(U) = (\sum a_i^2)\sigma^2 \tag{2.17}$$

Each Y_i in the sample mean has a coefficient called a_i of $1/n$. Equation (1.20) is applicable if the Y_i shared a variance of σ^2 and have no covariances. This case can be occurred when they have no dependencies. The coefficients' sum of squares is

$$\sum a_i^2 = n(\frac{1}{n})^2 = \frac{1}{n} \tag{2.18}$$

Then the mean's variance becomes

$$Var(\bar{Y}) = \frac{\sigma^2}{n} \tag{2.19}$$

Take the covariance between two linear functions. Let

$U = \sum a_i Y_i$ and $W = \sum d_i Y_i$.

$$Cov(U, W) = \sum a_i d_i \, Var(Y_i) + \sum \sum_{i \neq j} a_i d_j \, Cov(Y_i, Y_j) \tag{2.20}$$

If the Y_i are independent, the equation 1.20 modified to

$$Cov(U, W) = \sum a_i d_i \, Var(Y_i) \tag{2.21}$$

For each Y_i, take note that the associated coefficients are $1/n$ and $x_i / \sum x_j^2$, respectively. As a result, $\bar{Y}$ and $\hat{\beta}_1$ have a covariance as

$$Cov(\bar{Y}, \hat{\beta}_1) = \sum \left(\frac{1}{n}\right) \left(\frac{x_i}{\sum x_j^2}\right) Var(Y_i)$$

$$= \left(\frac{1}{n}\right) \left(\frac{x_i}{\sum x_j^2}\right) \sigma^2 = 0$$

Thus,

$$Var(\hat{\beta}_0) = Var(\bar{Y}) + (\bar{X})^2 Var(\hat{\beta}_1)$$

$$= \frac{\sigma 2}{n} + \bar{X}^2 \frac{\sigma^2}{\sum x_i^2} =$$

$$= \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right) \sigma^2 \tag{2.22}$$

Remind that when $X = 0$, the estimated value of Y is $\hat{\beta}_0$. Since

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \bar{Y} + \hat{\beta}_1 (X_i - \bar{X}) \tag{2.23}$$

we have

$$Var(\hat{Y}_i) = \left[\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum x_j^2}\right] \sigma^2 \tag{2.24}$$

According to above formula when $X_i = X$, The variance gets its minimum value which is σ^2/n . The formula $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$ is used for the estimation of the mean $\beta_0 + \beta_1 X_0$ of Y_0. The prediction aim will be attained when there is a reduced discrepancy between $\hat{Y}_0$ and the next observation Y_0. **Prediction error** is the difference $Y_0 - \hat{Y}_0$. The $E(\hat{Y}_0 - Y_0)^2$ is known as **mean squared error of prediction** . The variance of the deviation between $\hat{Y}_0$ and the subsequent observation Y_0 is

$$Var(Y_{pred0}) = Var(\hat{Y}_0 - Y_0) = Var(\hat{Y}_0) + \sigma^2 \quad = [1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}] \qquad (2.25)$$

For given regression equation $Y_{\text{i}} = \beta_0 + \beta_1 X_i$ for all i:1,....,n, if the variance of the equation, $Var(X_i) = E^2(X_i) - E(X_i^2)$, is equal to σ^2,it shows that the assumption is **homoscedasticity**. In contrast, where the variance of X_i varies for every i. This situation is known as **heteroscedasticity**. In this case, symbolically, it is possible to write $Var(X_i) = \sigma_i^2$

2.3 Properties of Least-Squares Estimators: The Gauss–Markov Theorem

The least-squares estimation includes some optimum properties which is stated by and Gauss-Markov theorem. The best linear unbiasedness property of an estimator should be taken into consideration. The ordinary least squares (OLS) estimator $\hat{\beta}_2$, is said to be a best linear unbiased estimator (BLUE) of $\beta_2$if the following conditions hold:

1. It is a random variable's linear function.

2. It is unbiased.

3. It has the lowest variance among all of these linear unbiased estimators.

Gauss–Markov Theorem: The least-squares estimators, in the class of unbiased linear estimators, have lowest variance, which means they are BLUE given the classic linear regression model's underlying assumptions.

2.4 The Coefficient Of Determination R: A Measure Of "Goodness Of Fit"

The coefficient of determination r^2 (two-variable case)or R^2 (multiple regression) is a goodness of fit of model to data. To compute this r^2,

$$Y_i = \hat{Y}_i + \hat{e}_i$$

or in the deviated form

$y_i = \hat{y}_i + \hat{u}_i$ (1.26)

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \quad (2.26)$$

$$\sum y_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2$$

Since $\sum \hat{y}_i \hat{u}_i = 0$ and $\hat{y}_i = \hat{\beta}_2 x_i$. The numerous sums of squares that may be found in (1.27) can be represented as $\sum y_i^2 = \sum (Y_i - \bar{Y})^2$ which is the total Y value variation with respect to the sample mean, often known as the total sum of squares (TSS).

$\sum y_i^2 = \sum (\hat{Y}_i - \bar{\hat{Y}})^2 = \sum (Y_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum x_i^2$ is the variation of the estimated Y values about their mean ($\bar{\hat{Y}} = \bar{Y}$), which properly may be called the sum of squares due to regression or in simple terms the explained sum of squares (ESS).

$\sum \hat{u}_i^2$ is the residual sum of squares (RSS). Thus, (2.27) is

$$\text{TSS} = \text{ESS} + \text{RSS} \quad (2.27)$$

Now dividing (2.27) by TSS on both sides, we obtain

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

$$= \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

Now r^2 is defined as

$$r^2 = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum (Y_i - \bar{Y})^2} = \frac{ESS}{TSS}$$

or, as an alternative,

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{RSS}{TSS}$$

The coefficient of determination is known as r^2. A regression line's quality of fit is often assessed using it. Note that

1. It is nonnegative.

2. Its limits are $0 \leq r^2 \leq 1$. When $\hat{Y}_i = Y_i$, the r^2 takes its maximum value 1 and that denotes a best quality fit, for each i. An r^2 takes it is minimum value zero means that there is no connection between the $\hat{Y}_i$ and Y_i. In this case, $\hat{Y}_i = \hat{\beta}_1 = \bar{Y}$, that is, the mean value of every Y value is the best predictor of any Y value. As a result, in this instance, the regression line and the X axis will be horizontal to each other.

The coefficient of correlation, which measures how tightly two variables are connected but differs conceptually from r^2, is a metric that is closely related to but conceptually significantly distinct from r^2. You may calculate it either from

$$r = \pm\sqrt{r2} \tag{2.28}$$

or based on its definement

$$r = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}} = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n\sum Y_i^2 - (\sum Y_i)^2][n\sum Y_i^2 - (\sum Y_i)^2]}} \tag{2.29}$$

which is referred to as the **sample correlation coefficient.** Here are some characteristics of r (seeFigure2.1).

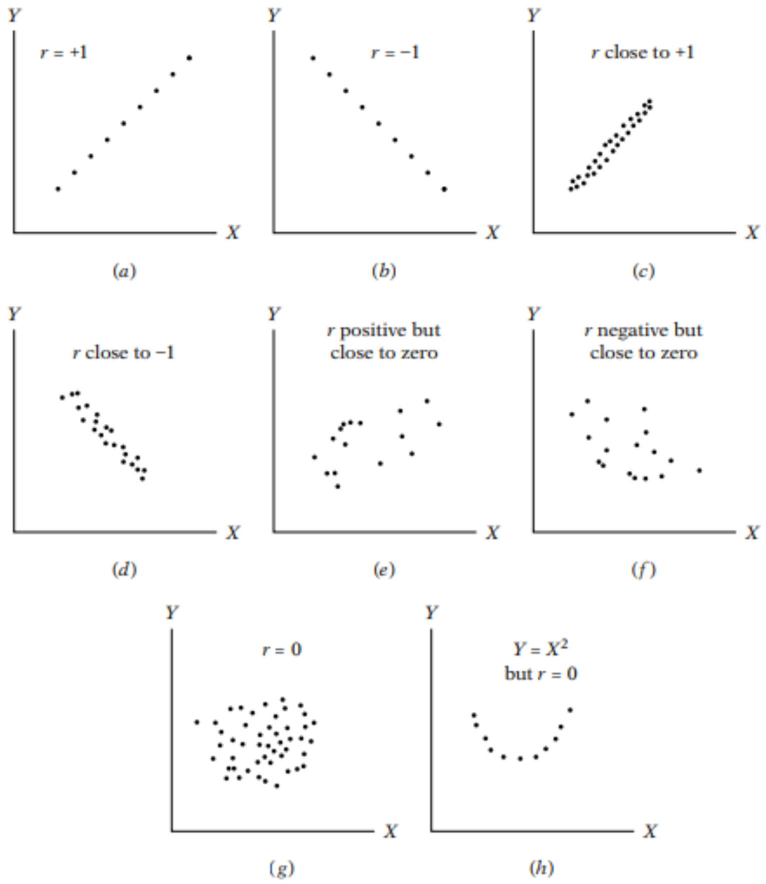

Figure 2.1 Correlation patterns

1. Depending on the sign of the term in the numerator (2.29), it might be either positive or negative.

2. It lies between −1 and +1; that is, $-1 \leq r \leq 1$.

3. It is symmetrical in nature; that is, $r_{XY} = r_{YX}$.

4. It doesn't care about scale or origin.

5. If X and Y are statistically independent , the correlation coefficient between them is zero; but the vise versa is not valid.

6. It is a measure of linear association only; for nonlinear relations the value has no meaning.

2.5 Hypothesis Testing

To test the null hypothesis $H_0 : \beta_1 = m$, where m is any constant of interest including zero (against the alternative hypothesis $H_a: \beta_1 \neq m, H_a: \beta_1 > m$, $or\ H_a: \beta_1 < m.$), the test statistic is

$$t = \frac{\hat{\beta}_1 - m}{s(\hat{\beta}_1)} \quad (2.30)$$

which is distributed as Student's t with n-p degrees of freedom under the assumptions of simple linear regression model where p is the number of parameters to be estimated.

3
MATRIX NOTATION

3.1 Fundamental Definitions

The components of a matrix are referred to as its elements. a_{1n} could be used to represent an elements general matrix.

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

The indices on the items indicate where the element is located by row and column, respectively. A matrix's rank is determined by how many rows (or columns) are linearly independent. The matrix

$$A = \begin{bmatrix} 1 & 1 & 4 \\ 3 & 1 & 10 \\ 5 & 2 & 17 \end{bmatrix}$$

has linear dependencies between its columns. The third column is created by multiplying the first column by three and combining it with the second column.

3.2 Specialized Matrix Types

A vector is a type of matrix formed with only one row, called row vector, or one column, called column vector. Thus, as an example,

$v = \begin{bmatrix} 1 \\ 9 \\ 4 \\ 8 \end{bmatrix}$ is a 4×1 column vector.

$\alpha = (\alpha 1\ \alpha 2\ \alpha 3)$ is a 1×3 row vector.

The matrix has equal columns and rows is called **square matrix.**

Diagonal matrix refers to the form of square matrix where all components are zero except for the main diagonal, which is made up of the elements a_{11} ,a_{22} , ..., a_{nn} and starts on the top left and moves to the lower right.

The symbol I_n designates an **identity matrix**, which is a matrix whose diagonal members are equivalent to one and other members equal to zero.

A square matrix is said to be **symmetric matrix** if the components form a *pattern* that is symmetrical around the diagonal line.

3.3 Operations of Matrix

When a matrix A is transposed, the rows of A become the columns of A', creating the matrix's transpose. If

$$A = \begin{bmatrix} 1 & 2 \\ 7 & 9 \\ 4 & 2 \\ 6 & 8 \end{bmatrix},$$

the transpose of A is $A' = \begin{bmatrix} 1 & 7 & 4 & 6 \\ 2 & 9 & 2 & 8 \end{bmatrix}$

It is commutative to add.: B + A =A + B.

A sets of vector multiplications can be viewed as the definition of matrix multiplication. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \text{ as } A = \begin{bmatrix} a_{1'} \\ a_{2'} \end{bmatrix}$$

where $a_{1'} = (a_{11} \;\; a_{12} \;\; a_{13})$ and $a_{2'} = (a_{21} \;\; a_{22} \;\; a_{23})$ are the 1×3 row vectors in A. Similar to this manner,

$$B = \begin{bmatrix} b_{11} & b_{11} \\ b_{21} & b_{11} \\ b_{31} & b_{11} \end{bmatrix} \quad B = (b_1 \; b_2)$$

$$AB = C = \begin{bmatrix} a_{1'}b_1 & a_{1'}b_2 \\ a_{2'}b_1 & a_{2'}b_2 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix},$$

Where

$$c_{11} = a_{1'}b_1 = \sum a_{1j}b_{j1} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}$$

$$c_{12} = a_{1'}b_2 = \sum a_{1j}b_{j2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

$$c_{21} = a_{2'}b_1 = \sum a_{2j}b_{j1} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31}$$

$$c_{22} = a_{2'}b_2 = \sum a_{2j}b_{j2} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32}$$

3.4 Multiple Regression in Matrix Notation

The linear additive model for relating a dependent variable to p independent variables is $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \mathcal{E}_i$. The independent variables p and Y were observed using the observational unit indicated by the subscript i. For the independent variable, use the second subscript. When the linear model contains the intercept β_0. Let $p' = (p+1)$. Here, we assume that $n > p'$.

Y : The dependent variable Y_i's $n\times 1$ column vector of findings

X: a column of ones joined on the p column vectors of the findings on the independent variables form the $n\times p'$ matrix.

β: the $p\times 1$ parameter vector; and

Ɛ: the $n\times 1$ random errors' vector.

With these, $Y = X\beta + \mathcal{E}$ or

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \ldots & X_{1p} \\ 1 & X_{21} & X_{22} & X_{23} & \ldots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \ldots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_n \end{bmatrix}$$

(nx1) (nxp') (p'x1) (nx1)

The joint pdf(probability density funciton) of $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ is

$$\prod_{i=1}^{n} 2\pi^{-1/2}\sigma^{-1}e^{-\mathcal{E}_i^2/2\sigma^2}$$

The random vector Ɛ is a vector $(\mathcal{E}_{1,}\ \mathcal{E}_2 \cdots \mathcal{E}_n)'$ made up of random variables $\mathcal{E}_i$.

Assuming that X and β components are constants, the Xβ and Y are vector of constants. Then, the constant vector Xβ and the random vector Ɛ combine to form the random vector Y. Assuming that $\mathcal{E}_i$ are separated $N(0,\sigma 2)$ random variables, we possess,

1. Y_i has a mean of $\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}$ and variance of σ^2, and it is a normal random variable;

2. Each of Y_i exists independently. When $i \neq j$, there is no covariance between Yi and Yj. The joint pdf(probability density function) of $Y_1, \ldots, Y_n$ is

$$(2\pi)^{-n/2}\sigma^{-n}e^{-\sum[Yi-(\beta0\ +\ \beta1\ Xi1\ +\ \beta2\ Xi2\ +\cdots+\ \beta p\ Xip)]^2/2\sigma^2} \qquad (3.1)$$

The likelihood function, which is used to determine which parameter values would have the highest likely of yielding a certain sample, is used to generate maximum likelihood estimators in the event that normality holds.

3.5 The Solution to the Normal Equations

The normal equations are shown in the matrix format below:

$$\hat{\beta} = (X'X)^{-1}(XY) \qquad (3.2)$$

In a p×p matrix created by the multiplication X'X, the other elements indicate the sums of the products of the independent variables, whereas the diagonal X'X components represent the total amount of the independent variables' squares. The typical expression for X'X is

$$\begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ip} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{ip} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i1}X_{ip} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{1p} & \sum X_{i1}X_{ip} & \sum X_{i2}X_{ip} & \cdots & \sum X_{ip}^2 \end{bmatrix}$$

In turn, make up the components of the matrix product X'Y:

$$X'Y = \begin{bmatrix} \sum Y_i \\ \sum X_{i1}\ Y_i \\ \sum X_{i2}\ Y_i \\ \vdots \\ \sum X_{ip}\ Y_i \end{bmatrix}$$

The singular answer to the ordinary equations can just be found when the reverse of X'X exists. The matrix X must therefore be of maximum column rank in order to for independent variables to not be linearly dependent.

3.6 The $\hat{Y}$ & Residuals Vectors

The dependent variable Y's vector of estimated means is calculated as follows.

$\hat{Y} = X\hat{\beta}$

It is helpful to formulate $\hat{Y}$ by inserting $[(X'X)^{-1}\ 'XY]$ for β. Thus,

$\hat{Y} = [X(X'X)^{-1}X]Y$

$= PY$

The residuals e is like below.

$e = Y - \hat{Y}$

The same way with $\hat{Y}$, PY may be used to write the equation in Y:

$e = Y - PY = (I - P)Y$

The least squares estimate as you remember reduces the squares of the residuals' sum, and that $e'e$ has been set to be the minimum. $(I - P)$ is symmetric and idempotent like P.

As a result Y has been split into two pieces: the residual e and the component of $\hat{Y}$ that is accounted for by the model.

3.7 Characteristics of Random Vectors' Linear Functions

The fact that e, $\hat{Y}$, and $\hat{\beta}$ are functions of the arbitrarily vector Y, indicates that they are all random vectors. These vectors are described in the earlier sections as linear functions AY of Y. The matrix A is

- $(X'X)^{-1}X'$ for $\hat{\beta}$,
- P for $\hat{Y}$, and
- $(I - P)$ for e.

In order to use the general characteristics of random vector's linear functions when researching the characteristics of $\hat{\beta}$, $\hat{Y}$, and e.

4
HETEROSCEDASTICITY TESTS

4.1 White's General Heteroscedasticity Test

The general test of heteroscedasticity proposed by White is easy to implement since it does not rely on the normality assumption. Take into account the below regression model with three-variable for the fundamental concept (It is straightforward to expand to the k-variable model.):

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i \quad (4.1)$$

The White test proceeds as follows:

Step 1. Given the data, we estimate (4.1) and obtain the residuals, $\hat{u}_i$.

Step 2. We then run the following (auxiliary) regression:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \quad (4.2)$$

In other words, the original X variables or regressors, their squared values, and the cross product(s) of the regressors are all used to regress the squared residuals from the initial regression. Regressors with higher powers may also be used. Regardless of whether original regression model contain the constant term or not, we add a constant term in this equation. We can obtain R^2 from auxiliary regression.

Step 3. It can be demonstrated that under the null hypothesis of no heteroscedasticity, sample size (n) times the R^2 derived by the auxiliary regression asymptotically follows the chi-square distribution with df equal to the number of regressors (apart from the constant term) in the auxiliary regression. That is,

$$n.R^2 \sim \chi_{df}^2 \quad (4.3)$$

where df is as previously described. There are 5 df in the example above because there are 5 regressors in the auxiliary regression.

Step 4. If the chi-square value obtained in (4.3) exceeds the critical chi-square value for the chosen level of significance, we obtain that the regression model is heteroscedastic. In the auxiliary regression (4.1), there is no heteroscedasticity if it does not surpass the crucial chi-square value, $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$.

In using the test, one must be careful because if a model has several regressors, theirs squared terms and cross products can quickly consume degrees of freedom. In additionally, the White test can be a test of (pure) heteroscedasticity or specification error or both. If there is cross product terms is auxiliary form, then it is testing both heteroscedasticity and specification bias. If there is no cross product terms, it is testing pure heteroscedasticity.

4.2 Park Test

Park formalizes the graphical method by suggesting that σ_i^2 is some function of the explanatory variable X_i. The functional form he suggested was

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{vi}$$

or

$$ln\sigma_i^2 = ln\sigma^2 + \beta \, ln \, X_i + v_i \qquad (4.4)$$

where v_i is the stochastic disturbance term.

Since σ_i^2 is generally not known, Park suggests using $\hat{u}_i^2$ as a proxy and running the following regression:

$$ln \; \hat{u}_i^2 = ln\sigma^2 + \beta \, ln \, X_i + v_i$$

$$= \alpha + \beta \, ln \, X_i + v_i \qquad (4.5)$$

If β appear statistically significant, it indicates that the data have heteroscedasticity. If it appears to be unimportant, homoscedasticity might be accepted. Thus, there are two stages to the Park test. We execute the OLS regression in the initial step while neglecting heteroscedasticity. From this regression, we derive $\hat{u}_i$ and in the following phase, we execute the regression (4.5)

4.3 Weighted Least Squares

Let's assume the linear model like

$$Y = X\beta + \mathcal{E}$$

with

$$Var(\mathcal{E}) = V\,\sigma^2$$

$$= Diag(a_1^2\; a_2^2 \ldots\ldots\ldots\ldots a_n^2\;)\sigma^2.$$

The variance of the random variable changes:

$$\sigma^2(cZ) = Var(\,cZ) = c^2 Var(Z)$$

$$= c^2[\sigma^2(Z)]$$

where c is a constant. The rescaled variable's variance is k^2, if the constant is selected to be like $c = k/\sigma(Z)$:

$$\sigma^2(cZ) = \left(\frac{k}{\sigma(Z)}\right)^2 \sigma^2(Z) = k^2 \quad (4.6)$$

$$Y_i = 1\,\beta_0 + X_{i1}\beta_1 + \cdots + X_{ip}\beta_p + e_i \quad (4.7)$$

where the i are uncorrelated , zero mean random variables. Assume that i's variance is $a_i^2\sigma^2$. Next, consider the scaled-down model

$$\frac{Y_i}{a_i} = \left(\frac{1}{a_i}\right)\beta_0 + \left(\frac{1}{a_i}X_{i1}\right)\beta_1 + \ldots\ldots\ldots\ldots \left(\frac{1}{a_i}X_{ip}\right)\beta_p + \left(\frac{e_i}{a_i}\right)$$

or

$$Y_i^* = X_{i0}^*\beta_0 + X_{i1}^*\beta_1 + \ldots\ldots\ldots\ldots.X_{ip}^*\beta_p + e_i^* \quad (4.8)$$

The following is the weighted regression matrix formulation. Create the matrix $V^{1/2}$ as $V^{1/2}V^{1/2} = V$. Y can be rescaled with common variances by W as stated below

$$W = (V^{\,1/2})^{-1}$$

$$= \begin{bmatrix} 1/a_1 & 0 & \cdots & 0 \\ 0 & 1/a_2 & & 0 \\ \vdots & \vdots & 1/a_3 & \vdots \\ 0 & 0 & \cdots & 1/a_n \end{bmatrix} \quad (4.9)$$

where the a_i are constants represents the proportionate variations in the variances of e_i. Premultiplying the model's two sides by W results in

$$WY = WX\beta + W \tag{4.10}$$

or

$$Y^* = X^*\beta + e^* \tag{4.11}$$

where $Y^* = WY, X^* = WX$, and $e^* = We$. Taking into account the variances of linear functions, the variance of e^* is

$$Var(e^*) = W[Var(e)]W' = WVW\sigma^2 = I\sigma^2 \tag{4.12}$$

since $WVW = (V^{1/2})^{-1}V^{1/2}V^{1/2}(V^{1/2})^{-1} = I$. Estimating β on Y^* and X^* can be done via ordinary least squares when the standard assumption of equal variances is satisfied.

The estimate of β using weighted least squares is

$$\hat{\beta}_W = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* \tag{4.13}$$

Alternatively,

$$\hat{\beta}_W = (X'W'WX)^{-1}(X'W'WY).$$

$$= (X'V^{-1}X)^{-1} - 1(XV - 1Y) \tag{4.14}$$

$$Var(\hat{\beta}_W) = \left(X^{*\prime}X^*\right)^{-1}\sigma^2 = (X'V^{-1}X)^{-1}\sigma^2. \tag{4.15}$$

The modified scale's fitted values are produced by using

$$Y^* = X^*\hat{\beta}_W,$$

$$= X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Y^* = P^*Y^* \tag{4.16}$$

$$\hat{Y}_W = W^{-1}\hat{Y}^* = X\hat{\beta}_W \tag{4.17}$$

Their respective variances are

$$Var(\hat{Y}^*) = X^*(X'V^{-1}X)^{-1}X^{*\prime}\sigma^2 = P^*\sigma^2 \tag{4.18}$$

and

$$Var(\hat{Y}_W) = X\left(X'^{V^{-1}}X\right)^{-1}X'\sigma^2 \tag{4.19}$$

$e^* = Y^* - \hat{Y}^*$ where e^* is observed residuals on the transformed scale and $e = Y - \hat{Y}_W$ on the actual scale. Note that $e = W^{-1}e^*$. Their variances are

$$Var(e^*) = [I - X^*(X'V^{-1}X)^{-1}X^{*\prime}]\sigma^2 \quad (4.20)$$

$$= (I - P^*)\sigma^2$$

and

$$Var(e) = [V - X\left(X'^{V^{-1}}X\right)^{-1}X']\sigma^2 \quad (4.21)$$

Thus,

$$V\sigma^2 = \begin{bmatrix} 1/r_1 & 0 & \cdots & 0 \\ 0 & 1/r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/r_n \end{bmatrix} \quad (4.22)$$

Using a weighted matrix, $Var(e^*) = I\sigma^2$ is

$$W = \begin{bmatrix} \sqrt{r_1} & 0 & \cdots & 0 \\ 0 & \sqrt{r_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{r_n} \end{bmatrix} \quad (4.23)$$

4.4 Box-Cox Transformation

Box and Cox describe a method for computing a power modification for the dependent variable using computation (1964), with the goal of obtaining a straightforward, normal, linear model that meets with the standard premises of least squares. Values of the power transformation (λ), σ^2, and β are obtained using the Box– Cox technique. First, the validation of the Box-Cox transformation should be verified, and standard proper procedures should be followed. Normality may not always come from the Box-Cox transformation, though. A family of transformations with parameters defined is used in the Box–Cox technique, in standard form, as

$$Y_i^{(\lambda)} = \begin{cases} \frac{Y_i^{\lambda}-1}{\lambda(\dot{Y})^{(\lambda-1)}} & for\ \lambda \neq 0 \\ Y\ln(Y_i) & for\ \lambda = 0 \end{cases} \quad (4.24)$$

where $\dot{Y}$ is defined as the geometric mean of the initial observations,

$$\dot{Y} = exp \sum[ln(Y_i)]/n$$

According to the Box-Cox technique, for some λ the $Y_i^{(\lambda)}$ assuage that independently and normally distributed with mean $X\beta$ and common variance σ^2. In other words $Y_i^{(\lambda)}$meet each and every standard theory of least squares premise. Under these presumptions, the maximum likelihood estimators of λ, β, and σ^2 are calculated. Least squares analysis is used to get the greatest probability solution for the updated data for λ, let's say $\lambda = -1 \; to \; 1$. Let $SS[Res(\lambda)]$ represents the residual sum of squares that was produced by fitting the model to $Y_i^{(\lambda)}$ for the specified value of λ and let $\sigma^2(\lambda) = \{SS[Res(\lambda)]\}/n$. For every option of λ, the probability is shown by

$$L_{max} = \frac{1}{2}\ln[\hat{\sigma}^2(\lambda)] \tag{4.25}$$

By maximizing the likelihood, the minimum value of the RSS can be obtained. $S[Res(\lambda)]$ is then plotted against to λ determine the value at which the minimum, $SS[Res(\lambda)]_{min}$, is attained in order to get the maximum probability solution for λ.

5
ENTROPY

Entropy is an indicator of reversibility; when there is no change of entropy, the process is reversible. An increase in entropy is a decrease of available energy. In general, an increase in entropy means decrease in order. Disorder in the sense of unpredictability based on a lack of knowledge of the positions and velocities of molecules (Pierce,1980).

The entropy of a statistical experiment, is a measure of uncertainty (Khinchin, 1957). So in a statistical sense, entropy and the amount of information are two closely related concepts. Uncertainty is not present after a statistical experiment is conducted. For some concepts and applications of statistical entropy, one can also refer to Cover&Thomas (2006) , Ash (1990), Reza(1994), Kullback(1996), and Rényi (2007).

5.1 Shannon Entropy

Let the discrete random variable X takes on the value $x_1, x_2,\ldots, x_k$ with respective probabilities $p_1, p_2,\ldots, p_k$ Shannon entropy is defined as

$$H_s = -\sum_{i=1}^{k} p_i \log p_i \tag{5.1}$$

Let $\widehat{H}$ is the estimator of Shannon entropy. It is calculated as

$$\widehat{H}_s = -\sum_{i=1}^{k} \hat{p}_i \log \hat{p}_i \tag{5.2}$$

Here $\hat{p}_i$ probabilities are estimated by maximum likelihood method. The variance of Shannon entropy is found as (Zhang Xing, 2013).

$$Var(\widehat{H}) \; = \frac{1}{n}\left(\sum_{i=1}^{K} p_i \ln^2 p_i - \; H^2\right) + \frac{K-1}{2n^2} \; + O(n^{-3}) \tag{5.3}$$

5.2 Rényi Entropy

Rényi entropy is defined as

$$H_R \; = \frac{\log \sum_{i=1}^{K} p_i^a}{1-a} \; for \quad a > 0 \quad \text{and} \quad a \neq 1 \tag{5.4}$$

The variance of Rényi entropy is given as follows (Pardo, 2006);

$$Var(\widehat{H}_R) = \frac{1}{n}\left[\left(\frac{a}{a-1}\right)^2 \left(\sum_{i=1}^{K} p_i^{2a-1} - \left(\sum_{i=1}^{K} p_i^{a}\right)^2\right)\right] \quad (5.5)$$

5.3 Tsallis Entropy

Tsallis (or Havrda-Charvat) entropy is known as

$$H_T = \frac{1-\sum_{i=1}^{K} p_i^a}{a-1} \; for\; a > 0 \quad and \quad a \neq 1 \quad (5.6)$$

The variance of this entropy estimator is (Pardo, 2006);

$$Var(H_T) = \frac{1}{n}\left[\left(\frac{a}{a-1}\right)^2 \left(\sum_{i=1}^{K} p_i^{2a-1} - \left(\sum_{i=1}^{K} p_i^{a}\right)^2\right)\right] \quad (5.7)$$

In the below table, the formulas of entropies for the various distributions can be found.

Table 5.1 The entropy formulas for distributions

Distribution	**Probability (density) function**	**Shannon Entropy**	**Rényi Entropy**	**Tsallis Entropy**
Uniform (continuous)	$\frac{1}{(b-a)}$	$\ln(b-a)$	$\ln(b-a)$	$\frac{1-(b-a)^{1-\alpha}}{\alpha-1}$
Exponential	$\lambda e^{-\lambda X}$	$1-ln\lambda$	$\frac{(\alpha-1)ln\alpha\lambda-\alpha ln\alpha}{1-\alpha}$	$\frac{\alpha-\lambda^{\alpha-1}}{\alpha(\alpha-1)}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2})$	$\ln(\sigma\sqrt{2\pi e})$	$\ln(\sigma\sqrt{2\pi e}) - \frac{ln\alpha}{2(1-\alpha)}$	$\frac{\sqrt{\alpha}-(\sigma\sqrt{2\pi})^{1-\alpha}}{\sqrt{\alpha}(\alpha-1)}$
Gamma	$\frac{\lambda^\theta e^{-\lambda x} x^{\theta-1}}{\Gamma(\theta)}$	$(1-\theta)ln\theta - ln\lambda + \theta - \frac{1}{2\theta} + \frac{1}{2} + \ln(\Gamma(\theta))$	$\ln\left(\frac{\sqrt{2\theta\pi e}}{\lambda}\right) - \frac{ln\alpha}{2(1-\alpha)}$ for large θ	$\frac{\sqrt{\alpha}-(\frac{\sqrt{2\theta\pi}}{\lambda})^{1-\alpha}}{\sqrt{\alpha}(\alpha-1)}$ for large θ

6
APPLICATION

We have generated and used some artificial heteroscedastic data given in appendix data-1, data-2, data-3 and data-4 We then fitted linear regression models, each having its own weighting methodology. Then we compared the goodness of fit statistics of each linear regression model, after weighting, and performing the weighted least squares technique. Note that initially (as a classical approach) we took the reciprocals of the sample standard deviations calculated at each X-level as the weights. Then we repeated the same procedure by taking the weights as the reciprocals of entropies of normal distribution calculated at the X-levels. Note that since Tsallis, and Rényi entropies have parametric nature, we performed various weighting techniques by giving different values to the α parameter. Finally, as an alternative to the weighted least squares we adopted Box-Cox approach.The three goodness of fit statistics used are R^2, R^2_{adj} and MSE.

From Figure1, and Figure 2, the heteroscedastic nature of data is obvious. These data were produced by adding random numbers whose variances are increasing systematically according to the levels of independent variables. Note that random numbers are generated from normal distributions with different variances. Note too that for the first data set, there are 5 dependent observations for each level of independent variable, and for the 2nd, 3rd, and 4th data sets there are 25 dependent observations generated. By checking regression results summarized in Tables 2-4, we may say that weighting has a positive effect on increasing R^2, R^2_{adj} and/or decreasing MSE . Besides we have observed that in many cases using the reciprocals of Tsallis, and Rényi entropies as the weights, produced much better goodness of fit statistics.

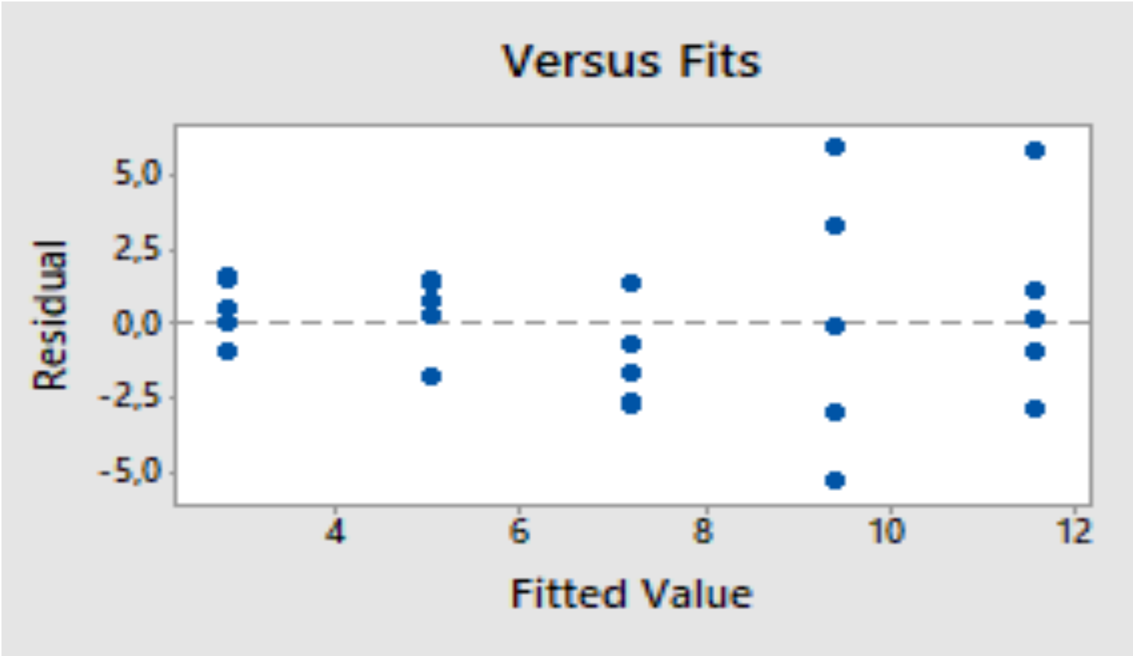

Figure 6.1 The scatter diagram of data-1

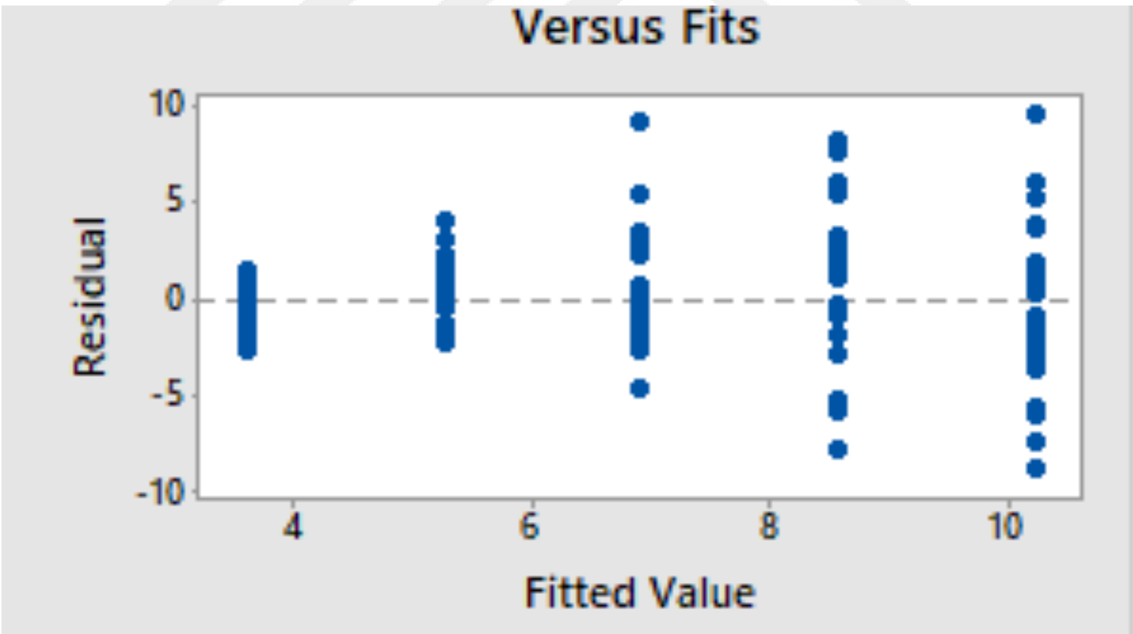

Figure 6.2 The scatter diagram of data-2

Table 6.1 Goodness of fit statistics model for data-1

Data-1	R^2	R^2adj	MSE
Classical	59.3%	57.53%	7.173
Weighted with OLS	65.84%	64.35%	1.052
Shannon	63.28%	61.68%	2.881
Renyi α=0,5	62.4%	60.77%	2.241
Renyi α=1,5	62.71%	61.09%	2.467
Box-cox	63.24%	61.65%	0.094
Tsallis α=1,5	61.23%	59.55%	4.064
Tsallis α=0,5	63.96%	62.39%	0.027
Tsallis α=1/9	66.27%	64.81%	0.021
Tsallis α=1/10	66.22%	64.75%	0.068
Renyi α=1/9	60.84%	59.14%	1.113

Table 6.2 Goodness of fit statistics model for data-2

Data-2	R^2	R^2adj	MSE
Classical	35.24%	34.71%	10.260
Weighted with OLS	44.63%	44.18%	1.589
Shannon	39.06%	38.57%	4.086
Renyi α=0,5	38.25%	37.75%	3.175
Renyi α=1,5	38.54%	38.04%	3.497
Box-cox	35.39%	34.87%	0.371
Tsallis α=1,5	36.88%	36.37%	5.810
Tsallis α=0,5	39.38%	38.89%	0.648
Tsallis α=1/9	44.24%	43.79%	0.032
Tsallis α=1/10	44.34%	43.89%	0.102
Renyi α=1/9	36.77%	36.25%	1.581

Table 6.3 Goodness of fit statistics model for data-3

Data-3	R^2	R^2adj	MSE
Classical	38.48%	37.98%	9.452
Weighted with OLS	52.52%	52.13%	0.952
Shannon	44.29%	43.84%	3.566
Renyi α=0,5	42.91%	42.45%	2.816
Renyi α=1,5	43.39%	42.93%	3.085
Box-cox	36.67%	36.16%	0.359
Tsallis α=1,5	41.12%	40.65%	5.242
Tsallis α=0,5	44.44%	43.99%	0.564
Tsallis α=1/9	52.3%	51.91%	0.021
Tsallis α=1/10	52.41%	52.03%	0.065
Renyi α=1/9	40.6%	40.12%	1.435

Table 6.4 Goodness of fit statistics model for data-4

Data-4	R^2	R^2adj	MSE
Classical	56.47%	56.27%	8.150
Weighted with OLS	61.55%	61.38%	0.925
Shannon	60.11%	59.93%	3.180
Renyi α=0,5	59.13%	58.95%	2.493
Renyi α=1,5	59.46%	59.28%	2.737
Box-cox	56.37%	56.18%	0.279
Tsallis α=1,5	58.49%	58.3%	4.570
Tsallis α=0,5	59.65%	59.47%	0.508
Tsallis α=1/9	62.22%	62.05%	0.020
Tsallis α=1/10	62.2%	62.04%	0.064
Renyi α=1/9	57.67%	57.48%	1.254

7
RESULTS AND DISCUSSIONS

The simplest model for regression analysis involves one independent variable and a constant. The variances of the equation may be equal for each level of independent variable or not. For equal variances the equation is called homoscedastic, otherwise the equation is called heteroscedastic. The heteroscedasticity could be detected visually when the data is shown in scatter diagram. On the other hand, for the numeric results White Test and Park's Test should be used to be more precise. In order to avoid from heteroscedasticity, generally weighting with error squares is one strategy. The Box-Cox transformation is also used to remedy. Alternatively, although it's not in common usage, weighting with reciprocal of entropy can be used, since entropy is a measure of uncertainty.In this study Shannon, Renyi and Tsallis entropies are used for determining the weights of The weighted Least Squares (WLS) procedure. As can be seen on the tables above, introducing Tsallis and Rényi entropies in weighting may improve goodness of test statistics of a regression model like the coefficient of determination, the adjusted coefficient of determination, and the mean squares error.

REFERENCES

Ash, R. B. (1990). “Information Theory”, Dover Publications, NY

Cover, T.M., Thomas J.A.(2006). “Elements of Information Theory”. Wiley Interscience (Second Edition). Hoboken.New Jersey, p45

Gujarati, D.M.(2004). ”Basic Econometrics”,The McGraw-Hill Companies fourth edition pp404,413.

Khincin, A.I (1957). ” Mathematical Foundations of Information Theory”. Dover Publications. NY, p7.

Kullback, S. (1996). “Information Theory and Statistics”. Dover Publications. NY

Neter, J, Kutner M.H. (1985). “Applied Linear Statistical Models”. Richard D.Irwin. Inc, p263.

Pardo, L.(2006). “Statistical Inference Based on Divergence Measures”.

Chapman&HALL/CRC, pp1-34.

Pierce, J.R. (1980). “An Introduction to Information Theory Symbols, Signals and Noise”. Dover Publications,Second, revised edition. NY, pp21-24

Rawlings, J.O., Pantula S.G., Dickey D.A.(1998). “Applied Regression Analysis”. Springer-Verlag.NY, pp1-10.

Renyi, A.(2007). “Probability Theory”. Dover Publications. New York , p23.

Reza,F.M. (1994). “An Introduction to Information Theory”. Dover Publications. NY,p268.

Ullah, A.(1996). “Entropy, Divergence and Distance Measures with Econometric Applications, Journal of Statistical Planning and Inference”. 49(1996), pp137-162

Upton,G., Cook I. (2006). Oxford Dictionary of Statistics,Second Edition. Oxford University Press.

Zhang, Xing (2013). “Asymptotic Normality of Entropy Estimators”.proposal submitted to the faculty of The University of North Carolina at Charlotte in partial fullfilment of the requirements for the degree of Doctor of Philosophy in Applied Mathematics, Charlotte)

APPENDIX

Data-1

X	Y
1,00	2,70
1,00	1,72
1,00	3,24
1,00	4,28
1,00	4,20
2,00	6,44
2,00	5,64
2,00	6,26
2,00	3,11
2,00	5,13
3,00	8,47
3,00	5,35
3,00	6,32
3,00	4,31
3,00	4,49
4,00	3,89
4,00	15,33
4,00	12,65
4,00	6,24
4,00	9,16
5,00	8,56
5,00	11,61
5,00	12,64
5,00	17,32
5,00	10,49

Data-2

X	Y
1,00	2,46
1,00	4,12
1,00	2,33
1,00	1,81
1,00	0,82
1,00	3,53
1,00	2,81
1,00	3,37
1,00	1,42
1,00	2,60
1,00	2,28
1,00	4,04
1,00	4,93
1,00	3,48
1,00	4,27
1,00	4,59
1,00	1,44
1,00	3,09
1,00	2,65
1,00	2,73
1,00	2,78
1,00	3,28
1,00	1,09
1,00	4,72
1,00	2,91
2,00	2,73
2,00	6,75
2,00	8,11
2,00	5,29
2,00	5,32
2,00	6,59
2,00	4,84
2,00	5,91
2,00	7,24
2,00	5,57
2,00	5,47
2,00	9,15
2,00	4,78
2,00	6,16
2,00	5,19
2,00	6,23
2,00	3,77
2,00	5,61
2,00	3,38
2,00	4,48
2,00	6,94
2,00	4,59
2,00	6,48
2,00	5,27
2,00	6,20
3,00	6,57
3,00	7,42
3,00	5,77
3,00	5,34
3,00	15,82
3,00	5,27
3,00	6,80
3,00	6,79
3,00	4,05
3,00	5,57
3,00	9,65
3,00	2,14
3,00	10,26

Data-3

X	Y
5,00	12,54
5,00	17,02
5,00	13,08
5,00	6,20
5,00	0,93
5,00	6,76
5,00	9,32
5,00	11,55
5,00	4,83
5,00	6,53
5,00	9,75
5,00	14,97
5,00	0,70
5,00	6,86
5,00	13,65
5,00	10,37
5,00	15,28
5,00	10,70
5,00	6,13
5,00	14,25
5,00	8,41
5,00	12,01
5,00	16,71
5,00	5,06
5,00	10,96
4,00	6,42
4,00	5,88
4,00	10,80
4,00	12,96
4,00	10,78
4,00	14,48
4,00	6,78
4,00	7,67
4,00	8,61
4,00	13,07
4,00	6,38
4,00	5,79
4,00	10,36
4,00	8,67
4,00	6,79
4,00	8,74
4,00	9,80
4,00	15,02
4,00	8,49
4,00	14,33
4,00	12,14
4,00	3,63
4,00	12,36
4,00	1,04
4,00	3,49
3,00	7,04
3,00	5,88
3,00	3,78
3,00	9,80
3,00	14,16
3,00	5,37
3,00	3,55
3,00	6,94
3,00	2,82
3,00	10,93
3,00	4,06
3,00	4,20
3,00	5,93

Data-4

X	Y
1,00	2,89
1,00	3,19
1,00	2,01
1,00	4,76
1,00	2,43
1,00	2,94
1,00	4,30
1,00	3,33
1,00	4,12
1,00	1,47
1,00	3,63
1,00	2,99
1,00	3,68
1,00	2,37
1,00	3,67
1,00	2,02
1,00	5,51
1,00	3,66
1,00	4,08
1,00	3,32
1,00	4,00
1,00	3,22
1,00	4,31
1,00	3,39
1,00	3,53
2,00	6,14
2,00	5,58
2,00	5,14
2,00	5,89
2,00	7,85
2,00	5,40
2,00	4,96
2,00	3,91
2,00	3,64
2,00	7,75
2,00	8,80
2,00	2,89
2,00	5,26
2,00	3,17
2,00	4,22
2,00	4,56
2,00	6,87
2,00	4,00
2,00	5,51
2,00	3,98
2,00	5,03
2,00	3,78
2,00	4,79
2,00	3,46
2,00	2,81
3,00	9,18
3,00	6,01
3,00	4,36
3,00	6,30
3,00	7,17
3,00	13,33
3,00	5,90
3,00	12,44
3,00	6,67
3,00	0,25
3,00	9,77
3,00	9,09
3,00	11,89

Figure apx.1 Data

3,00	4,60
3,00	4,54
3,00	6,67
3,00	6,22
3,00	9,93
3,00	4,75
3,00	12,13
3,00	7,00
3,00	10,10
3,00	10,09
3,00	4,93
3,00	8,98
4,00	16,57
4,00	5,55
4,00	10,46
4,00	3,20
4,00	9,67
4,00	13,77
4,00	9,78
4,00	15,98
4,00	11,30
4,00	7,43
4,00	7,98
4,00	16,64
4,00	14,49
4,00	2,52
4,00	7,78
4,00	11,71
4,00	7,55
4,00	9,37
4,00	11,13
4,00	2,54
4,00	10,30
4,00	0,63
4,00	6,40
4,00	9,48
4,00	8,12
5,00	13,91
5,00	2,71
5,00	6,88
5,00	13,89
5,00	10,88
5,00	19,69
5,00	11,45
5,00	9,15
5,00	7,55
5,00	1,31
5,00	9,05
5,00	16,04
5,00	4,45
5,00	7,99
5,00	11,89
5,00	8,72
5,00	10,34
5,00	10,67
5,00	13,68
5,00	4,35
5,00	6,29
5,00	7,77
5,00	8,57
5,00	15,31
5,00	4,07

3,00	10,73
3,00	5,27
3,00	10,85
3,00	12,08
3,00	9,29
3,00	8,75
3,00	2,28
3,00	10,30
3,00	7,22
3,00	3,82
3,00	6,10
3,00	4,10
2,00	5,22
2,00	8,32
2,00	4,32
2,00	5,57
2,00	5,95
2,00	6,12
2,00	4,72
2,00	5,78
2,00	6,20
2,00	3,43
2,00	6,06
2,00	5,37
2,00	7,18
2,00	2,50
2,00	2,22
2,00	4,71
2,00	6,39
2,00	4,26
2,00	3,73
2,00	7,95
2,00	6,43
2,00	3,67
2,00	6,86
2,00	4,71
2,00	6,15
1,00	3,09
1,00	2,03
1,00	3,40
1,00	3,02
1,00	1,49
1,00	2,42
1,00	3,63
1,00	2,51
1,00	3,10
1,00	2,11
1,00	1,77
1,00	3,39
1,00	4,41
1,00	2,94
1,00	1,61
1,00	4,07
1,00	6,07
1,00	4,45
1,00	2,53
1,00	2,83
1,00	2,40
1,00	3,08
1,00	4,17
1,00	3,19
1,00	3,01

3,00	7,82
3,00	9,22
3,00	9,99
3,00	7,10
3,00	8,43
3,00	7,79
3,00	2,94
3,00	9,37
3,00	2,63
3,00	2,83
3,00	10,87
3,00	3,38
4,00	5,08
4,00	16,01
4,00	8,05
4,00	16,92
4,00	10,13
4,00	16,23
4,00	11,79
4,00	8,35
4,00	13,84
4,00	10,11
4,00	4,20
4,00	7,71
4,00	7,85
4,00	4,51
4,00	9,97
4,00	8,61
4,00	9,37
4,00	8,32
4,00	8,48
4,00	9,73
4,00	7,49
4,00	2,86
4,00	11,70
4,00	12,75
4,00	14,98
5,00	20,01
5,00	15,44
5,00	11,93
5,00	13,00
5,00	20,57
5,00	16,85
5,00	10,39
5,00	14,46
5,00	12,03
5,00	22,22
5,00	8,63
5,00	11,92
5,00	17,18
5,00	13,80
5,00	8,91
5,00	10,50
5,00	15,52
5,00	17,63
5,00	12,94
5,00	9,44
5,00	15,61
5,00	12,69
5,00	7,38
5,00	10,28
5,00	17,79
4,00	4,98
4,00	6,98
4,00	10,94

Figure apx.2 Data (still)

4,00	9,99
4,00	10,24
4,00	2,33
4,00	8,46
4,00	5,78
4,00	9,65
4,00	8,91
4,00	8,72
4,00	9,74
4,00	10,55
4,00	5,95
4,00	11,66
4,00	19,10
4,00	10,84
4,00	5,62
4,00	13,66
4,00	11,47
4,00	4,43
4,00	11,06
4,00	7,59
4,00	12,73
4,00	12,45
3,00	4,09
3,00	7,14
3,00	6,80
3,00	1,64
3,00	4,28
3,00	6,22
3,00	9,65
3,00	9,99
3,00	2,82
3,00	6,72
3,00	4,77
3,00	1,22
3,00	6,12
3,00	6,59
3,00	4,21
3,00	9,62
3,00	4,91
3,00	0,92
3,00	8,66
3,00	5,04
3,00	4,85
3,00	7,53
3,00	10,18
3,00	5,58
3,00	11,98
2,00	6,63
2,00	6,35
2,00	0,67
2,00	5,02
2,00	5,58
2,00	6,51
2,00	5,81
2,00	7,37
2,00	3,33
2,00	8,07
2,00	2,50
2,00	5,46
2,00	4,11
2,00	5,24
2,00	1,62
2,00	3,47
2,00	6,54
2,00	2,10

Figure apx.3 Data (still)

2,00	2,55
2,00	3,84
2,00	4,07
2,00	7,17
2,00	6,89
2,00	8,83
2,00	4,88
1,00	2,70
1,00	3,53
1,00	3,92
1,00	3,26
1,00	2,23
1,00	3,90
1,00	2,10
1,00	3,32
1,00	1,73
1,00	2,87
1,00	1,64
1,00	3,50
1,00	2,09
1,00	3,45
1,00	2,01
1,00	3,72
1,00	2,85
1,00	2,80
1,00	3,11
1,00	2,11
1,00	1,56
1,00	4,23
1,00	3,56
1,00	3,87
1,00	2,35

Figure apx.4 Data (still)

PUBLICATIONS FROM THE THESIS

Papers

Cin, H.C., Evren A.A. (2019). "Determining Weights By Entropy Measures In Case of Heteroscedasticity". IJEAS. ISSN:2394-3661. Issue-5. May 2019.