Modelling and Analysis of Stock Market Index Returns using Normal and Student T Distributions

Paul Michel, Candidate number: 250696

March 2023

1 Introduction

This report analyses two different models used to model the log returns for a stock market index on 1859 consecutive trading days (which excludes weekends and holidays).

The first method was to model the log returns as realisations of independent and identically distributed random variables with a Normal distribution. The second method was to model log returns as realisations of independent and identically distributed random variables with a Student T distribution.

2 Descriptive statistics

We first estimate the parameters of these distributions using the method of moments. We use the Normal distribution, $N(\mu, \sigma^2)$, for the first model. Using the method of moments to estimate the parameters of this model, we let $\bar{Y}$ and S^2 denote the sample mean and sample variance of the log returns, respectively. Then, the method of moments estimates and their values (rounded to 3 significant figures) are: $\hat{\mu} = \bar{Y}$, $\hat{\sigma}^2 = S^2$. The values of these estimates are (rounded to 3 significant figures): $\hat{\mu} = 4.32 \times 10^{-4}$, $\hat{\sigma}^2 = 6.33 \times 10^{-5}$. Figure 1 shows the distribution of the log returns for the first model. From this figure, we can assess the realism of this model which shows that the model approximates the data distribution closely. However, Figure 2, showing the quantile-quantile plot for the same distribution, reveals more about the model. Despite the Normal distribution showing a good approximation in the middle of the distribution, the tails of the data distribution are much heavier than the tails of the Normal distribution.

Therefore, the Normal distribution would be a bad choice for modelling log returns.

The second model considers modelling the log returns as realisations of independent and identically distributed random variables with a Student T distribution, $Stu(\alpha, \beta, \gamma)$, where $\beta > 0$ and $\gamma > 4$. The method of moments estimates for a Student T distribution with parameters α, β, and γ are given by: $\hat{\alpha} = \bar{Y}$, $\hat{\beta} = \sqrt{\frac{S^2(\gamma-2)}{\gamma}}$, $\hat{\gamma} = \frac{6-4x}{x-3}$. Where $\bar{Y}$ is the sample mean, S^2 is the sample variance, and x is the sample kurtosis. The values of these estimates are (rounded to 3 significant figures): $\hat{\alpha} = 4.32 \times 10^{-4}$, $\hat{\beta} = 6.57 \times 10^{-3}$, $\hat{\gamma} = 6.28$. Figure 3 shows the distribution of the log returns for this model. From this figure, it can be seen that the model approximates the data distribution closely. Digging deeper, as seen in Figure 4, based on the quantile-quantile plot for this distribution, Student T distribution is a better model. This is because the Student T distribution model captures the heavier tails for the model and hence provides a better fit for the data.

3 Conclusion

Using the preferred model, Student T distribution, we can calculate the value at risk and the probability that the price of a stock market index on a said day reduces by at least 3% of its value in one day. The value at risk is -1.99×10^{-2} (rounded to 3 significant figures) and the required probability is 1.47×10^{-3} (rounded to 3 significant figures). It is important however to note that the trustworthiness of these calculations and model should be taken with some caution. Since we are dealing with a statistical model that relies on assumptions, there may be other factors that affect the accuracy such as market conditions, unexpected events, etc. Hence, the model should not be used as a precise outlook of the future.

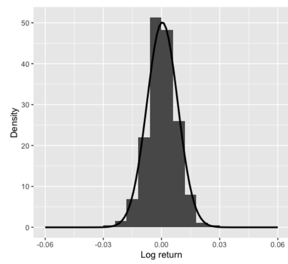

Figure 1: Distribution of the daily log returns for first model

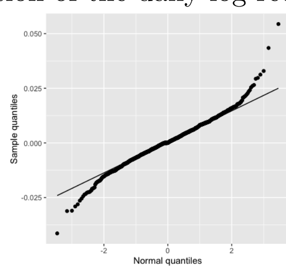

Figure 2: Quantile-quantile plot of log returns against a Normal distribution

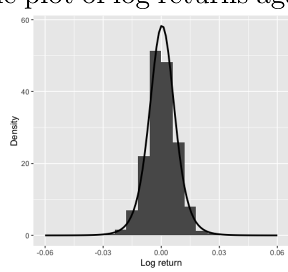

Figure 3: Distribution of the daily log returns for second model

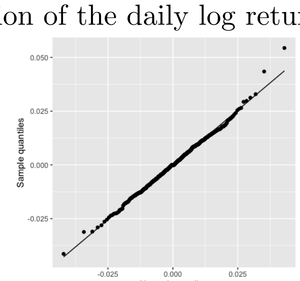

Figure 4: Quantile-quantile plot of log returns against a Student T distribution

Analysis on a study assessing the effectiveness of two treatments for arthritis

Paul Michel, Candidate number: 250696

March 2023

1 Introduction

In this report, we will analyse a study that aims to determine the effectiveness of two treatments for arthritis for adults in the UK. This study's sample comprised of adult patients in Devon who were referred for treatment for arthritis during a particular month. Patients were split into two groups: a control and an intervention group. The control group received the standard treatment while the intervention group received the new treatment. The sample was divided so that each patient was twice as likely to be put into the intervention group. Each patient's degree of flexion in their joints was measured before and after the three-week treatment and an increase in degrees was used to indicate improvement in their condition.

2 Strengths and weaknesses

2.1 Strengths

In this study, patients were randomly allocated to each of the two groups, the control and intervention groups. This helps to reduce selection bias and improve the validity of the study.

Another strength is that the study measured the degrees of flexion before and after the treatment. This reduces biases in the difference over the three-week period, decreasing the chances of a change being due to other factors.

2.2 Weaknesses

However, there are weaknesses to this study. Firstly, the study only comprised of adult patients in Devon. Therefore, the study population is unrepresentative of the target population, adults in the UK, giving rise to generalisation bias.

Moreover, the study only determined the effectiveness of each treatment over a three-week period. This disregards the possibility of long-term effects of the treatment and hence may skew the results of the study by not giving a conclusive result after each treatment.

3 Effects of the treatments

To evaluate the effectiveness of each treatment, we can use point estimates and confidence intervals for each treatment.

3.1 Control group

The sample mean and standard deviation of the degrees in flexion differences from before and after the treatment for the control group are $\bar{x} = -3.38$ and $s_1 = 5.31$ respectively. The approximate 95% confidence interval for the difference in mean degrees in flexion for the control group is

$\bar{x} \pm z_p\sqrt{\frac{s_1^2}{n_1}} = -3.38 \pm 1.9600\sqrt{\frac{5.31^2}{24}} = (-5.50, -1.26)$ rounded to 3 significant figures, since $z_p = 1.9600$ is the 2.5% quantile of the N(0,1) distribution. The point estimate indicates that the degree of flexion is about 3.38 degrees greater after the treatment in comparison to before.

3.2 Intervention group

The sample mean and standard deviation of the degrees in flexion differences from before and after the treatment for the intervention group are $\bar{y} = -6$ and $s_2 = 6.41$. The approximate 95% confidence interval for the difference in mean degrees in flexion for the intervention group is

$\bar{y} \pm z_p\sqrt{\frac{s_2^2}{n_2}} = -6 \pm 1.9600\sqrt{\frac{6.41^2}{52}} = (-7.74, -4.26)$ rounded to 3 significant figures, since $z_p = 1.9600$ is the 2.5% quantile of the N(0,1) distribution. The point estimate indicates that the degree of flexion is about 6 degrees greater after the treatment in comparison to before.

4 Comparing the effects of both groups

As seen from comparing the mean difference between both groups, the intervention saw a more significant increase in mean degrees of flexion. Moreover, the point estimates that have been calculated from the respective confidence intervals indicate a greater effect in the intervention group on the degrees of flexion

1 Appendix: R code

1.1 Report 1

1.1.1 Normal distribution: Methods of moments estimates, Histogram and Quantile-quantile plot

```
library(tidyverse)
finance
a=mean(finance$logret)
b=var(finance$logret)
ggplot(finance) +
  geom_histogram(aes(x = logret, y = ..density..),
                 breaks = seq(-0.06, 0.06, 0.006)) +
  stat_function(geom = "line", fun = dnorm,
                args = list(mean = a, sd = sqrt(b)), lwd
                    = 1) +
  lims(x = c(-0.06, 0.06)) +
  labs(x = "Log␣return", y = "Density")

ggplot(finance, aes(sample = logret)) +
  stat_qq(distribution = qnorm) +
  stat_qq_line(distribution = qnorm) +
  labs(x = "Normal␣quantiles", y = "Sample␣quantiles")
```

1.1.2 Student T distribution: Methods of moments estimates, Histogram and Quantile-quantile plot

```
alpha = mean(finance$logret)
(6-4*kurtosis(finance$logret))/(3-kurtosis(finance$logret
   ))
gamma = (6-4*kurtosis(finance$logret))/(3-kurtosis(
   finance$logret))
c = (var(finance$logret)*(gamma-2))/(gamma)
beta = sqrt(c)

ggplot(finance) +
  geom_histogram(aes(x = logret, y = ..density..),
                 breaks = seq(-0.06, 0.06, 0.006)) +
  stat_function(geom = "line", fun = dstudent,
                args = list(a=alpha, b=beta, g=gamma),
                    lwd = 1) +
  lims(x = c(-0.06, 0.06)) +
```

```
  labs(x = "Log␣return", y = "Density")

ggplot(finance, aes(sample = logret)) +
  stat_qq(distribution = qstudent, dparams = list(a=alpha
     , b=beta, g=gamma)) +
  stat_qq_line(distribution = qstudent, dparams = list(a=
     alpha, b=beta, g=gamma)) +
  labs(x = "Normal␣quantiles", y = "Sample␣quantiles")
```

1.1.3 Value at risk and Probability that the price Xt reduces by at least 3% of its value in one day

```
qstudent(0.01, alpha, beta, gamma)
pstudent(log(0.97), alpha, beta, gamma)
```

1.2 Report 2

1.2.1 Mean and standard deviation of differences in degrees of flexion for control and intervention group, confidence interval for both groups

```
arthritis
arthritis %>% group_by(Group) %>% summarise(mean(Before-
   After))
arthritis %>% group_by(Group) %>% summarise(sd(Before-
   After))
-3.38+1.9600*sqrt((5.31^2)/(24))
-3.38-1.9600*sqrt((5.31^2)/(24))
-6-1.9600*sqrt((6.41^2)/52)
-6+1.9600*sqrt((6.41^2)/52)
```