Basics of Multiple Regression & Underlying Assumptions

1. INTRODUCTION

Multiple linear regression –
models the linear relation b/w dependent variable & multiple independent variables.
- o Multiple regression predicts returns, improves portfolio, and explains security drivers.
- o If misused, can yield inaccurate predictions.
- o The analyst specifies the variables and use software to estimate the model.

2. USES OF MULTIPLE LINEAR REGRESSION

- Financial and economic relationships are complex and require rigorous statistical and theoretical scrutiny.
- Multiple regression helps to identify relationships between variables and test theories.

The Regression Process

The objective is to use the variation of independent variable(s) to explain the variation in the dependent variable.

Is the dependent variable continuous? → No – use logistic regress
Yes – use traditional regression model

Does the model meet the key assumptions? → No – adjust the model
Yes – examine model's goodness of fit

Is the overall fit significant? → No – adjust the model
Yes – determine is this model optimal

Is the model optimal? → No – adjust the model
Yes – use the model for analysis

3. THE BASICS OF MULTIPLE REGRESSION

A multiple linear regression model has the following general form:

$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_kX_{ki} + \varepsilon_i, i = 1, 2, \ldots n$

- **Slope coefficient,** measures how much Y, changes when X_j, changes by one unit, holding other independent variables constant.
  - **Intercept** is the value of Y when independent variables are all equal to zero.
  - A regression equation has k slope coefficients and one intercept i.e., k + 1 regression coefficients.

4. ASSUMPTIONS UNDERLYING MULTIPLE LINEAR REGRESSION

Five Key Assumptions	
Linearity	Relation between dependent and the independent variables is linear.
Homoskedasticity	The variance of residuals is the same for all observations
Independence of errors	Residuals are uncorrelated across observations
Normality	Residuals must be normally distributed
Independence of independent variables	Independent variables are not random and have no linear relationship.

Normal Q-Q Plot: A Q-Q plot is used to compare the distribution of a variable to a normal distribution.