

# A Comparison of Three Estimation Methods In Linear Regression Analysis

Xianghong Luo

Loyola University Chicago, Chicago 60660, USA [gznyyb@hotmail.com](mailto:gznyyb@hotmail.com)

**Keyword:** Linear Regression, Least Ordinary Square, Method of Moment, Maximum Likelihood Estimate, Hypothesis Testing, R-square

**Abstract.** The paper begins with an introduction of some crucial definitions apropos of regression analysis. Then it discusses briefly the concept of R-square that verifies the accuracy of a regression model and of Hypothesis Testing that tests hypothesis made concerning the population. The main part of the paper then focuses on three estimation methods that estimate the parameters of a regression model: Ordinary Least Square, Method of Moments, and Maximum Likelihood Estimation. The paper concludes with a discussion on the advantages and disadvantages of each method and the possible applications of linear regression.

## 1. Introduction

Regression Analysis is an important topic under the study of statistics because it not only estimates the relationship between x-values and y-values of the data points on a graph but also is able to predict data's future behavior according to the known data. Simply stated, Linear Regression Analysis (LRA) is the process of generating a linear equation ( $\hat{y} = \hat{\alpha} + \hat{\beta}x_t$ ), called the Regression Equation or Regression Model, that best expresses all the given data points. In order to distinguish it from Population Regression Function (PRF), which will be discussed later, Regression Equation is also called Sample Regression Function (SRF). Note that in the equation above, the accents on top of  $\hat{\alpha}$  and  $\hat{\beta}$ , called the parameters, specify that they are estimated values. Since this paper will focus primarily on Linear Regression Analysis, regression equations discussed in this paper contain only linear relation. There is a linear relation between y and x when y is a linear function of x ( $y = ax + b$  for instance) or when y is a linear function of parameters ( $y = ax^2 + b$  for instance).

In addition to estimating relationship between the x-values and y-values of the data points, a linear equation, called the Population Regression Function (PRF) in the form of  $y_t = \alpha + \beta x_t + u_t$ , can also incorporate more information including that of the relationship between the linear equation and the data points [2]. Note that the subscript t indicates that y, x, and u in this equation are accurately represented as the y, x, and u values of the given data points where  $t = 0, 1, 2, 3\dots n$  and n is number of data points. Note also that  $u_t$  in a PRF is called the Stochastic Error Terms or simply Error Terms. They are differences between y-values of the linear equation and those of the data points and their calculations are crucial components of various estimation methods, which the paper will discuss at length. One can readily observe that PRF is the sum of the SRF and the error terms.

## 2. Estimation Methods

Estimation Methods are methods through which Regression Analysis is conducted to generate a linear equation based on the data points given on a graph. Generally used estimation methods include Ordinary Least Squares (OLS), Method of Moments (MoM), and Maximum Likelihood Estimate (MLE).

### 2.1 Ordinary Least Squares

The basic principle utilized by OLS is the minimization, through Calculus, of sum of stochastic error terms [1]. The sum of error terms can be computed through the following formula:

$$S = \sum_{t=1}^n (y_t - (\alpha x_t + \beta))^2$$

where  $n$  denotes the number of data points and  $y_t$  and  $x_t$  the y-values and x-values of the data points. According to the basic principles of Calculus,  $S$  is minimized when the partial derivative of  $S$  in terms of  $x$  and in terms of  $y$  are set to equal to 0:

$$\sum_{t=0}^n 2 * ((\alpha x_t + \beta) - y_t) * x_t = 0$$

$$\sum_{t=0}^n 2 * ((\alpha x_t + \beta) - y_t) = 0$$

Then we can by solving  $\alpha$  and  $\beta$  obtain the parameters that are necessary for generating a linear equation:

$$\alpha = \frac{1}{n^3 - n} * [12 * \sum_{t=1}^n (t * y_t) - 6 * (n + 1)(\sum_{t=1}^n (y_t))]$$

$$\beta = \frac{1}{n^2 - n} * [(4 * n + 2) * \sum_{t=1}^n (y_t) - 6 * (\sum_{t=1}^n (t * y_t))]$$

## 2.2 Method of Moments

MoM utilizes the relationship between moments, which is defined as the expectation of the powers of a random variable, and unknown parameters to estimate the values of parameters [4]. If  $y_t$  is set to be a random variable  $X$ , its moments can be calculated and be set to equal to randomly assigned letters:

$$\alpha_1 = E[X], \alpha_2 = E[X^2], \dots, \alpha_k = E[X^k]$$

The number of moments that needed to be calculated (the  $k$  value in the above list) depends on the number of unknown parameters. MoM estimates that the moments of the population are equal to those of the sample. Moreover, the moments of the sample can be calculated according to Law of Large Numbers and when they are set to equal to moments of the population obtained above, the following equations can be obtained:

$$A_1 = \frac{1}{n} \sum_{i=1}^n X_i = \alpha_1, \dots, A_k = \frac{1}{n} \sum_{i=1}^n X_i^k = \alpha_k$$

If  $X$ , and by extension moments of  $X$ , is a function of the unknown parameters, parameters can be estimated by solving the system of equations listed above.

## 2.3 Maximum Likelihood Estimate

The basic principle of MLE is to seek parameters that maximize the Likelihood Function. In other words, MLE assumes that the value of the parameters is true when it makes the observed sample most probable [3]. MLE is applicable whenever the probability distribution of the error terms is known. In the example provided by this paper, error terms are normally distributed with 0 average and constant standard deviation. One can readily conclude that y-values of the data points are also normally distributed whose average is the linear equation. The probability functions of the y-values of the data points are, therefore, as follows:

$$P(y_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} \Sigma(y_t - \beta - \alpha x_t)^2}$$

All of these functions can be combined to form a combined probability function, called the Likelihood function:

$$P(y_1, y_2, y_3, \dots, y_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \Sigma(y_t - \beta - \alpha x_t)^2}$$

In order to simplify the resulting function, one can take the natural log of it and as a result of this simplification, the Likelihood Function becomes:

$$P(y_1, y_2, y_3, \dots, y_n) = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum (y_t - \beta - \alpha x_t)^2$$

Similar to the method of OLS, one can now take the partial derivative of  $P$  in terms of  $\beta$  and  $\alpha$

and set the two resulting equations to 0. One can then obtain the values of these two parameters, which maximizes the Likelihood Function:

$$\beta = \frac{\sum x_t^2 \sum y_t - \sum x_t \sum y_t x_t}{n \sum x_t^2 - (\sum x_t)^2}$$

$$\alpha = \frac{n \sum y_t x_t - \sum y_t \sum x_t}{n \sum x_t^2 - (\sum x_t)^2}$$

### 3. Concept of R-sqaure and Hypothesis Testing

#### 3.1 R-square

After a linear equation is generated through LRA, techniques such as the calculation of R-square can be employed to verify its accuracy. R-square is calculated through the following equation:

$$R^2 = \frac{ESS}{TSS}$$

where the range of R-square goes from 0 to 1 inclusively. Note that ESS is called The Explained Sum of Squares and TSS The Total Sum of Squares [2]. The formula for calculating ESS is as follows:

$$\sum (y_t - \bar{y})^2$$

where  $y_t$  are the y-values of all data points and  $\bar{y}$  is the average of all y-values on the linear equation. The formula for calculating TSS is as follows:

$$\sum (\hat{y} - \bar{y})^2$$

where  $\hat{y}$  is the y-values on the linear equation. Generally speaking, the value of R-square is proportional to how accurately the linear equation represents all the given data points.

One can also employ the technique of Hypothesis Testing (HT) to verify assumptions made concerning the whole population based on the data given by a sample culled from the population [5]. The basic procedure of conducting a HT is 1. A hypothesis, called the Null Hypothesis (NH), is made based on given circumstances and an Alternative Hypothesis (AH) is suggested should the NH be proven as invalid, 2. Determine the veracity of the NH according to the information provided by the sample, and 3. Conclusions are made whether to accept the NH or to reject it in favor of the AH. The technique of HT makes use of the concept of small probability. It assumes that if an event, with a small enough probability of it ever occurring, were conceived under the NH assumption, NH is rejected if this event turns out to be true or is accepted if it is otherwise.

The technique of HT can be carried out through two methods: Test of Significance Approach (TSA) and Confidence Interval Approach (CIA).

#### 3.2 Test of Significance Approach

The TSA calculates a t value, which is subject to t distribution, according to the formula  $t = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$ , where  $\hat{\beta}$  is the parameter of the sample estimated by using one of the estimation methods mentioned below,  $\beta^*$  the NH of the parameter of the population, and  $SE(\hat{\beta})$  the standard deviation of the estimated parameter of the sample. If the resulting t value is located within a region that identifies it as a small probability event, meaning that there is small probability of ever obtaining this t value, then the NH is rejected. Formulas for calculating the components within the t-value formula are included below. Formula for calculating the standard deviation of  $\beta$  is as follows:

$$\sqrt{\frac{\sum(\hat{u}_t)^2}{T-2}} \sqrt{\frac{1}{\sum x_t^2 - N\bar{x}^2}}$$

where  $\sqrt{\frac{\sum(\hat{u}_t)^2}{T-2}}$  is the standard deviation of the error terms,  $\hat{u}_t$  the error terms, N the number of data points,  $x_t$  the x values of all the sample points, and  $\bar{x}$  the average of the x values of all the

sample points. A NH  $\beta = \beta^*$  and an AH  $\beta \neq \beta^*$  are proposed before values of all the components for calculating the t value are obtained.

After the t value is obtained through the above method, a standard t value is established against which the obtained t value can be measured. Note that significance, which is the probability that we might be wrong to reject the null hypothesis, is usually set to 5 percent, which corresponds to 95% confidence, meaning that one is 95% sure that it is right to accept the NH. One can, therefore, refer to a t value distribution table for the standard t value at N-1 degrees of freedom and 95% confidence. One is to reject the NH and accept the AH if the obtained t value is bigger than the standard t value, indicating the event conceived under the NH assumption happens within the 5 percent area under the t distribution curve, namely, the event happens in spite of its being a small probability event. Otherwise, one is to accept the NH and reject the AH [2].

### 3.3 Confidence Interval Approach

Similar to the TSA, one needs to obtain the estimated value and standard deviation of  $\beta$  as well as to establish a NH,  $\beta = \beta^*$ , and an AH,  $\beta \neq \beta^*$ . Then one needs to assign a confidence (as mentioned above, usually 95%) to this test and to obtain the standard t value by referring to the t distribution table, which can then be plugged into the following formula to establish a confidence range:

$$(\hat{\beta} - t * SE(\hat{\beta}), \hat{\beta} + t * SE(\hat{\beta}))$$

If the NH is within the range indicated above, then the NH can be accepted. One is to reject the NH if it did not fall into the range.

### 3.4 Conclusion

Since even small probability event has a tiny probability of ever occurring, it is possible to make two mistakes when using the two methods above. The first possible mistake is that the NH is in fact correct, but one had rejected the NH because it has fell in the five percent area [6]. The second possible mistake is that the NH is actually incorrect, but one had accepted the NH due to the fact that it falls into the confidence range.

## 4. Summary

Each of the three methods discussed above were conceived in situation when the other two are not applicable. The advantage of MoM and OLS is their simplicity. In the case when probability distribution is either inaccessible or can only be obtained through complex computer analysis, MoM and OLS are by far simpler than MLE since they can be carried out readily by hand. MLE, however, yields more accurate result, so OLS becomes inferior to the MLE when the probability distribution to which the parameter is subject is known [3]. This is because OLS, although the assumption that the error terms are subject to normal distribution is not required, the estimation is more accurate with the assumption [2]. MLE, however, does not come with this limitation. Moreover, the accuracy of MoM decreases as the sample size gets smaller, while the accuracy of MLE is not affected by varying sample size. The estimation methods discussed above were traditional means through which one can estimate the behavior and properties of the population without the necessity of obtaining the data with size of or near the whole population. Since these methods are only estimation, the concept of R-square and the Test of Significance Approach were developed to verify their accuracy.

Since linear regression is one of the most important topic under the study of statistics, it is not surprising that its application is prevalent in many fields that rely on the analysis of collected data. The most apparent case would be linear regression's ability to generate a trend line that delineates the linear behavior of scattered data. Analysts can also verify the proposed trend line by collecting their data to see if it corresponds to the proposed linear relationship [7]. Another application of linear regression, of course, can be seen in the field of economics. For instance, a simple yet much relevant relationship between gross domestic product growth and unemployment expressed by Okun's Law is a linear and negative correlation [8].

## References

- [1] Leng, Ling, et al. "Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science." 2007:012084-012088.
- [2] Logan, Murray. "Correlation and Simple Linear Regression." *Methods in Molecular Biology* 404.3(2010):143-64.
- [3] Myung, In Jae. "Tutorial on maximum likelihood estimation." *Journal of Mathematical Psychology* 47.1(2003):90-100.
- [4] Wooldridge, Jeffrey M. "Applications of generalized method of moments estimation." *The Journal of Economic Perspectives* 15.4 (2001): 87-100.
- [5] Bakan, David. "The test of significance in psychological research." *Psychological bulletin* 66.6 (1966): 423.
- [6] Shannon, Harry, and Stephen Walter. "Basic statistics for clinicians: hypothesis testing." *Can Med Assoc J* 152.1 (1995): 27.
- [7] Vallin, Jacques, and France Meslé. "The Segmented Trend Line of Highest Life Expectancies." *Population and Development Review* 35.1 (2009): 159-87. Web.
- [8] Knotek, Edward S., II. "How Useful is Okun's Law?" *Economic Review - Federal Reserve Bank of Kansas City* 92.4 (2007): 73,103,3. *ProQuest*. Web. 7 Oct. 2016.