Mikailalsys Journal of
Advanced Engineering International

e-ISSN : 3046-9694
p-ISSN : 3046-8914

Index : Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/MJAEI.v2i2.5401

Regression Analysis and Forecasting with Regression Model in Economics

Nand Kishor Kumar[1*], Raj Kumar Shah[2], Suresh Kumar Sahani[3]
[1]Tri-Chandra Campus, Tribhuvan University, Nepal; [2]Ratna Raj Laxmi Campus, Exhibition Road, Nepal; [3]MIT Campus, T.U. Janakpurdham-45600 Nepal
nandkishorkumar2025@gmail.com; rajkumarshah90111@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Mar 13, 2025	Mar 28, 2025	Apr 9, 2025	Apr 14, 2025

Abstract

This work aims to provide a mathematical model that can be applied to prediction and defines this relationship. It helps economists understand how different factors influence economic indicators such as GDP, inflation, unemployment, and market trends. Forecasting using regression models provides valuable insights for policy-making, business strategies, and economic planning.

Keywords: Regression analysis, Forecasting Models, Regression model, Lines of Regression, Regression Equation, Multiple Regression Equation

Introduction

Regression analysis is a statistical technique usesd to model and analyze the relationships between a dependent variable (target) and one or more independent variables (predictors). It helps in predicting outcomes, identifying trends, and understanding how variables interact. It also determines the nature and strength of relationship between two variables.

Thus, regression is the estimation of unknown values or prediction of one variable from known values of other variables. For examples, the yield of a crop depends on the amount of rainfall, expenditure of a person depends on his income etc. The regression analysis confined to the study of only two variables at a time is called simple regression.

The known value which is used for prediction (or estimation), is called independent (or regressor or predictor or explanator) variable and the unknown value which is to be estimated (predicted) by known value is called dependent (or regressed, explained) variable. For examples: height of sons, yield of crops, expenditure etc. are dependent variables whereas height of fathers, amount of rainfall, income etc. are independent variables (Gujarati, & Porter,2009).

Regression analysis involves examining the relationship between a dependent variable (the outcome of interest) and one or more independent variables (factors that influence the outcome). The goal of this paper is to establish a mathematical model that describes this relationship and can be used for predictive purposes.

Preliminaries

Lines of Regression

A line fitted to a set of data points to estimate the relationship between two variables is called regression line. In other words, the device used for estimating the value of one variable from the value of the other consists of a line through the points, drawn in such a manner as to represent the average relationship between two variables. Such a line is called the line of regression.

A line fitted by the method of least squares is the line of best fit. A line of regression gives the best estimate of one unknown variable for any given value of the other variable. As simple regression consists two variables only there are always two lines of regression, one of Y on X and the other of X on Y. The line of regression of Y on X is used to estimate (or predict) the value of dependent variable Y for any given value of independent variable X. Similarly, the line of regression of X on Y gives the best estimate for the value of dependent variable X for any given value of independent variable Y.

Regression lines are expressed algebraically by means of equation called regression equations (or prediction equations or estimating equations). Since there are two regression lines, there are also two equations namely the regression equation of Y on X and the regression equation of X on Y. The regression equation of Y on X is used to describe the change in the value of Y for given change in the value of X whereas the regression equation of X on Y is used to describe the change in the value of X for given variations in the value of Y (Chatterjee, & Hadi, 2015).

Regression Equation of Y on X

The equation of regression line where the dependent variable Y is determined by the independent variable X is

$$Y = a + bX \quad (1)$$
$$a = Y - \text{ intercept}$$

$b =$ slope of the regression line (i.e. it measures the change in Y per unit change in X) or regression coefficient of Y on X which is denoted by b_{yx}.

According to the assumption of least squares, two normal equations for estimating two numerical constants a and b are given

$$\sum Y = n\,a + b\sum X \quad (2)$$

and

$$\sum XY = a\sum X + b\sum X^2 \quad (3)$$

where n is the number of pair observations. Solving these two normal equations, we get

$$a = \bar{Y} - b\bar{X} \quad (4)$$

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \quad (5)$$

Now, putting the values of a and b in equation (1), we get the required estimated regression equation of Y on X which is as follows.

$$\hat{Y} = a + bX \quad (6)$$

where $\hat{Y} =$ estimated or calculated value of Y.

Thus, the line obtained by substituting the values of numerical constants a and b in (1) is called line of best fit.

Regression equation of X on Y

Let the regression equation of X on Y be

$X = a' + b'Y$ (7)

where X = dependent variable

Y = independent variable

The numerical constants a' and b' can be obtained by solving following two normal equations.

$$\Sigma X = na' + b'\Sigma Y$$

(8)

$$\Sigma XY = a'\Sigma Y + b'\Sigma Y^2$$

(9)

and

Solving these two normal equations, we get

and

$$a' = \overline{X} - b'\overline{Y}$$

(10)

$b' = \frac{n\sum XY-(\Sigma X)(\Sigma Y)}{n\sum Y^2-(\sum Y)^2}$ (11)

The numerical constant b' is also called regression coefficient of X on Y which is denoted by b_{xy}. Now, substituting the values of a' and b' in equation (7), we get the required estimated regression equation of X on Y as

$\hat{X} = a' + b'Y$ (12)

This derivation can be summarized in table

General Equations	$Y = a + bX$	$X = a' + b'Y$
Normal Equations (i)	$\sum Y = na + b\Sigma X$	$\sum X = na' + b'\Sigma Y$
(ii)	$\sum XY = a\sum X + b\Sigma X^2$	$\sum XY = a'\sum X + b'\Sigma X^2$

Now taking an example of two regression equations from the following data.

X:	1	2	3	4	5
Y:	1	3	5	7	9

Let the regression equation of Y on X be

$$Y = a + bX$$

Then, two normal equations estimating a and b are

$$\left.\begin{array}{r} \Sigma Y = na + b\Sigma X \\ \text{and } \Sigma XY = a\Sigma X + b\Sigma X^2 \end{array}\right\}$$

Also, let the regression equation of X on Y be

$$X = a' + b'Y$$

Then, two normal equations estimating a and b are

$$\left.\begin{array}{r} \sum \quad X = na' + b'\Sigma Y \\ \text{and } \Sigma XY = a'\Sigma Y + b'\Sigma Y^2 \end{array}\right\}$$

Calculation of regression equations -

X	Y	X^2	Y^2	XY
1	1	1	1	1
2	3	4	9	6
3	5	9	25	15
4	7	16	49	28
5	9	25	81	45
$\sum X = 15$	$\sum Y = 25$	$\sum X^2 = 55$	$\sum Y^2 = 165$	$\sum XY = 95$

Here, $n = 5, \Sigma X = 15, \Sigma Y = 25, \Sigma X^2 = 55, \Sigma Y^2 = 165, \Sigma XY = 95$

Putting these values, we get

and

$$5a + 15b = 25$$
$$a + 3b = 5$$
$$15a + 55b = 95$$
$$3a + 11b = 19$$

Solving, we get

$$3a + 11b = 19$$
$$3a + 9b = 15$$
$$- \quad - \quad -$$
$$2b = 4$$
$$b = 2.$$

Putting the value of b in (5), we get

$$a + 3 \times 2 = 5$$
$$a = -1.$$

Substituting the values of a and b in (1), we get

$\hat{Y} = -1 + 2X$. This is the required estimated regression equation of Y on X.

Again, from (4), we get

$$5a' + 25b' = 15$$
$$25a' + 165b' = 95$$
$$5a' + 33b' = 19$$

or

Solving these two equations, we get

$$a' = \frac{1}{2} \text{ and } b' = \frac{1}{2}$$

Putting these values of a' and b' , we get

$\hat{X} = \frac{1}{2} + \frac{1}{2}Y$. This is the required estimated regression equation of X on Y.

Multiple Regression Analysis

Assuming that the variables are closely related, we can estimate the unknown value of one variable from the given or known values of the other variables. Multiple regression analysis is a logical extension of the simple linear regression analysis. In multiple regression analysis, instead of a single independent variable, two or more independent variables are used to estimate the unknown values of a dependent variable. The fundamental concepts in multiple regression are similar to those of simple regression. The following are the main purposes of multiple regression analysis.

(1) To establish a regression equation which provides estimates of the dependent variable from the values of two or more independent variables.

(2) To obtain measures of error involved in using this regression as a basis for estimation of the dependent variable (i.e. to examine the multiple regression standard error of estimate)

(3) To measure the coefficient of multiple determination or the proportion of variation in the dependent variable which is explained by the independent variable.

Multiple Regression Equation

The procedure for studying multiple regression is similar to the one we have for simple regression, with the difference that the other variables are added in the regression equation.

This regression model assumes that the relationship between each independent variable and the dependent variable is linear in nature and the regression line being calculated by the method of least squares.

The multiple regression equation describes the average relationship between one dependent variable and two or more independent variables and this relationship is very much useful for estimating (or predicting) the dependent variable. Thus, a multiple regression equation of X_1 on X_2 and X_3 is an equation for estimating a dependent variable X_1 from two independent variables X_2 and X_3.

The multiple regression equation of dependent variable X_1 on two independent variables X_2 and X_3 is given by

$$X_1 = a_1 + b_1X_2 + b_2X_3 \qquad (13)$$

Here, b_1 and b_2 are known as coefficient of net regression which implies that the regression of the dependent variable on the particular independent variable is measured while holding the value (s) of the other variable (s) constant.

The values of the constants a_1, b_1 and b_2 can be obtained by solving following three normal equations simultaneously obtained by the method of least squares (Stock, & Watson,2019).

$$\left.\begin{aligned} \sum X_1 &= na_1 + b_1\sum X_2 + b_2\sum X_3 \\ \sum X_1X_2 &= a_1\sum X_2 + b_1\sum X_2^2 + b_2\sum X_2X_3 \\ \sum X_1X_3 &= a_1\sum X_3 + b_1\sum X_2X_3 + b_2\Sigma X_3^2 \end{aligned}\right\} \qquad (14)$$

Solving these three normal equations simultaneously after substituting the values of $n, \Sigma X_1, \Sigma X_2, \sum X_3, \sum X_2^2, \sum X_3^2, \sum X_1X_2, \sum X_1X_3$ and $\sum X_2X_3$, we get the values of a_1, b_1 and b_2. Putting the values of a_1, b_1 and b_2 in equation. we get required estimated multiple regression equation of X_1 on X_2 and X_3. Also, the multiple regression equation of dependent variable X_3 on two idependent variables X_1 and X_2 is given by

$$x_3 = a_1 + b_1X_1 + b_2x_2$$

Then, following three normal equations can be obtained by method of hast squares.

$$\left.\begin{array}{l} \Sigma X_3 = na_1 + b_1\Sigma X_1 + b_2\Sigma X_2 \\ \Sigma X_1X_3 = a_1\Sigma X_1 + b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 \\ \Sigma X_2X_3 = a_1\Sigma X_2 + b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 \end{array}\right\}$$

Solving these three normal equations simultaneously, we get values of $4, b_1$ and b_2. Substituting the values of a_1, b_1 and b_2 in (5), we get the sulired estimated regression equation of X_3 on X_1 and X_2 (Greene,2018).

Method and Discussion

Regression analysis is widely used in forecasting to predict future values based on historical data. It helps establish relationships between a dependent variable (what you want to predict) and one or more independent variables (predictors).

Forecasting with Regression Model

The regression model is an explicitly multivariate model, in which variables are explained and forecast on the basis of their own history with their related variables (Wooldridge, 2020).

$$Y_t = \beta_0 + \beta_1X_t + \varepsilon_t, \text{ where, } \varepsilon_t \sim N(0, \sigma^2)$$

i. **Conditional Forecasting Models**

A conditional forecasting model is one that can be used to produce forecasts for a variable of interest, conditional upon assumptions about other variables. With the regression model,

$$y_t = \beta_0 + \beta_1x_t + \varepsilon_t$$
$$\varepsilon_t \text{ is } N(0, \sigma^2)$$

our h-step ahead conditional forecast for y, given that the h-step value of x is is $x^*T + h$

$$y_{T+h,T} \mid x_{T+h}^* = \beta_0 + \beta_1x_{T+h}^*$$

ii. **Unconditional Forecasting Models**

Frequently, we do not want to make forecasts of Y conditional upon assumptions about x, rather, we just want the best possible forecast of y-an unconditional forecast. To get an unconditional forecast from a regression model, we often encounter the forecasting the right-hand-side variables problem. That is, to get an optimal unconditional point forecast for Y, we cannot insert an arbitrary value for future x, rather, we need to insert the optimal point forecast, $X_{T+h,T}$, which yields the unconditional forecast (Box, Jenkins, & Reinsel, 2015).

$$Y_{T+h,T} = \beta_0 + \beta_1 X_{T+h,T}$$

We usually don't have such a forecast for x and the regression model at hand doesn't help us.

a. **Averaging methods and Exponential Smoothing Methods**

We suppose that we are currently at time t and that we wish to use the data up to this time, ie $Y_1, Y_2, \dots, Y_{t-1}, Y_t$, to make forecasts $F_{t+1}, F_{t+2}, \dots, F_{t+m}$ of future values of Y.

However, it is convenient to follow this convention it is important to realize at the outset that his distinction is artificial in that all the methods. They are thus, all similar to the moving averages considered in this paper. The difference is that the averages are used for forecasting rather than for describing past data.

b. **Averaging Methods**

The moving average forecast of order k, which we write as $MA(k)$, is defined as

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^{t} Y_i$$

This forecast is only useful if the data does not contain a trend-cycle or a seasonal component. In other words, the data must be stationary. Data is said to be stationary if Y_t which is a random variable, has a probability distribution that does not depend on t. A convenient way of implementing this forecast is to note that (Dougherty, 2016).

$$F_{t+2} = \frac{1}{k} \sum_{i=t-k+2}^{t+1} Y_i = F_{t+1} + \frac{1}{k}(Y_{t+1} - Y_{t-k+1})$$

This is known as an updating formula as it allows a forecast value to be obtained from the previous forecast value by a simpler calculation than using the defining expression. The only point of note is that moving average forecasts give a progressively smoother forecast as the order increases, but a moving average of large order will be slow to respond to real but rapid changes. Thus, in choosing k, a balance has to be drawn between smoothness and ensuring that this lag is not unacceptably large.

c. **Single Exponential Smoothing Method**

The single exponential forecast or single exponential smoothing (SES) is defined as

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

where α is a given weight value to be selected subject to $0 < \alpha < 1$. So, F_{t+1} is the weighted average of the current observation, Y_t, with the forecast, F_t, made at the previous time point $t-1$.

Repeated application of the formula yields

$$F_{t+1} = (1-\alpha)^t F_1 + \alpha \sum_{j=0}^{t-1} (1-\alpha)^j Y_{t-j}$$

Shows the dependence of the present forecast on Y_t, Y_{t-1} , Y_{t-2} ⋯ ⋯falls in exponential way.

Applications of Regression Analysis in Economic Forecasting

GDP Forecasting: Predicting future economic growth based on indicators like investment, consumption, and trade balance.

Inflation Prediction: Estimating future inflation trends using variables like money supply, interest rates, and consumer demand.

Unemployment Rate Analysis: Understanding how changes in economic conditions affect employment levels.

Stock Market Analysis: Predicting stock price movements based on historical performance and economic indicators.

1. **Challenges and Considerations in Economic Forecasting**

Data Limitations: Availability and accuracy of economic data can impact model reliability.

Multicollinearity: High correlation among independent variables may distort results.

External Shocks: Unpredictable events such as natural disasters or policy changes can affect forecasts.

Model Overfitting: Complex models may fit past data well but fail to predict future trends accurately.

2. Steps in Forecasting with Regression Analysis

Step 1: Collect and Prepare Data

Gather historical data on the dependent and independent variables.

Clean and preprocess the data (handle missing values, remove outliers, etc.).

Step 2: Choose the Type of Regression Model

Simple Linear Regression (one independent variable)

Multiple Linear Regression (multiple independent variables)

Polynomial Regression (for nonlinear relationships)

Time Series Regression (if data is time-dependent)

Step 3: Train the Model

Split the data into training and testing sets.

Fit the regression model using statistical software (e.g., Python, R, Excel).

Estimate model parameters (a, b) using methods like Ordinary Least Squares (OLS).

Step 4: Validate the Model

Evaluate the model's accuracy using metrics like:

R^2coefficient of determination) – measures how well the model explains variance in Y.

RMSE (Root Mean Square Error) – measures prediction error.

p-values and t-tests – assess the significance of predictors.

Step 5: Make Forecasts

Use the trained model to predict future values.

Plug new values of X into the model equation to estimate Y.

Step 6: Interpret and Improve Forecasts

Compare predictions with actual results.

Adjust the model by adding new variables, transforming data, or using different regression techniques.

Conclusion

Regression analysis is a powerful tool in economics for understanding relationships between variables and making forecasts. By leveraging regression models, economists and policymakers can make informed decisions that drive economic stability and growth. However, careful model selection, validation, and consideration of external factors are essential for improving forecasting accuracy.

References

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). McGraw-Hill.

Wooldridge, J. M. (2020). Introductory Econometrics: A Modern Approach (7th ed.). Cengage Learning.

Stock, J. H., & Watson, M. W. (2019). Introduction to Econometrics (4th ed.). Pearson.

Greene, W. H. (2018). Econometric Analysis (8th ed.). Pearson.

Chatterjee, S., & Hadi, A. S. (2015). Regression Analysis by Example (5th ed.). Wiley.

Enders, W. (2014). Applied Econometric Time Series (4th ed.). Wiley.

Dougherty, C. (2016). Introduction to Econometrics (5th ed.). Oxford University Press.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.