Part 2. Data Analysis (30%)

Stock Quant Report - Model Estimation and Testing

• Given an asset and an index series undertake a graphical and descriptive statistical analysis of your logarithmic stock returns. Are the returns normally distributed?

(Hint: you can use graphical procedures, moments & normality tests)

• Use log-returns on your asset (R_t) and on the index (Rm_t) in order to estimate the market model regression:

$$R_t = a + bRm_t + u_t$$

Discuss and interpret the regression output (coefficients, standard errors, t-stats, *p*-values) and basic diagnostics (R^2, F-stat).

• Are the assumptions of OLS valid in terms of residual normality and heteroskedasticity? Is your asset/stock "neutral" in terms of systematic risk? (Hint: test if b=1)

Part 3. Presentation (20%)

"Bitcoin is a cryptocurrency, a digital asset designed to work as a medium of exchange that uses cryptography to control its creation and management, rather than relying on central authorities. The presumed pseudonymous Satoshi Nakamoto integrated many existing ideas from the cypherpunk community when creating bitcoin. In January 2009, the bitcoin network came into existence with the release of the first open source bitcoin client and the issuance of the first bitcoins." Source: https://en.wikipedia.org/wiki/History_of_bitcoin

On Friday 15 September 2017, Financial Times (FT) published an Editorial with title "In cryptocurrencies, tech and speculation meet" (see PDF file). As an analyst, you are asked to provide expert advice on financial trading, i.e. need to prepare a non-technical short presentation for the CEO of your company with only five (5) power-point slides with the following structure: 1 slide: setting the scene; 3 slides: facts & analysis; 1 slide: recommendations. Everything should be documented: please use an Appendix to include all sources (links & references) you have used.

can reject the null hypothesis and conclude that there is a significant linear relationship between X^2 and Y because the correlation coefficient is significantly different from zero.

- **H_0: $\beta_3 = 0$**
  H_1: $\beta_3 \neq 0$

  We firstly calculate for another time the test statistic, where β^* is the β enter the null hypothesis:

$$test\ statstic = \frac{\widehat{\beta_3} - \beta^*}{se(\widehat{\beta_3})} = \frac{0.07-0}{0.03} = 2.3333$$

  Test statistics derived in this way can be shown to follow a t- distribution with T – 2 degrees of freedom. The degrees of freedom for the model are **298**.

  Given a significance level of 5%, α = 0.05, we can determine a rejection region and a non-rejection region for a 2-sided test. We use as always a t-table to obtain a critical value with which we can compare the test statistic. The critical values are **-1.9679** and **+1.9679**. The test statistic lies in the rejection region. To conclude, we can reject the null hypothesis, which means that there is a significant linear relationship between P and Y because the correlation coefficient is significantly different from zero.

- **H_0: $\beta_4 = 0$**
  H_1: $\beta_4 \neq 0$

  The coefficient's β_4 p-value is **0,002**. According to the theory, if the p-value is less than the significance level (α = 0,05), we reject the null hypothesis. The p-value for the β_4 coefficient is less than the significance level, since p-value=0,002 < α = 0,05. Therefore, we can reject the null hypothesis and conclude that there is a significant linear relationship between Q and Y because the correlation coefficient is significantly different from zero, something which was expected since it is evident previously that β_4 was highly statistically significant (in significance level of 1%).

To sum up, we can present the above results to the following table:

	t – test	**p – value**	**Significant at 5%**	**Significant at 1%**
X	2.500	-	YES	NO
X^2	- 2.300	-	YES	NO
P	2.330	-	YES	NO
Q	-	0.002	YES	YES
Constant	1.733	-	NO	NO

against the below alternative hypothesis:

H_1: $\beta_1 + \beta_2 \neq 0$

According to the null hypothesis, it is set that **$\beta_1 + \beta_2 = 0$.** Consequently, the *t* statistic should now be based on whether the estimated sum $\hat{\beta}_1 + \hat{\beta}_2 = 0$ is significantly different from 0 in order to be able to reject H_0.

There are two procedures to perform the test with a single linear combination of parameters.[5] The first one based on the covariance matrix of the estimators, in contrast to the second one where the model is reparametrized by introducing a new parameter:

<u>1st Procedure:</u> Using Covariance Matrix of Estimators

Taking into account, firstly, the sampling error in our estimators, we must standardize this sum by dividing by its standard error:

$$t_{\widehat{\beta_1}+\widehat{\beta_2}} = \frac{\widehat{\beta_1} + \widehat{\beta_2}}{se(\widehat{\beta_1} + \widehat{\beta_2})}$$

We also have:

$$se(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{var\,(\widehat{\hat{\beta}_1 + \hat{\beta}_2})}$$

Where:

$$\widehat{var(\widehat{\beta_1} + \widehat{\beta_2})} = \widehat{var(\widehat{\beta_1})} + \widehat{var(\widehat{\beta_2})} + 2 * \widehat{covar(\widehat{\beta_1}, \widehat{\beta_2})}$$

Hence, to calculate $se\left(\widehat{\beta_1} + \widehat{\beta_2}\right)$, we need information on the estimated covariance of estimators. Most of the econometric software can display estimates of the covariance matrix of the estimator vector. If we have the covariance, we can calculate the t statistic. If the test statistic lies in the rejection, we can reject the null hypothesis. Finally, we have to notice that in this case the suitable type of test in order to exam the estimated sum $\hat{\beta}_1 + \hat{\beta}_2$ is a two-sided t-test.

<u>2nd Procedure:</u> Reparametrizing the Model by Introducing a New Parameter

[5] https://www.coursehero.com/file/27139011/4Hypothesistestinginthemultipleregressionmodel-1pdf/

In order to test for normality, we should use the residuals of the model regression. A residual of the regression is the difference between the actual values, which are the points in the left plot of Figure 7, and the predicted values, which fall on the red line. One of the assumptions of linear regression is that the residuals are drawn from a normal distribution; another way of saying this would be that the plot a histogram would be normal-like.

Figure 7: Graph of log_asset and fitted values

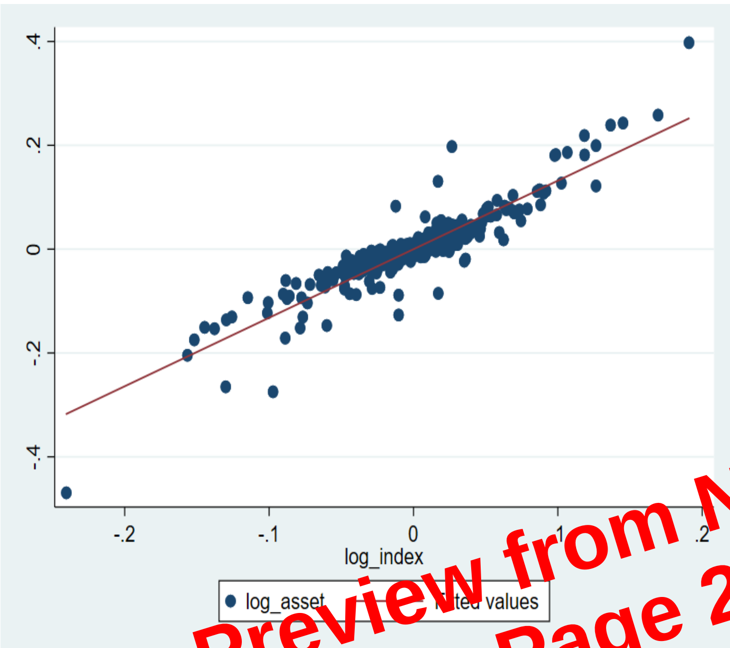
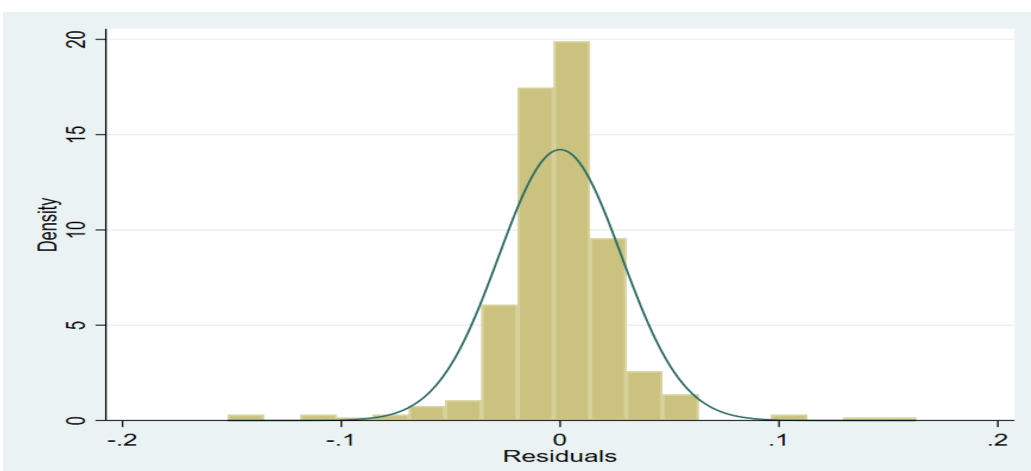


Figure 8: Graph of residuals

The histogram is one graphical way to say that the data comes from a normal distribution. Still, the histogram can be deceptive since changing the number of bins alter the shape of the distribution, and this may lead to some confusion. A bell-shaped curve shows the normal distribution of the series.



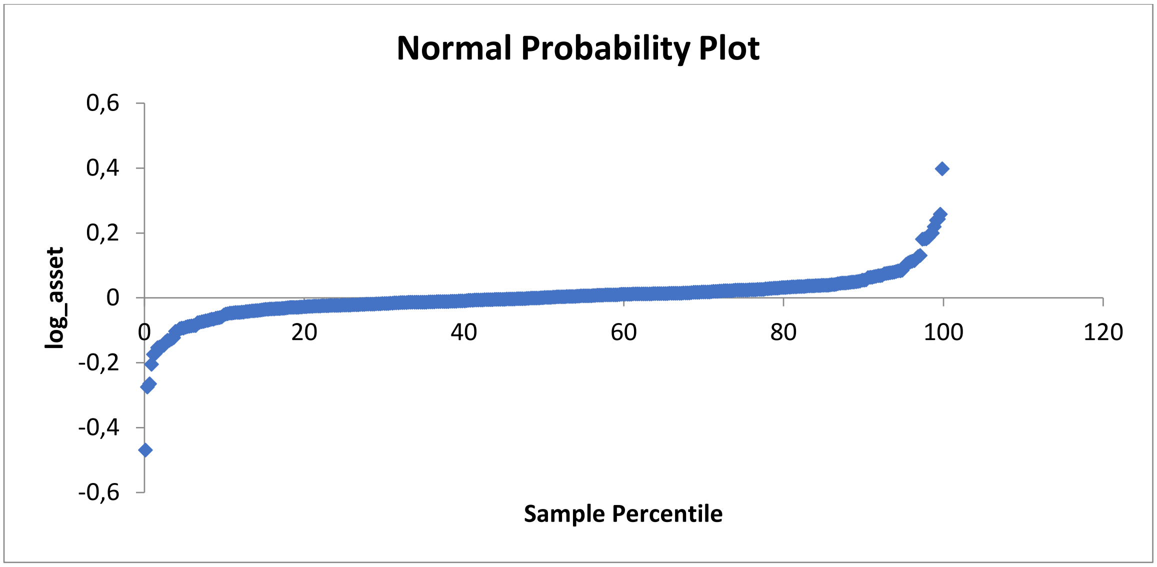


Figure 11: Normal Probability Plot

As we can observe from Figure 11, the data do not form a line. Therefore it could be concluded that the data are not normally distributed.

c. Quantile-quantile plot

An alternative to using a histogram is to use a quantile-quantile plot. A quantile-quantile plot (qq plot) is a scatter plot where we plot the dataset values vs normal distribution values for quantiles determined from the dataset. The y-coordinate of a qq plot is the dataset values, the x coordinates are values from the normal distribution.

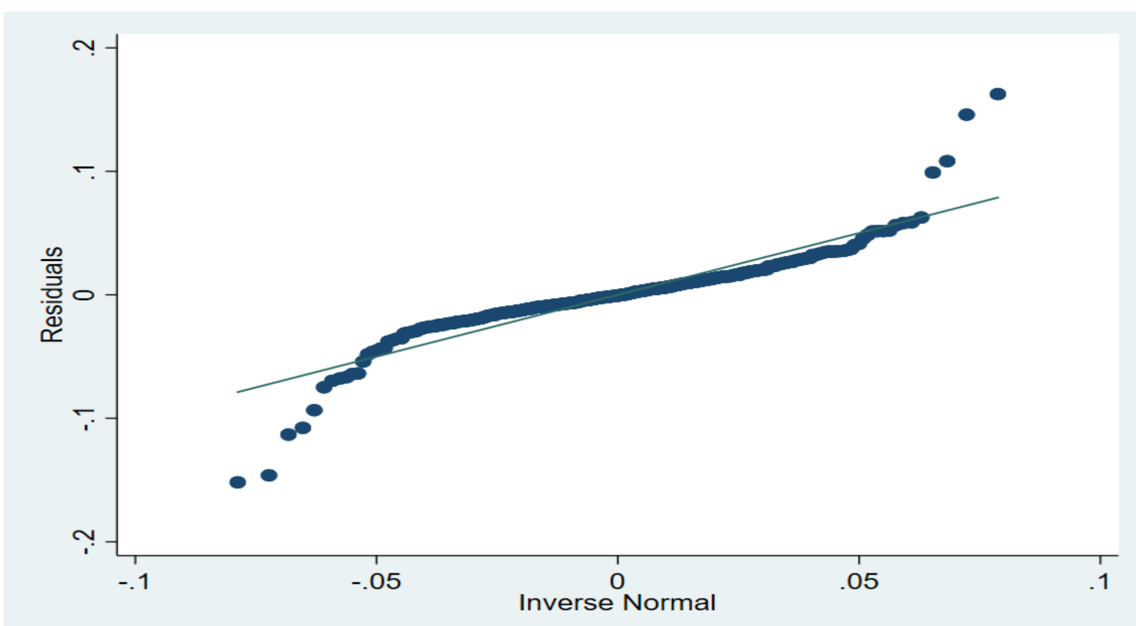


Figure 12: QQ plot – residuals