Level II of the CFA® 2025 Exam

Study Notes - Quantitative Methods

Offered by AnalystPrep

Last Updated: Feb 28, 2025

Table of Contents

Reading 1: Basics of Multiple Regression and Underlying Assumptions

Los 1 (a) Describe the types of investment problems addressed by multiple linear regression and the regression process

Multiple linear regression describes the variation of the **dependent variable** by using **two or more independent variables**. When used properly, it can improve predictions. However, if used incorrectly, it can create spurious relationships that can undermine predictions.

Typically, a multiple regression model takes the following form:

$$Y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \ldots + b_kX_{k,i} + \epsilon_i$$

Where:

Y_i = Dependent variable.

b_0 = Intercept term.

$b_1, b_2, \ldots, b_k$ = Slope coefficients.

$X_{1,i}, X_{2,i}, \ldots, X_{k,i}$ = Independent variables.

ϵ_i = Error term.

n = Number of observations.

A regression equation has k slope coefficients and k + 1 regression coefficients.

The intercept term is defined as the value of the dependent variable when the independent variables are zero. On the other hand, the slope coefficient is defined as the estimated change in the dependent variable given a one-unit change in the independent variable, keeping the other independent variables constant.

Researchers can use multiple regression to test existing theories, identify relationships between variables, or forecast.

The researcher must specify the model to determine the good-fit criteria for the regression

model, including an independent variable. Once the regression model has been specified, it must be estimated and analyzed to ensure it satisfies all the key assumptions.

It is equally noteworthy that a researcher can use multiple regression to test existing forecasting theories. Alternatively, multiple regression can further be used to identify relationships between variables after the model is tested and deemed acceptable for out-of-sample performance.

A single factor cannot adequately explain or forecast the complex world of investments. Due to their complexity, statistical tests and fundamental justification are critical in the exhaustive explanation of financial and economic relations.

There are several ways to use multiple regression, including:

- A company's profitability, growth, revenue, and market share are variables that an investor is interested in. These variables can predict if a company will run into financial difficulties.
- An analyst seeking to determine how a company's stock price and trading volume change daily can correlate them. An analyst can use linear regression to determine the relationship between variables.

Question

Which of the following *most* accurately detects whether the underlying assumptions of multiple linear regression models are satisfied?

A. Scatter plots.
B. Residual plots.
C. Diagnostic plots.

Solution

The correct answer is **C**.

Diagnostic plots for multiple regression show the prediction errors against predicted values. Therefore, they are useful in the determination of the gaps that a researcher should address in their data to improve the accuracy of their predictions. Using diagnostic plots, a researcher can determine whether the assumptions of multiple linear regression are valid.

A is incorrect. The scatterplot is useful for detecting nonlinear relationships between dependent and independent variables.

B is incorrect. A residual plot is an effective tool for detecting violations of homoskedasticity and error independence.

LOS 1 (b) Formulate a multiple linear regression model, describe the relation between the dependent variable and several independent variables, and interpret estimated regression coefficients

Example: Multiple Regression in Investment World

James Chase, an investment analyst, wants to determine the impact of inflation rates and real rates of interest on the price of the US Dollar index (USDX).

Chase uses the multiple regression model below:

$$P = b_0 + b_1 INF + b_2 IR + \epsilon_t$$

Where:

P = Price of USDX.

INF = Inflation rate.

IR = Real rate of interest.

ϵ_t = Error term.

The regression of the price of USDX on inflation and real interest rates generates the following results:

	Coefficients	Standard Error	t Stat	P-value
Intercept	81	7.9659	10.1296	0.0000
Inflation rates	−276	233.0748	−1.1833	0.2753
Real interest Rates	902	279.6949	3.2266	0.0145

Chase can express the multiple regression equation as follows:

$$P = 81 - 276INF + 902IR$$

The regression coefficient estimate of the inflation rate is negative. This indicates that an increase in the inflation rates causes a decrease in the price of the US Dollar index (USDX).

Furthermore, the positive real rate of interest coefficient implies that an increase in the price of USDX accompanies the real interest rate.

The t-statistic indicates that only the real interest rate variable is significant at the 5% significance level.

The **intercept term** is defined as the value of the dependent variable when the independent variables are zero. On the other hand, each **slope coefficient** is the estimated change in the value of the dependent variable for a one-unit change in the value of the respective independent variable, keeping the other independent variables constant. Slope coefficients are also called **partial slope coefficients**.

Continuing with this example:

$$P = 81 - 276INF + 902IR$$

Where:

P = Price of the US Dollar index.

INF = Annual inflation rate.

IR = Annual real rate of interest.

The regression equation is interpreted as follows:

The intercept term of 81 implies that the price of USDX is $81 when both the inflation rate and real interest rate are 0.

A 1% increase in the inflation rate leads to a $276 decrease in the price of USDX, keeping real interest rates constant. On the other hand, a 1% increase in the real interest rate leads to a $902 increase in the price of USDX, keeping the inflation rate constant.

Question

Adil Suleman, CFA, wishes to establish the possible drivers of a company's percentage return on capital (ROC). Suleman identifies performance measures such as the profit margin (%), sales, and debt ratio as possible drivers of ROC.

He obtains the following results from the regression of ROC on profit margin (%), sales, and debt ratio.

SUMMARY OUTPUT

	Coefficients	Standard Error	t Stat	P-value
Intercept	8.6531	0.9174	9.4323	0.0000
Sales	0.0009	0.0005	1.7644	0.0922
Debt ratio	0.0229	0.0165	1.3880	0.1797
Profit Margin (%)	0.2996	0.0564	5.3146	0.0000

Which independent variable(s) is (are) most likely statistically and significantly different from zero at the 5% significance level, assuming the sample size is 25?

A. Profit margin.
B. Sales and profit margin.
C. Sales and debt ratio.

Solution

The correct answer is **A**.

An independent variable is statistically significant if its p-value is less than the significance level, in this case, 5% or 0.05. Therefore, only the profit margin is statistically and significantly different from zero at the 5% significance level.

B and C are incorrect. At a 5% significance level, only the profit margin is statistically significantly different from zero.

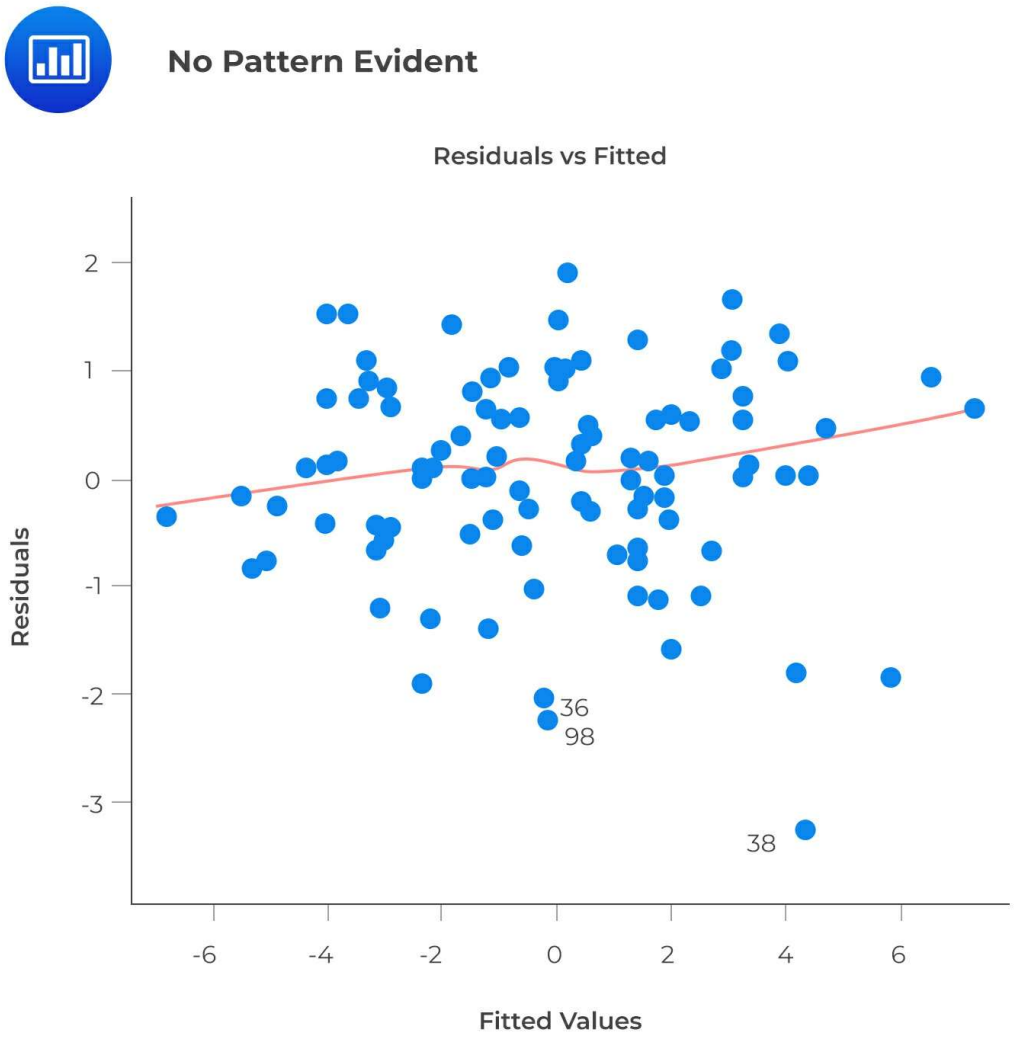
Los 1 (c) Explain the assumptions underlying a multiple linear regression model and interpret residual plots indicating potential violations of these assumptions

The following assumptions are used to build multiple regression models:

- The relationship between the dependent variable, and the independent variables, is linear.
- The independent variables are not random.
- There is no definite linear relationship between two or more independent variables. A high correlation between two or more independent variables is known as multicollinearity.
- The expected value of the error term, conditional on the independent variables, is equal to 0.
- The variance of the error term is equal for all observations. This is known as the homoskedasticity assumption.
- The error term is uncorrelated across all observations.
- The error term is normally distributed.

The following assumptions are investigated, and their outcomes are analyzed (if violated) as follows:

i. ***Linearity***: The regression algorithm would mathematically fail to capture the trend when fitted to a nonlinear, non-additive data set.



A prediction based on unobserved data will also be incorrect. The model can capture the nonlinear effect by including polynomial terms.

This plot may show non-linearity if any pattern (such as a parabolic shape) appears. In other words, the model fails to capture nonlinear effects.