Practical 1

Melvil Deleage, Jeff Macaraeg

March 31, 2025

Table of contents

Part 1 - Financial Returns & Normality

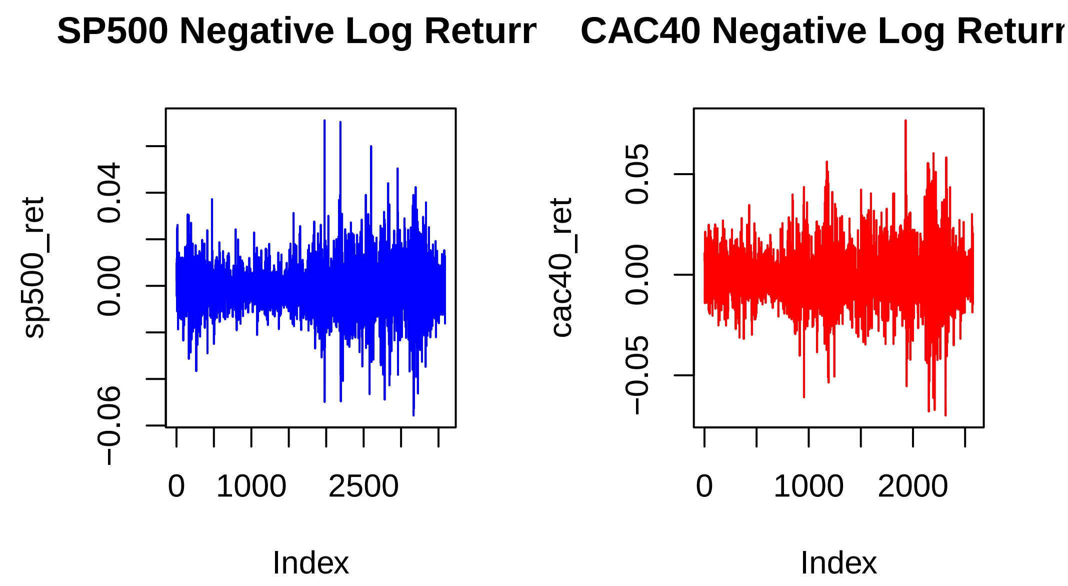
Question a)

> Read in the data. Then, assess the stationarity of the (raw) stock indices.

As a result, we can see that for SP500, the p-value is 0.9044 and for CAC40, the p-value is 0.9676. As they are both higher than 0.05, we can conclude that both series are not stationary.

Question b)

> Create a function to transform the daily stock indices into their daily negative log returns counterparts. Plot the latter series and assess their stationarity. To compare the series, also plot the negative log returns on a common scale to all indices.

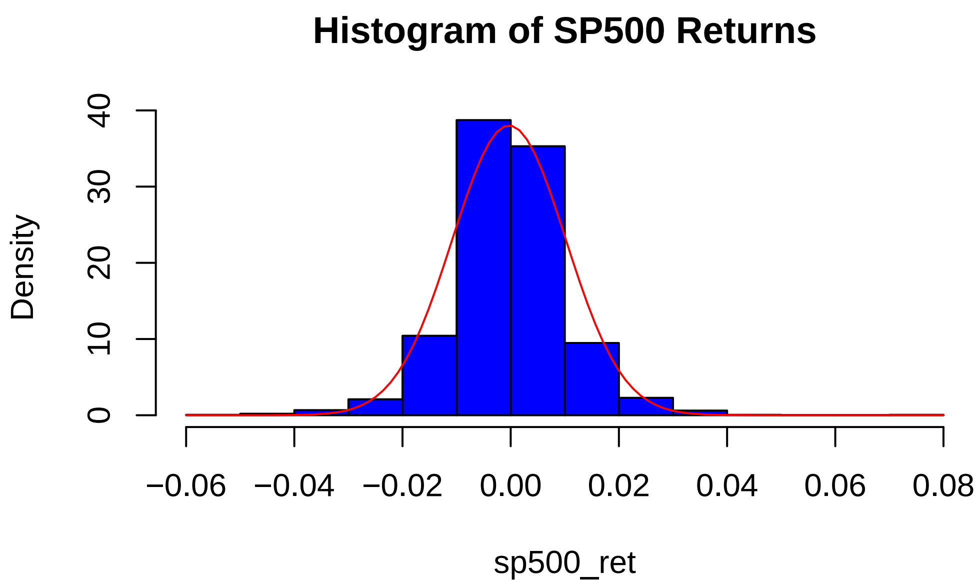


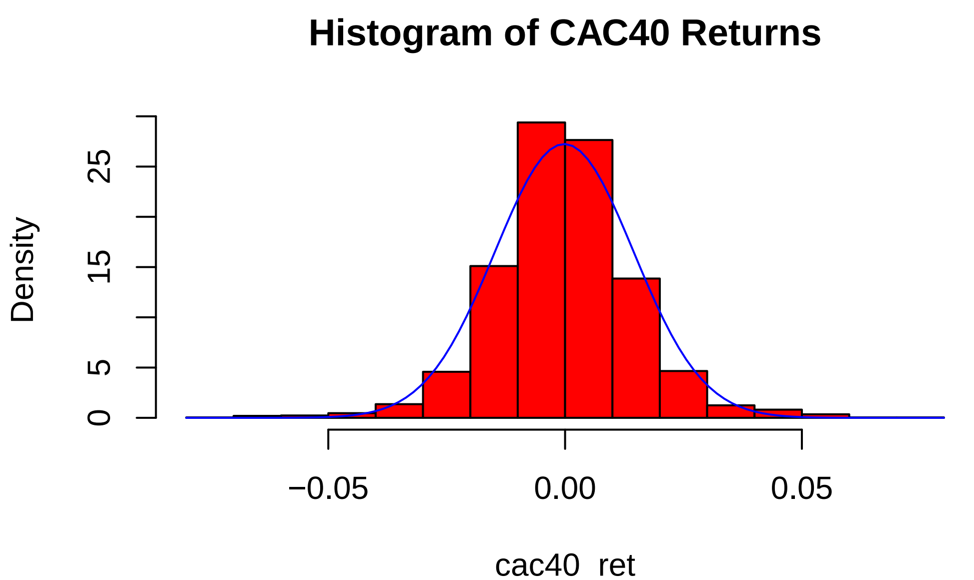

```
[1] "ADF Test for SP500:"

	Augmented Dickey-Fuller Test

data:  sp500_ret
Dickey-Fuller = -14.843, Lag order = 15, p-value = 0.01
alternative hypothesis: stationary

[1] "ADF Test for CAC40:"

	Augmented Dickey-Fuller Test

data:  cac40_ret
Dickey-Fuller = -13.34, Lag order = 13, p-value = 0.01
alternative hypothesis: stationary
```

As both plots seem to not show trends, periodic cycles and a stable variance, the series seem to be stationary. To verify this, the ADF test shows that both p-values (p-value = 0.01) are lower than 0.05, so we reject the hypothesis and thus, both series are stationary.

Question c)

Draw histograms of the negative log returns and compare them to the Normal distribution. What do you observe?

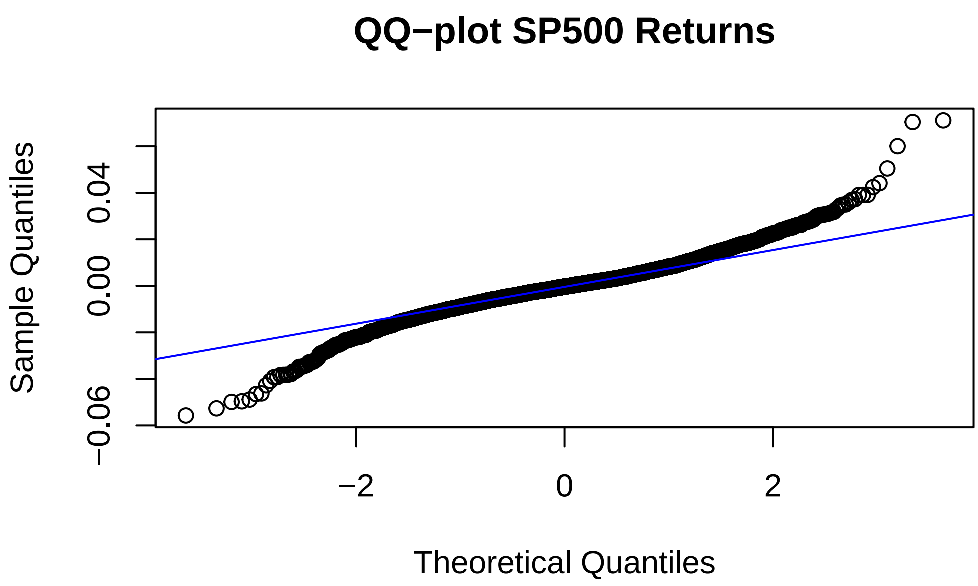


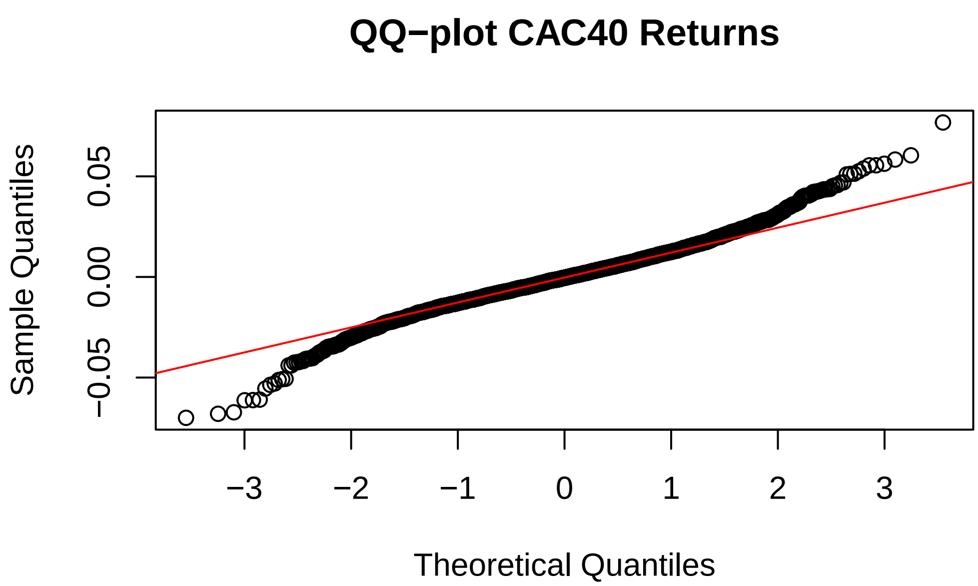


Both histograms have bell-shaped distributions, but are not perfectly aligned with the normal curve. ALso, the tails seeem to go further than what the normal disctribution curve predicts, which can indicate a higher probability of extreme returns than expected in a normal model.

Question d)

Check the normality assumption of the negative log returns using QQ-plots. What is your conclusion?





On both the lower and upper extremes, the points deviate from the line, showing heavier tails than a normal distribution. So, the negative log returns don't seem to follow normal distribution.

Question e)

> Formally test the normality assumption of the negative log returns using an Anderson-Darling testing procedure. Do you reject the Normal hypothesis?

```
    Anderson-Darling normality test

data:  sp500_ret
A = 29.24, p-value < 2.2e-16
```

```
    Anderson-Darling normality test

data:  cac40_ret
A = 10.33, p-value < 2.2e-16
```

As both p-values are lower than 0.05 (both at 2.2e-16), we reject the null hypothesis, meaning that the returns are not normally distributed.

Question f)

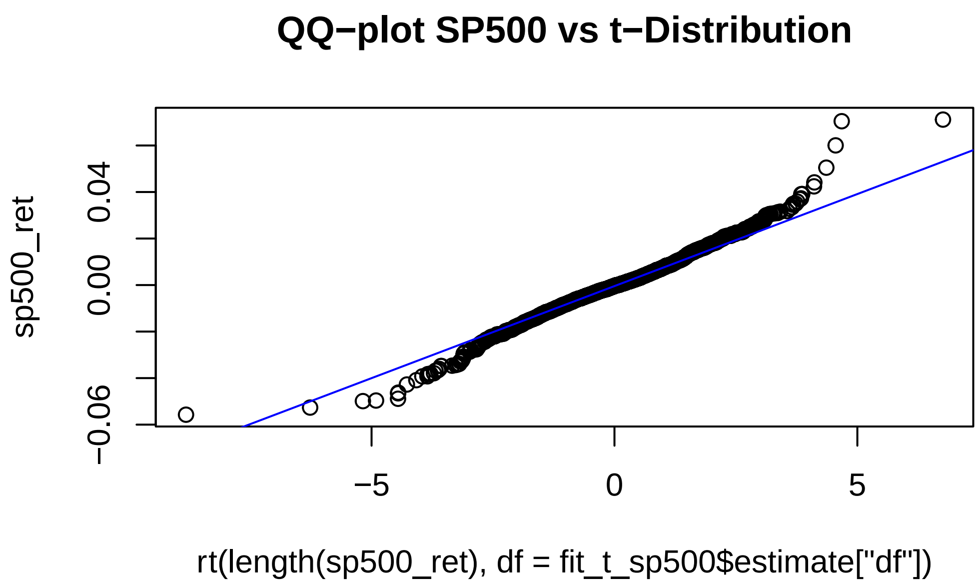
> Use the fitdistr() function from the MASS package in order to obtain the (maximum-likelihood estimated) parameters of distributions you could imagine for the negative log returns. Try to fit at least two different distributions on the data and, using an information criteria (such as the AIC), decide which distributional framework fits best for each of the series.

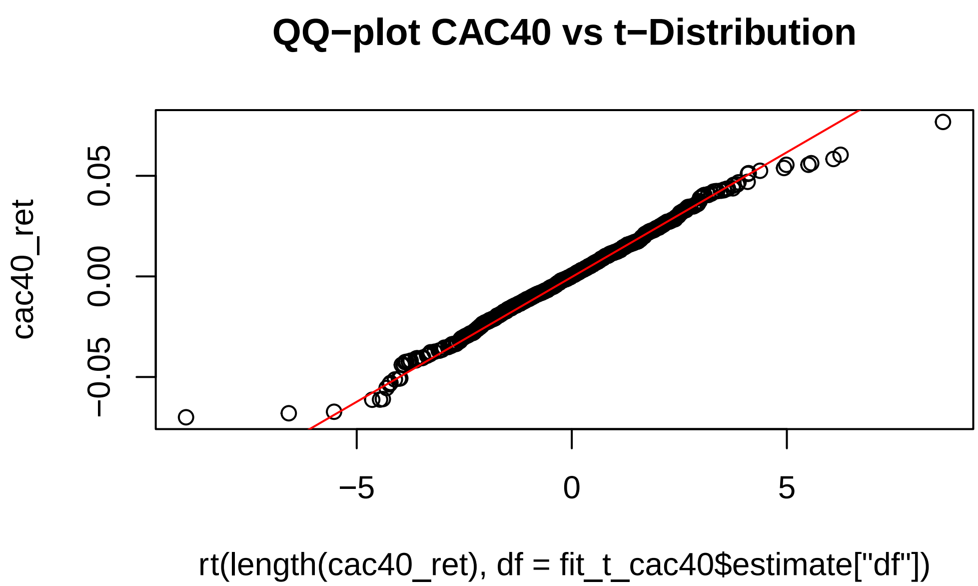
```
       mean            sd
  -0.0003139511    0.0104904994
 ( 0.0001751582) ( 0.0001238556)

        m               s               df
  -0.0003823438    0.0083396816    6.5020438529
 ( 0.0001642079) ( 0.0011490812) ( 5.6443703001)

       mean            sd
  -0.0001723074    0.0146403328
 ( 0.0002884550) ( 0.0002039685)

        m               s               df
  -0.0003374078    0.0116334321    5.2499036218
 ( 0.0002632892) ( 0.0002745656) ( 0.5356059623)

[1] -22510.5

[1] -22898.7

[1] -14447.55

[1] -14626
```

As the Student model yields lower AIC than the normal model, it indicates a better balance between the goodness-of-fit and the complexity. It is a more appropriate choice of model.

Question g)

If this has not been done in (f), fit a t-distribution to the negative log returns using fitdistr(). Using a QQ-plot for each of the series, decide whether the fit is better than with a Normal distribution, based on your answer in (d).



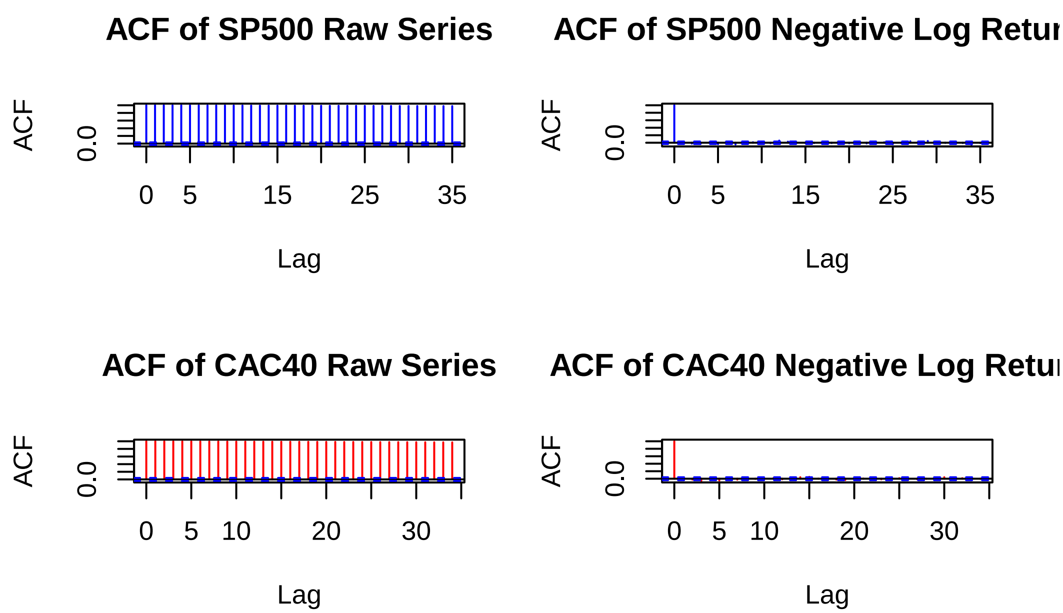


Despite some deviations from points in the extremes, the QQ-plots confirm that a Student's t-distribution provides a better fit for both the SP500 and CAC40 log returns compared to a normal distribution. The heavier tails of the t-distribution captures the extremes better. It is also consistent with question d).

Part 2 - Financial time series, volatility and the random walk hypothesis

Question a)

Plot the ACF of all the series in Part 1 (i.e. the raw series as well as the negative log returns). What do you observe?



The raw series show high autocorrelation at all lags, implying non-stationarity. For the negative log returns, it shows little to no autocorrelation, suggesting that the daily returns are close to uncorrelated.

Question b)

Use a Ljung-Box procedure to formally test for (temporal) serial dependence in the series. What is your conclusion?

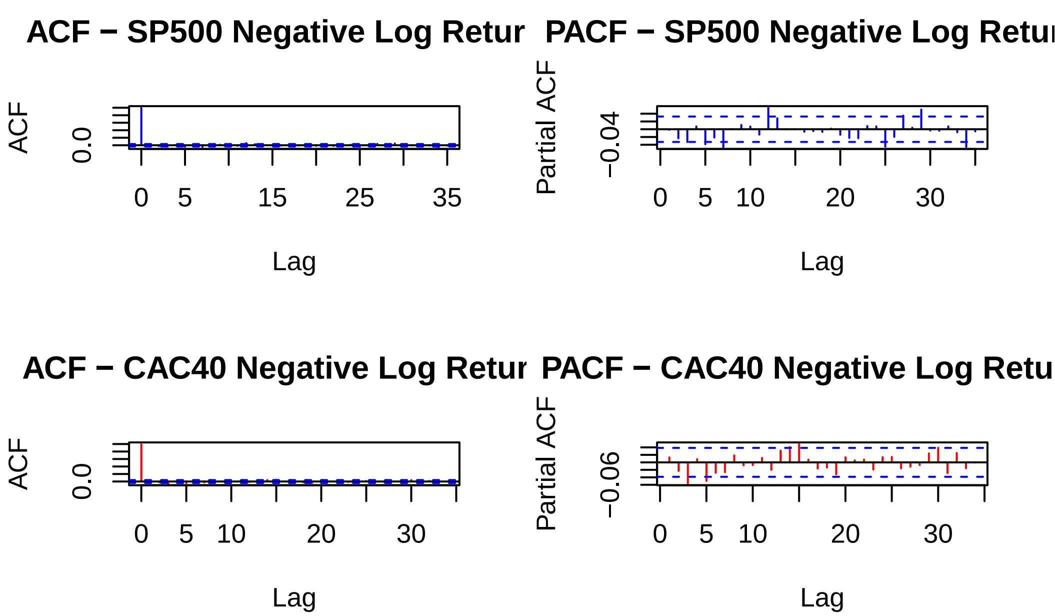
```
    Box-Ljung test

data:  sp500
X-squared = 71035, df = 20, p-value < 2.2e-16

    Box-Ljung test

data:  sp500_ret
X-squared = 38.931, df = 20, p-value = 0.0068

    Box-Ljung test

data:  cac40
X-squared = 50889, df = 20, p-value < 2.2e-16
```

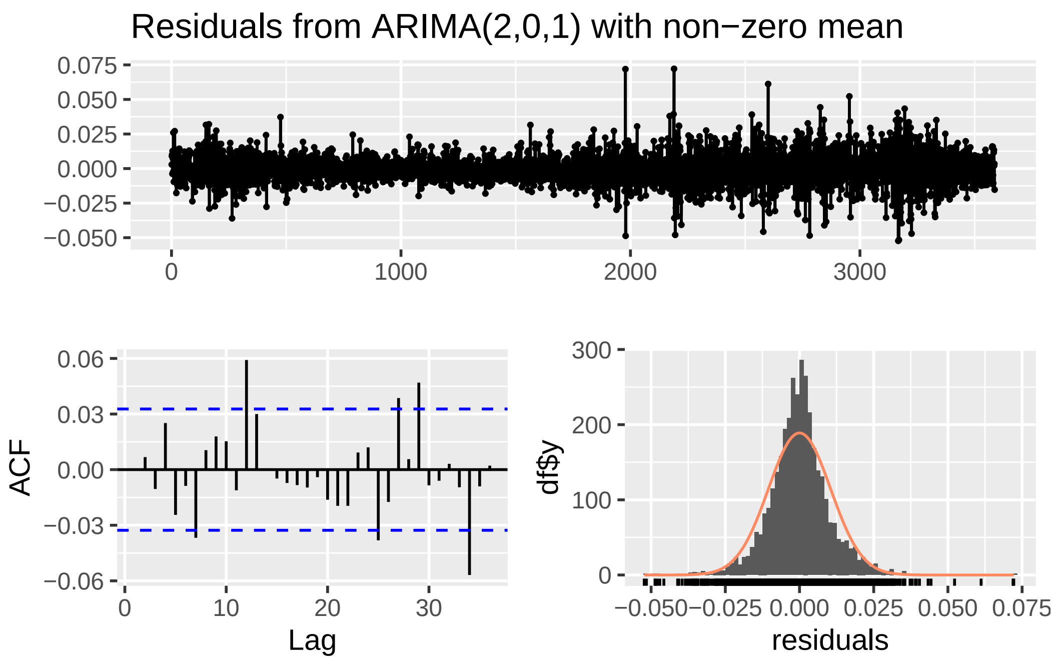
```
	Box-Ljung test

data:  cac40_ret
X-squared = 41.079, df = 20, p-value = 0.003639
```

Looking at the raw series for both CAC40 and SP500, we obtain a p-value that is lower than 0.05 (both at 2.2e-16), so we reject the null hypothesis and the raw series show autocorrelation. For both negative log returns, the p-values are higher than the raw series (0.0068 for SP500, 0.003639 for CAC40), but still smaller than the p-value, so again they show autocorrelation.

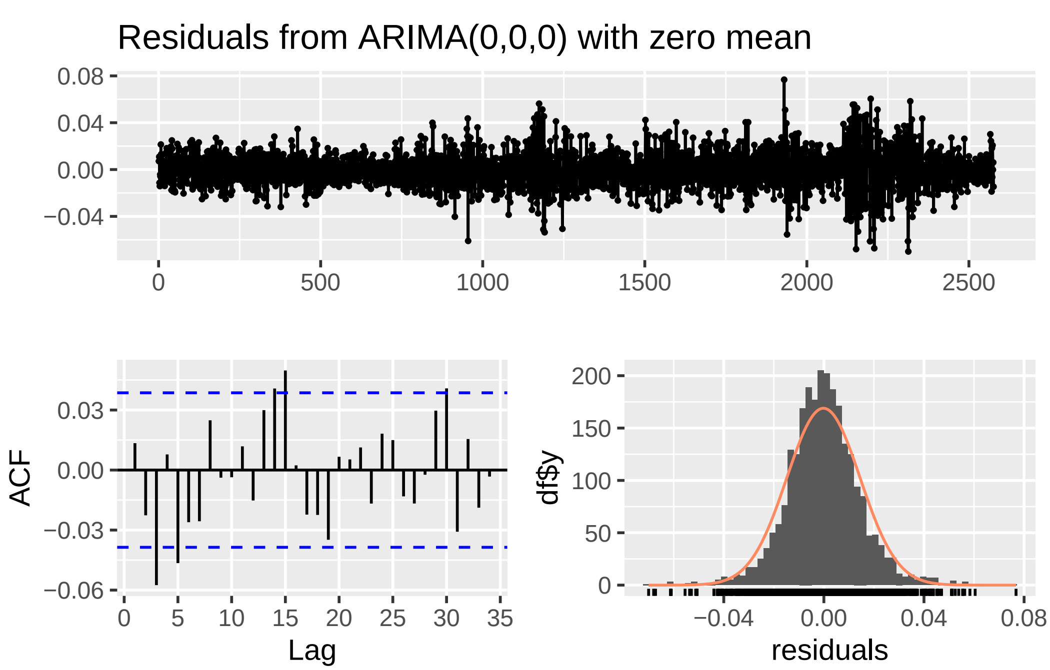
Question c)

> Propose ARIMA models for each of the negative log returns series, based on visualisation tools (e.g. ACF and PACF). Select an ARIMA model using auto.arima() (forecast package) to each of the negative log returns series. Comment on the difference. Assess the residuals of the resulting models.






```
    Ljung-Box test

data:  Residuals from ARIMA(2,0,1) with non-zero mean
Q* = 12.498, df = 7, p-value = 0.08534

Model df: 3.   Total lags used: 10
```



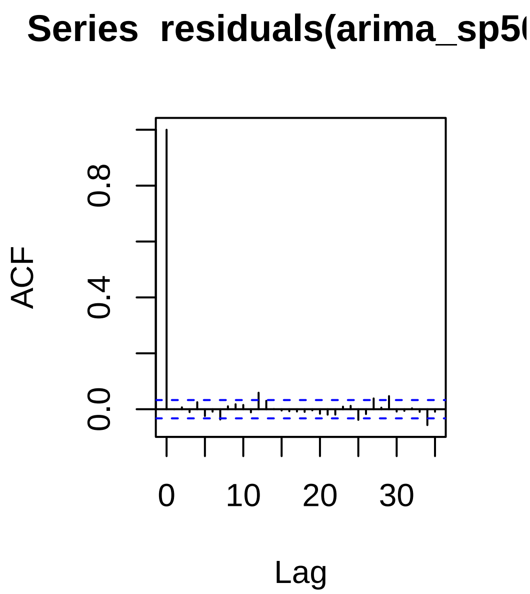

```
    Ljung-Box test

data:  Residuals from ARIMA(0,0,0) with zero mean
Q* = 21.191, df = 10, p-value = 0.0198
```

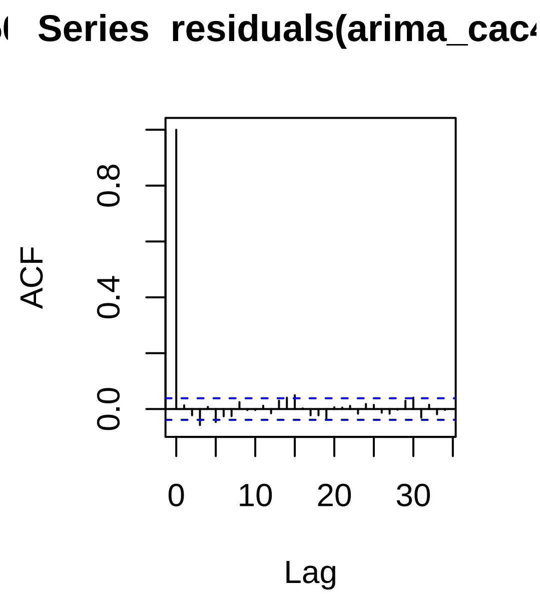
```
Model df: 0.   Total lags used: 10

Series: sp500_ret
ARIMA(2,0,1) with non-zero mean

Coefficients:
         ar1      ar2      ma1    mean
      0.7796  -0.0291  -0.7827  -3e-04
s.e.  0.1035   0.0180   0.1024   2e-04

sigma^2 = 0.0001099:  log likelihood = 11261.47
AIC=-22512.94   AICc=-22512.92   BIC=-22482.01

Training set error measures:
                       ME       RMSE         MAE  MPE MAPE      MASE
Training set 2.252009e-06 0.01047816 0.00748605 -Inf  Inf 0.6910122
                      ACF1
Training set -0.0003543668


Series: cac40_ret
ARIMA(0,0,0) with zero mean

sigma^2 = 0.0002144:  log likelihood = 7225.6
AIC=-14449.19   AICc=-14449.19   BIC=-14443.34

Training set error measures:
                        ME       RMSE         MAE MPE MAPE      MASE       ACF1
Training set -0.0001723074 0.01464135 0.01089581 100  100 0.7046982 0.01343285
```

The ACF & PACF for both SP500 & CAC40 negative log returns show almost no significant autocorrelation, indicating that the returns behave in a similar way to white noise.It is also confirmed by checking the residuals. We can also observe that the auto.arima() function selected very simple models.

Question d)

> Assess the residuals of the resulting models from (c), both their raw values and their absolute values, through visual tools (such as the ACF) and formal tests (e.g. Ljung-Box). What do you conclude about the independence assumption?

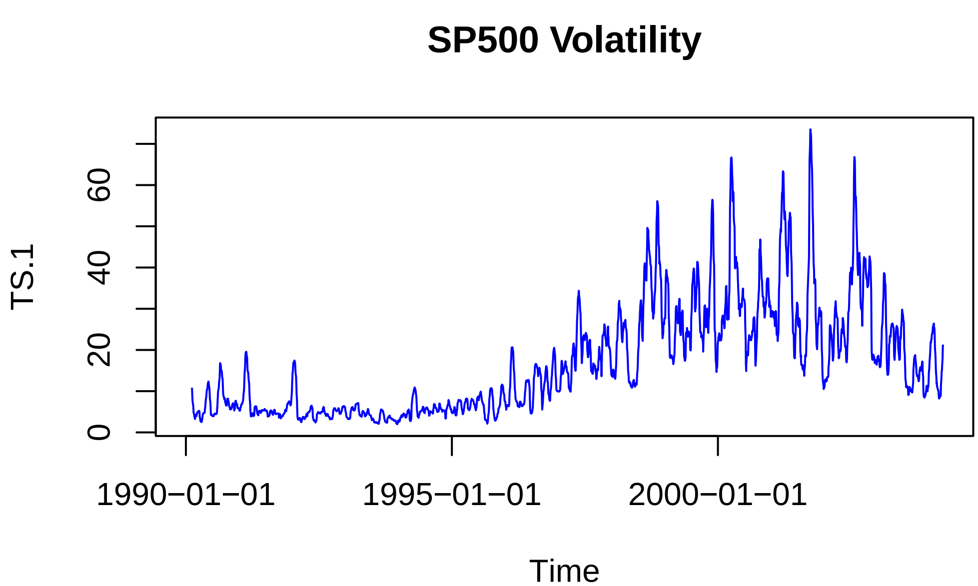


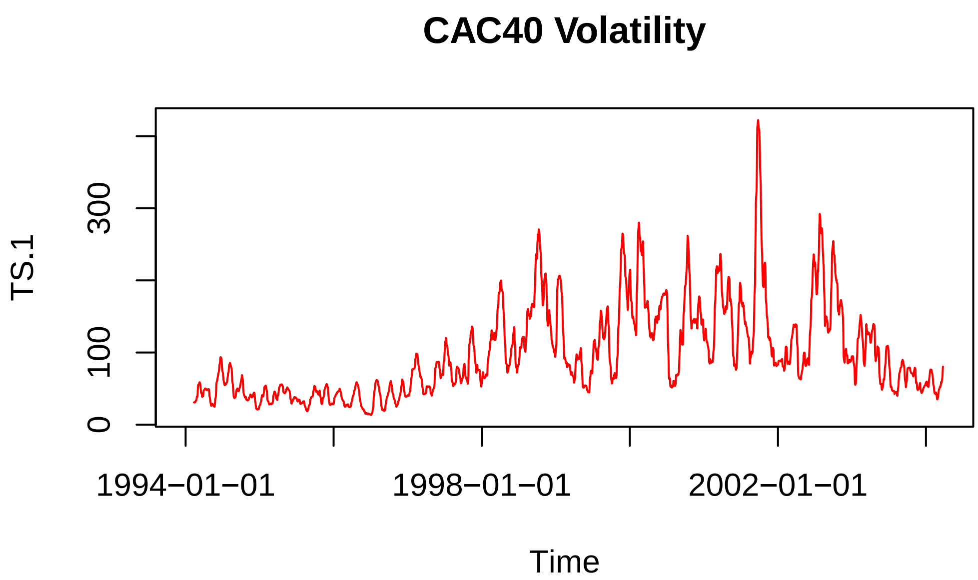


The independence assumption holds for the raw residuals but not for the volatility patterns (autocorrelation in the absolute residuals). So, further modeling could be used to fully capture the dynamics of the residuals.

Question e)

Plot the volatility of the raw series of indices. What is your conclusion on the homoscedasticity assumption?





The data shows heteroskedasticity, as the volatility (or variance) is not constant over time.

Question f)

> Residual serial correlation can be present when fitting a GARCH directly on the negative log returns. Hence, in order to circumvent this problem, it is possible to use the following two-step approach: • fit an ARMA(p,q) on the negative log returns; • fit a GARCH(1,1) on the residuals of the ARMA(p,q) fit. Proceed with the above recipe. Assess the quality of the above fit.

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 1), data = sp500_ret, trace = FALSE)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x1531267f0>
 [data = sp500_ret]

Conditional Distribution:
 norm

Coefficient(s):
         mu        omega       alpha1        beta1
-5.2127e-04   5.8543e-07   5.9959e-02   9.3606e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
```

```
mu     -5.213e-04   1.326e-04   -3.930 8.48e-05 ***
omega   5.854e-07   1.730e-07    3.384 0.000715 ***
alpha1  5.996e-02   7.873e-03    7.615 2.64e-14 ***
beta1   9.361e-01   8.194e-03  114.235  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 11726.55    normalized:  3.269179

Description:
 Mon Mar 31 18:26:16 2025 by user:


Standardised Residuals Tests:
                                  Statistic     p-Value
 Jarque-Bera Test   R    Chi^2  634.1311707 0.000000000
 Shapiro-Wilk Test  R    W        0.9846834 0.000000000
 Ljung-Box Test     R    Q(10)   17.5172908 0.063673028
 Ljung-Box Test     R    Q(15)   33.6290930 0.003837944
 Ljung-Box Test     R    Q(20)   35.4927658 0.017631499
 Ljung-Box Test     R^2  Q(10)    5.9096209 0.822798459
 Ljung-Box Test     R^2  Q(15)    9.2105927 0.866251163
 Ljung-Box Test     R^2  Q(20)   13.2884867 0.864666017
 LM Arch Test       R    TR^2     8.5347893 0.742067692

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-6.536128 -6.529231 -6.536131 -6.533670


Title:
 GARCH Modelling

Call:
 garchFit(formula = ~garch(1, 1), data = cac40_ret, trace = FALSE)

Mean and Variance Equation:
 data ~ garch(1, 1)
<environment: 0x1476a0ba8>
 [data = cac40_ret]

Conditional Distribution:
 norm

Coefficient(s):
         mu        omega       alpha1        beta1
-4.5271e-04   2.1959e-06   6.6767e-02   9.2304e-01

Std. Errors:
 based on Hessian

Error Analysis:
```

```
         Estimate  Std. Error  t value Pr(>|t|)
mu     -4.527e-04   2.356e-04   -1.922   0.0546 .
omega   2.196e-06   7.203e-07    3.049   0.0023 **
alpha1  6.677e-02   8.655e-03    7.714 1.22e-14 ***
beta1   9.230e-01   9.827e-03   93.926  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 7494.964    normalized:  2.909536

Description:
 Mon Mar 31 18:26:16 2025 by user:


Standardised Residuals Tests:
                                 Statistic      p-Value
 Jarque-Bera Test   R    Chi^2  35.7558904 1.720707e-08
 Shapiro-Wilk Test  R    W       0.9968568 3.631423e-05
 Ljung-Box Test     R    Q(10)  12.1728824 2.736528e-01
 Ljung-Box Test     R    Q(15)  20.6911526 1.469907e-01
 Ljung-Box Test     R    Q(20)  22.2871604 3.251340e-01
 Ljung-Box Test     R^2  Q(10)  11.9996781 2.850780e-01
 Ljung-Box Test     R^2  Q(15)  12.5832455 6.344530e-01
 Ljung-Box Test     R^2  Q(20)  14.2805083 8.159934e-01
 LM Arch Test       R    TR^2   13.0086585 3.684117e-01

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.815966 -5.806876 -5.815970 -5.812671
```

By applying a GARCH model on the ARMA model, the mean and volatility dynamics are well captured. The GARCH parameters are highly significant for both series. The final models for both SP500 and CAC40 effectively remove serial correlation from the residuals and capture the volatility clustering, making them appropriate for modeling on negative log returns.

Question g) FAUX

> Use the garchAuto.R script in order to fit a GARCH on the residuals of the ARMA(p,q) from (g). Assess the quality of the fit.

A FINIR