Empirical Finance

- Empirical Assignment -

Lasse Mejlholm Larsen
Maarten Moes
Thomas Schnack Poulsen

Aalborg University
Finance

AALBORG UNIVERSITY
STUDENT REPORT

Finance
Aalborg University Business School
https://www.business.aau.dk/

Title:
Empirical Finance

Theme:
Empirical Assignment

Project Period:
Spring Semester 2022

Project Group:
Fri-69465-13

Participant(s):
Lasse Mejlholm Larsen
Maarten Moes
Thomas Schnack Poulsen

Supervisor(s):
Douglas Eduardo Turatti

Copies: 1

Page Numbers: 26

Date of Completion:
November 15, 2022

Contents

Chapter 1

Predictability of Asset Returns

Question 1.1 What is the RW Hypothesis for asset prices? What is the random walk model?
The RW hypothesis for asset prices is,

$$E[P_{t+1}|P_t, P_{t-1} \dots] = P_t$$

The random walk hypothesis states that the best prediction of the price of an asset tomorrow is the price today considering the price history of the asset.

Question 1.2 It is usually stated that under the RW Hypothesis the best prediction for prices tomorrow are the prices of today. Show this result using the random walk model
The Random Walk Model is defined by,

$$P_t = P_{t-1} + \epsilon_t$$

This means the price changes is given by,

$$\triangle P_t = P_t - P_{t-1} = \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2)$$

Because $\triangle P_t = \epsilon_t$ where ϵ_t is a white noise process, which is purely random with no correlation, then the Random Walk hypothesis implies that the complete path of an asset's prices is random. Therefore the prices doesn't revert to anything - not reverting to any fundamental value and no mean reversion.
Since all prices are purely random, the price changes will vary, but they will fluctuate around 0, which is the sum of random shocks or returns. This means the best forecast of an asset's price tomorrow is the asset's price today.

Question 1.3 What is the expected value for asset returns under the RW Hypothesis? Explain.
Because the price change has an expected value of 0, so does the returns. Thereby returns have a zero mean.

$$E[(P_{t+1} - P_t)/P_t] \approx E[ln(P_{t+1}/P_t)] = 0$$

Question 1.4 Under the RW hypothesis asset prices are said to have no long-term value. Can you explain this sentence?
According to the Random Walk prices can be written as,

$$P_t = P_0 + \sum_{t=1}^{T} \epsilon_t$$

This value cannot converge to any quantity, because they are a sum of multiple white noise shocks with weight equal to 1. Therefore the price will increase or decrease randomly.

Let r_t be some asset log-returns, and they are known to follow the model below,

$$r_t = 0.002 + 0.55r_{t-1} + 0.35r_{t-2} + \epsilon_t, \quad \epsilon_t \sim N(0, 0.0002)$$

it is known that the log returns were
-0.0313 **on 03/05**
0.0131 **on 02/05**
0.0032 **on 01/05**
-0.0151 **on 30/04**
The next three questions refer to this model.

Question 1.5 On this basis provide the best forecast of the daily log-returns for this stock on 04/05
The model above suggests that asset returns are described by an AR(2) or Autoregressive of order 2,

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} \quad \epsilon \sim WN(0, \sigma^2)$$

$$r_t = 0.002 + 0.55r_{t-1} + 0.35r_{t-2} + \epsilon_t, \quad \epsilon_t \sim N(0, 0.0002)$$

The best 1-step ahead forecast is the conditional expectation of $\hat{r}_{t+1}$,

$$\hat{r}_{04/05} = E_t[r_{04/05}] = \phi_0 + \phi_1 r_{03/05} + \phi_2 r_{02/05}$$

$$\hat{r}_{04/05} = 0.002 + 0.55 \cdot -0.0313 + 0.35 \cdot 0.0131 = -0.01063$$

Question 1.6 Following the information on 1.5, provide the confidence interval of the forecast.
The forecasting error associated to the estimated log-return on 04/05 is

$$\epsilon_{04/05} = r_{04/05} - \hat{r}_{04/05}$$

Since the forecasting error is just the error term, then the conditional variance is:

$$var_t[r_{t+1}] = \sigma^2$$

Assuming normality, the 95% confidence interval of the forecast is described by,

$$[\phi_0 + \phi_1 r_t + \phi_2 r_{t-1} - 1.96\sigma; \phi_0 + \phi_1 r_t + \phi_2 r_{t-1} + 1.96\sigma]$$

Question 1.7 What is the best trading position for 04/05 on the basis of your answer in 1.5?
The best trading position for 04/05 is to be short, since the expected return of the asset on date 04/05 is negative.

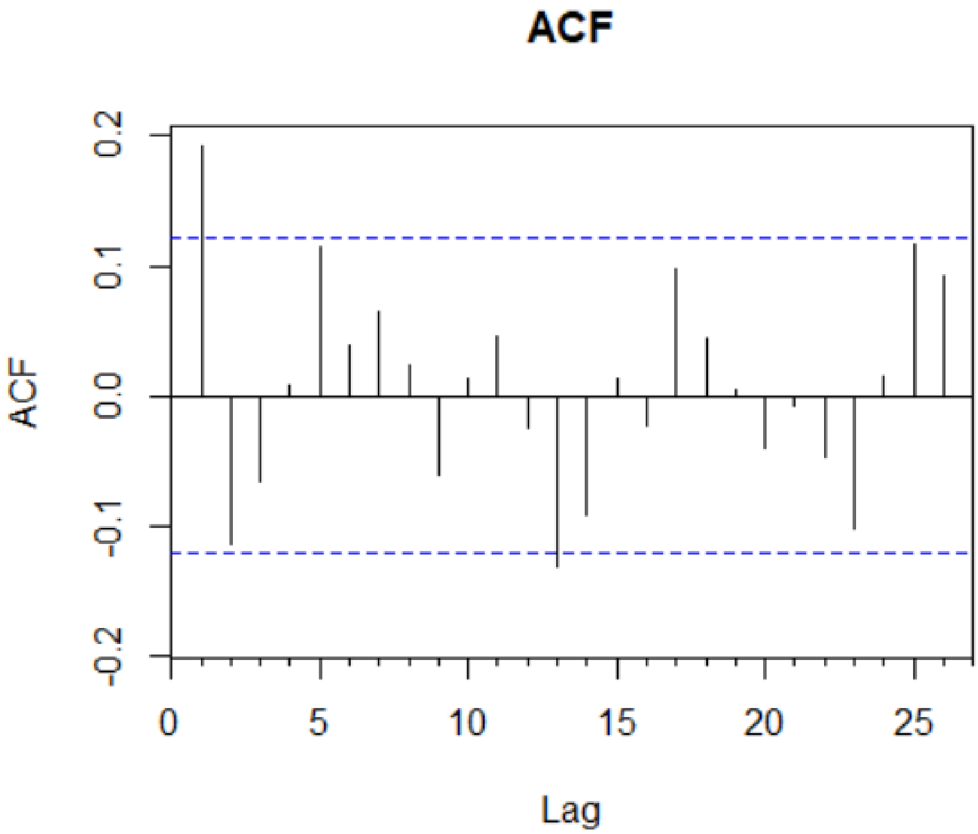
The file exchange.xls contains 262 observations of weekly log-returns for the dollar-euro exchange rate from 02/01/2015 to 03/01/2020. The following questions refer to this dataset.

Question 1.8 Is this data a white noise process? How can you test this? What are the implications if the log-returns for the dollar-euro exchange rate are a white noise process?
It can be tested if this data is a white noise process using the Ljung-Box test. If the log-returns for the dollar-euro exchange rate are a white noise process, then we can't forecast future returns based on historical returns, because there are not enough evidence to confirm a potential autocorrelation.

Question 1.9 What is the autocorrelation function? Plot the sample autocorrelation function for this data set. What is the interpretation of this sample ACF?
An autocorrelation function explains whether time series data points, on average, are related to the prior data points. According to the plot of the sample autocorrelation function below, its illustration indicates a MA(1), since the rule of thumb, is that the ACF plot helps identifying MA terms and since there is one spike with a decay afterwards.
The following code in r was used to calculate the ACF,

```
Acf(exchange,lag = 26, main="ACF")
```

Figure 1.1: ACF



Question 1.10 Test the autocorrelations jointly by using the Ljung-Box test. Analyze the results.

The following code in r, was used to calculate the Ljung-Box test,

```
Box.test(exchange, lag = 26, type = c("Ljung-Box"))
```

It can be tested by the Ljung-Box Test whether the dollar-euro exchange rate is a white noise process or not. With a lag of 26, the test provided a p-value of 0.0224, which means the evidence points towards a rejection of the null, which might indicate that autocorrelation occurs. When autocorrelation occurs, future observations can be forecasted by the historical observations, whereby the dollar-euro exchange rate might be autocorrelated.

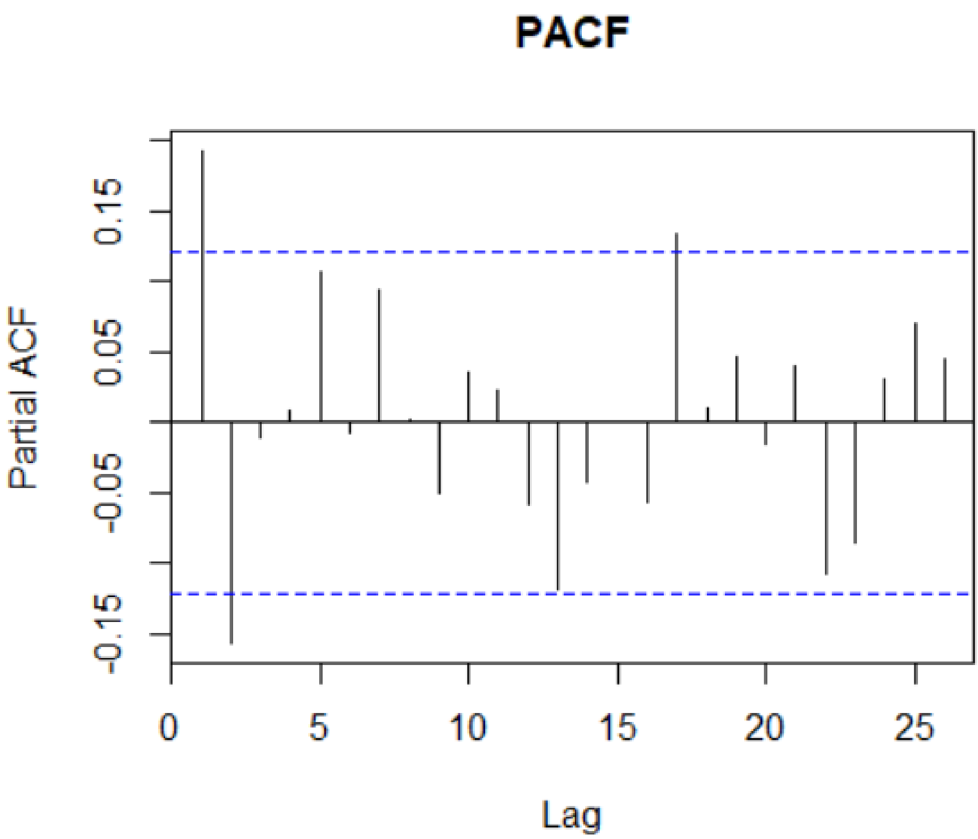
Question 1.11 What is the partial autocorrelation function? Plot the sample partial autocorrelation function for this dataset. What is the interpretation of this sample PACF?

This code in r was used to calculate the PACF,

```
pacf (exchange, lag = 26, main="PACF")
```

The partial autocorrelation function explains whether time series data points, on average, are related to the prior data points with correlations of intervening data points removed. The illustration of the PACF indicates an AR(2) model, since the rule of thumb for PACF, is that the PACF helps identifying AR terms. Furthermore there is two spikes with a decay afterwards.

Figure 1.2: PACF



Question 1.12 Estimate a time series model according to these ACF and PACF. Justify your choice. Analyze results.

The ACF has one spike with a decay afterwords indicating a MA(1) while the PACF indicates an AR(2), which hints an ARMA(2,1) model.

```
# estimate an ARMA(2,1)
model2 <- arima(exchange, order=c(2,0,1))
```

```
print(model2)
```

by using this command in R, it computes the estimated time series model with regards to the data chosen.

$$r_t = -0.0273 + 0.2839r_{t-1} + -0.1696r_{t-2} + \epsilon_t + -0.0638_1\epsilon_{t-1}$$

Question 1.13 How can we test if the model proposed in 1.12 is able to fully capture the autocorrelation in the data? Analyze the results.
By using the Ljung -Box test, we tested if the residuals were autocorrelated on the ARMA(2,1) model,

```
Box.test(model2$residuals, lag = 26, type = c("Ljung-Box"))
```

The test computed a p-value of 0.6121, which means that we do not reject the null hypothesis of the Ljung-Box test, which indicates that the residuals are not autocorrelated. This is the outcome supports our model, Since the residuals are independent, which indicates that the model does not show lack of fit.

Question 1.14 Perform an expanding window forecasting study between your model proposed in 1.12, and the ARMA(1,1). Use the first 100 observations for the estimation window. Use only 1-step ahead forecasts. Compute the RMSE. Which model has the best predictive power?
We have used the following code to compute the RMSE of both models.

```
# Forecasting study:
# arma(2,1)
err_arma21 <- expforecast (exchange, 2, 1, 100, 1)
# 2= order ar, 1= order ma, 100= end of estimation window,
1=step-ahead
# calculate rmse with the vector of forecasting errors
rmse_arma21 <- sqrt(mean(err_arma21^2))
# arma(1,1)
err_arma11 <- expforecast (exchange, 1, 1, 100, 1)
rmse_arma11 <- sqrt(mean(err_arma11^2))
#approximate white noise (only intercept, no ar/ma term)
err_wn <- expforecast (exchange, 0, 0, 100, 1)
rmse_wn <- sqrt(mean(err_wn^2))
print(rmse_arma21)
print(rmse_arma11)
print(rmse_wn)
```

The R-code provided the following RMSE values:

$$\text{ARMA}(2,1) = 0.714409$$

$$\text{ARMA}(1,1) = 0.7044253$$

$$\text{WN} = 0.7182221$$

The results show that the ARMA(1,1) has the best predictive power, because the data points are more concentrated around the line of best fit.

Chapter 2

Volatility Modeling

Question 2.1 What is volatility clustering? Can you provide some financial reasoning for this phenomenon?
Volatility clustering is a term for how volatility tends to cluster together. It should be understood as a concept that high volatility tends to be followed by high volatility and low volatility tends to be followed by low volatility These changes can be of both signs. Furthermore periods with high volatility are followed by periods with low volatility.
There can be several financial reasons for this stylized fact in the financial marks. One is the situation where neither the fundamental or technical (chartist) dominates the market. Furthermore it can occur when traders switch more often between those two strategies.[2]

Question 2.2 What are leverage effects? Can you provide some financial reasoning for this phenomenon?
Leverage effects can be understood as the phenomenon of which volatility tends to react differently to a large price increase than a large price drop. A financial reason for this could be found within the theory of behavioral finance. The term loss aversion is the theory, that investors perceive loss differently to gains and are more adverse to recognize losses than gains. .[1]

Question 2.3 Write the following models (mean, variance and distribution assumption),

ARMA(1,1)-GARCH(2,1) with in mean effects.

$$r_t = \phi_0 + \phi_1 r_{t-1} + \mathrm{c}\,\sigma_t^2 + \epsilon_t + \theta_1 \epsilon_{t-1}$$
$$\epsilon_t = v_t \sqrt{\sigma_t^2}, \quad v_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \beta_1\sigma_{t-1}^2$$

ARMA(3,2)-GARCH(1,1)-t with zero-mean.

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$$
$$\epsilon_t = v_t\sqrt{\sigma_t^2}, \quad v_t \sim t(v)$$
$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

Question 2.4 What are the differences between conditional and unconditional volatility in the context of ARCH/GARCH models?

Conditional volatility is the volatility of a random variable given some extra information, while unconditional volatility is the general volatility of a random variable without some extra information.

The volatility at time t is the conditional volatility. It is conditioned on the past values of itself and it is equal to,

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2$$

The unconditional variance/volatilty is constant and also defines the long-term variance and is calculated as the mean of conditional variances.

$$\text{Unconditional volatility for ARCH} = E[\sigma_t^2] = \overline{\sigma}^2 = \frac{\alpha_0}{1-\alpha_1}$$

$$\text{Unconditional volatility for GARCH} = E[\sigma_t^2] = \overline{\sigma}^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$$

Question 2.5 What is the unconditional volatility defined by this GARCH? (Consider volatility as the standard deviation).

Consider the estimated GARCH model for some asset log-returns,

$$r_t = \epsilon_t, \epsilon_t \sim N(0, \sigma_t^2),$$

$$\epsilon_t = u_t\sqrt{\sigma_t^2}, u_t \sim N(0,1),$$
$$\sigma_t^2 = 0.002 + 0.07 r_{t1}^2 + 0.84\sigma_{t1}^2$$

To calculate this we use the following formula:

$$\overline{\sigma}^2 = \frac{\alpha_0}{1-\alpha_1-\beta_1}$$

$$\frac{0.002}{1-0.07-0.84} = 0.022222$$

$$\sqrt{0.022222} = 0.14907$$

The unconditional/long-run volatility is 14.9%

Question 2.6 Now, consider that $r_t = -0.01$ and $\sigma_t^2 = 0.02$. Will the market at time $t+1$ be tranquil or turbulent when compared to the long-run? Explain

$$\sigma^2{}_{t+1} = 0.002 + 0.07(-0.01^2) + 0.84(0.02) = 0.018807$$
$$\sqrt{0.018807} = 0.137139$$

The expected level of variance for the time t+1 is 1.88% meaning that the expected level of volatility is 13.71% This implies that the market is more tranquil at time t+1 compared to the long-run.

Question 2.7 Calculate the Value-at-Risk for time $t+1$ if you have a financial position of W in this asset and with 95% probability level.
To calculate this, the following formula will be applied:

$$\text{VaR} = W(\mu + z_{1-\alpha}\sigma_{t+1})$$

$$\text{VaR} = W(0 + 1.6448 \cdot \sqrt{0.018807}) = 0.225566W$$

The result of this means that there is a 95% probability for losses smaller than 22.56% or 5% probability for losses greater than 22.56% of the investment.

The file exchange.xls contains 262 observations of weekly log-returns for the dollar-euro exchange rate from 02/01/2015 to 03/01/2020. The following questions refer to this dataset.

Question 2.8 Is there autocorrelation in the squared log-returns? Apply the LB test.

```
Box.test(my_ts^2, lag = 26, type = c("Ljung-Box"))
```

The Ljung-Box test resulted in a p-value < 0.05 and thereby a rejection of the null, therefore there is enough evidence to say that there is autocorrelation in the squared log-returns.

Question 2.9 Apply the ARCH-LM Test on the log-returns. Explain results.

```
test <- ArchTest (my_ts, lags=12, demean = FALSE)
```

The Arch-LM test computes a p-value below 0.05, which rejects the null hypothesis, which means there is volatility clustering , and thereby the conditional variance is time-varying.

Question 2.10 Estimate an ARMA(1,1)-GARCH(1,1) with Gaussian distribution and zero mean. Re-estimate with student-t. Are both models able to fully capture the volatility clustering effects present on the data? Explain results.

For the estimation of the models, the following code was used in r,

```
#Estimate model one with ARMA (1,1)-GARCH(1,1)
With gaussian distribution
and zero mean.
spec <- ugarchspec(variance.model = list(model= "sGARCH",
garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(1, 1), include.mean = FALSE ),
distribution.model = "norm")
model1 <- ugarchfit(spec, my_ts)
print(model1)
#Estimate model two with ARMA (1,1)-GARCH(1,1)
With Student-t distribution
and zero mean.
spec2 <- ugarchspec(variance.model = list(model = "sGARCH",
garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(1, 1), include.mean = FALSE ),
distribution.model = "std")
model2 <- ugarchfit(spec2, my_ts)
print(model2)
```

Gaussian:	P-values	P-values	P-values
Optimal parameters	Lag 1	Lag 5	Lag 9
Weighted LB-test std. residuals	0.7762	1.0000	0.9881
Weighted LB-test std. squared residuals	0.6527	0.8886	0.7961
Optimal parameters	Lag 3	Lag 5	Lag 7
Weighted ARCH LM Test	0.9800	0.6316	0.6303

Student-t:	P-values	P-values	P-values
Optimal parameters	Lag 1	Lag 5	Lag 9
Weighted LB-test std. residuals	0.7752	1.0000	0.9877
Weighted LB-test std. squared residuals	0.6402	0.8855	0.7916
Optimal parameters	Lag 3	Lag 5	Lag 7
Weighted ARCH LM Test	0.9760	0.6360	0.6281

All p-values regarding both the Ljung-Box and ARCH LM test are non-significant. This means there is no sign of either autocorrelation, which could indicate volatility clustering, or ARCH effects. Therefore, it is concluded that both models are able to fully capture the volatility clustering effects.

Question 2.11 Compare the in-sample fit of both models. Which model has the best overall fit? Can you relate this result with the estimated parameters from results in 2.10?
The following code was used to calculate the information criteria values for the two models,

```
infocriteria(model1)
infocriteria(model2)
```

Distribution:	Gaussian	Student-t
Information Criteria		
Akaike	2.530430	2.537574
Baysian	2.598529	2.619291
Shibata	2.529720	2.536556
Hannan-Quinn	2.557801	2.570418

The model with the overall best fit is the ARMA(1,1)-GARCH(1,1) with Gaussian distribution and zero mean. This is justified as this models has the lowest values regarding the four information criteria, which means it has the least loss of information.
This result agrees with the fact that most of the estimated parameters regarding ARMA(1,1)-GARCH(1,1) with Gaussian distribution and zero mean in 2.10, has the highest p-values, which might indicate that it has a better fit.

Question 2.12 Giving your answers in 2.10 and 2.11, calculate the Value-at-Risk for the best model for one week ahead if you have a financial position of W in the dollar-euro exchange rate and with 95% probability level. The following code was used in r,

```
# Estimate VaR
# level of significance of var
p <-0.05
#Forecast our volatility(std) only 1 step ahead,
since the data is weekly.
fore <-ugarchforecast(model1,n.ahead = 1)
#This command gives us the the forecasted volatility t+1
fore@forecast$sigmaFor
```

```
# get the (1-p) quantile from standard normal distribution
quantile_n <- qnorm((1-p), mean = 0, sd = 1)
# var estimation
var_est <- 0 + quantile_n*0.5104809
print(var_est)
```

$$\text{VaR} = W(\mu + z_{1-\alpha}\sigma_{t+1})$$

$$\text{VaR} = W(0 + 1.6448 \cdot 0.5104809) = 0.839664W$$

The VaR for the best model for one week ahead is 83.97%, which means there is a 5% probability of losing more than 83.97% of the investment and a 95% probability of losing less.

Chapter 3

Factor Models

Consider the general 1-factor model for k assets

$$r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_i^2)$$

for all $i = 1, \ldots k$ and $t = 1, \ldots T$.
The next three questions refer to the model above.

Question 3.1 State the assumptions so the above model is truly a factor model.
For this single factor model, $r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}, \epsilon_{i,t} \sim N(0, \sigma_i^2)$, f_t is a known and observed factor and we have to estimate α_i, β_i and ϵ_i^2 for each asset.
We assume that there are k assets and T time periods. Let $r_{i,t}$ be the excess return of asset i in the time period t, the excess return is calculated by $\text{return}_{i,t}$ - riskfree_t.
α_i is a constant that represents the intercept f_t for $i = 1, \ldots$ and β_i is the factor loading for asset i on the ith factor, and $\epsilon_{i,t}$ is the factor of asset i.
$\epsilon_{i,t}$ is the error term and in this case the idiosyncratic asset return that is only specific for asset i and time t.
For the common factors $f_{j,t}$ we assume,

$$E(f_{j,t} = \mu_{j,f}, \text{for } j = 1, \ldots, m$$

$$var(f_{j,t}) = \sigma_{j,f}^2, \text{for } j =, \ldots, m$$

$$cov(f_{j,t}, f_{h,t}) = \sigma_{j,h} \text{ for } j = 1, \ldots, m \text{ and } h = 1, \ldots, m$$

And the asset specific factor $\epsilon_{i,t}$ is a white noise series and uncorrelated with the common factors $f_{j,t}$ and other specific factors. We assume,

$$E(\epsilon_{i,t}) = 0 \text{ for all } i \text{ and } t$$

$$cov(f_{j,t}, \epsilon_{i,s}) = 0 \text{ for all } j, i, t \text{ and } s$$

$$cov(\epsilon_{i,t}, \epsilon_{j,s}) = \sigma_i^2 \text{ only if } i = j \text{ and } t = s$$
$$cov(\epsilon_{i,t}, \epsilon_{j,s}) = 0 \text{ if } i \neq j \text{ and } \ t \neq s$$

These assumptions imply that the common factors are uncorrelated with the specific factors, and that the specific factors are uncorrelated among each other.

Question 3.2 What is the variance of the asset i under the above general factor model?

Since the model above general 1-factor model only has one factor $m = 1$ and we are looking at asset i, in $r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}$. We calculate the variance by,

$$var(r_{i,t}) = var(\alpha_i + \beta_i f_t + \epsilon_{i,t})$$

We use the formula for variance that looks like, $var(ax + by) = a^2 var(x) + b^2 var(y) + a\ b\ cov(x, y)$, which leads to this formula, for the general 1-factor model

$$var(r_{i,t}) = \beta_i^2 \sigma_f^2 + \sigma_i^2$$

$\beta_i^2 \sigma_f^2$ is a proportion of the factor variance. β_i^2 converts the factor variance into an asset variance. All assets in the portfolio have some proportion due to the factor, thus, this term is the common variance. σ_i^2 is the idiosyncratic variance, which means that it is specifically the variance of asset i.

Question 3.3 What is the covariance between asset i and asset j under the above general factor model?

The covariance between asset i and asset j is,

$$\text{cov}[r_i, r_j] = \beta_i \beta_j \sigma_f^2$$

This implies that the assets are only related through their factors.

Consider the 1-factor model for excess returns of k assets in a portfolio,

$$r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}$$
$$\epsilon_{i,t} \sim W\ N(0, \epsilon_i^2) \text{ and } i = 1, \dots k$$

The next three questions refer to this model

Question 3.4 Suppose f_t represents long-term interest rates. How can you estimate α_i and β_i for all assets in the portfolio? Explain.

First step would be to gather all excess returns for k assets in the portfolio.

Thereafter use OLS to estimate α and β for each asset k.

Question 3.5 Now, suppose that f_t represents the effects of price-to-book ratio and β_i is the value of the price-to-book ratio for company i at time t. How can you estimate this model for all assets in the portfolio and over time? Explain.
The model that will be estimated is a fundamental factor model that use an asset specific attribute (price-to-book ratio), which is used to construct a common factor.

We use the Barra factor model and assume that excess returns and factor realizations are mean corrected and treat observable asset specific attributes, like price-to-book ratio, as factor betas, β_i. The common factors f_t are unobservable, but will be estimated. We would estimate $\hat{f}_t$ with OLS for every asset k at time t,

$$\hat{f}_t = cov(\beta_t, \tilde{r}_t)/var(\beta_t)$$

Where β_t gathers all betas, and $\tilde{r}_t$ all excess returns for time t.

The excel file "factordata.xlsx" contains time series representing two asset daily log-returns and a potential factor, a total of 500 observations. The following questions refer to this dataset.

Question 3.6 Estimate the following 1-factor model,

$$r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_i^2), \quad i = 1, 2$$

For both assets. Analyze results.

For the first asset, the following code in r was used:

```
ft <- my_data$factor
r1 <- my_data$asset1
r2 <- my_data$asset2
model1 <- lm(r1 ~ ft)
```

which resulted in the following output,

$$r_{1,t} = -0.97962 + 0.42333 f_t + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_i^2), \quad i = 1, 2$$

For the second asset, the following code in r was used, with the defined variables from above:

```
model2 <- lm(r2 ~ ft)
```

which resulted in the following output,

$$r_{2,t} = -1.10208 + 0.77814 f_t + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_i^2), \quad i = 1, 2$$

The α and β in both estimated models are significant. The characteristics of the models differ by the fact that the factor loading for asset one is 0.423033 compared to 0.77814 of asset two, whereby the factor has a higher impact for the return of asset two.

Question 3.7 Do the usual diagnostics on the regression results in 3.6. Test for both assets. Explain results.
The usual diagnostics that were used are the ACF test, Ljung-Box test, ARCH-LM test and Breusch-Godfrey test, which was done with code in r, upon further continuation of the code in question 3.6.

This is the code for asset 1:

```
Acf(model1$residuals, lag.max= 26)
Box.test(model1$residuals, lag = 26, type = c("Ljung-Box"))
ArchTest(model1$residuals, lags= 12)
bgtest(model1, order = 12)
```

This is the code for asset 2:

```
Acf(model2$residuals, lag.max= 26)
Box.test(model2$residuals, lag = 26, type = c("Ljung-Box"))
ArchTest(model2$residuals, lags= 12)
bgtest(model2, order = 12)
```

Figure 3.1: ACF model 1

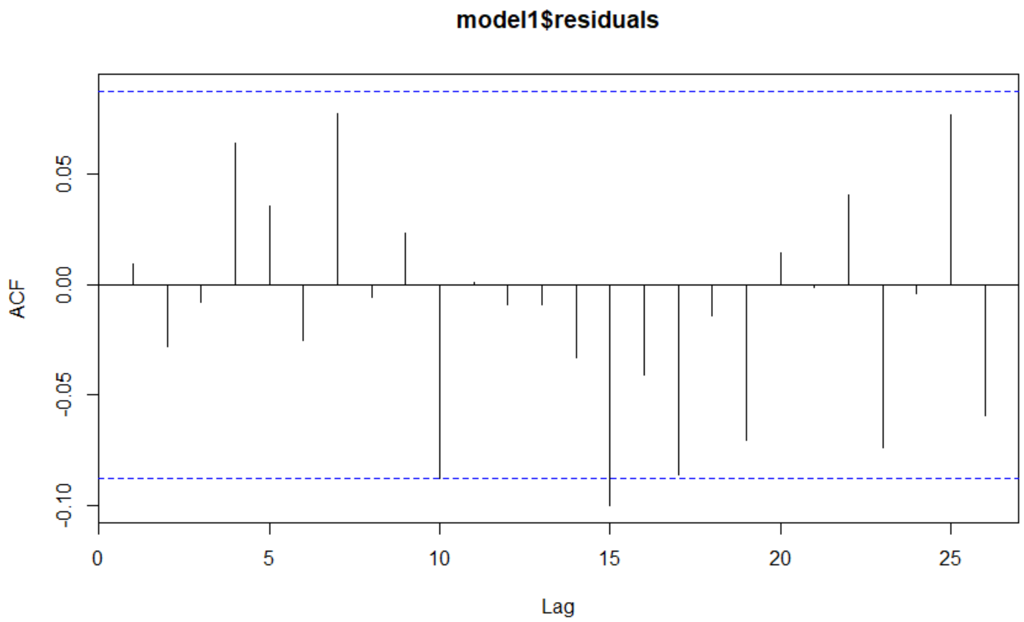
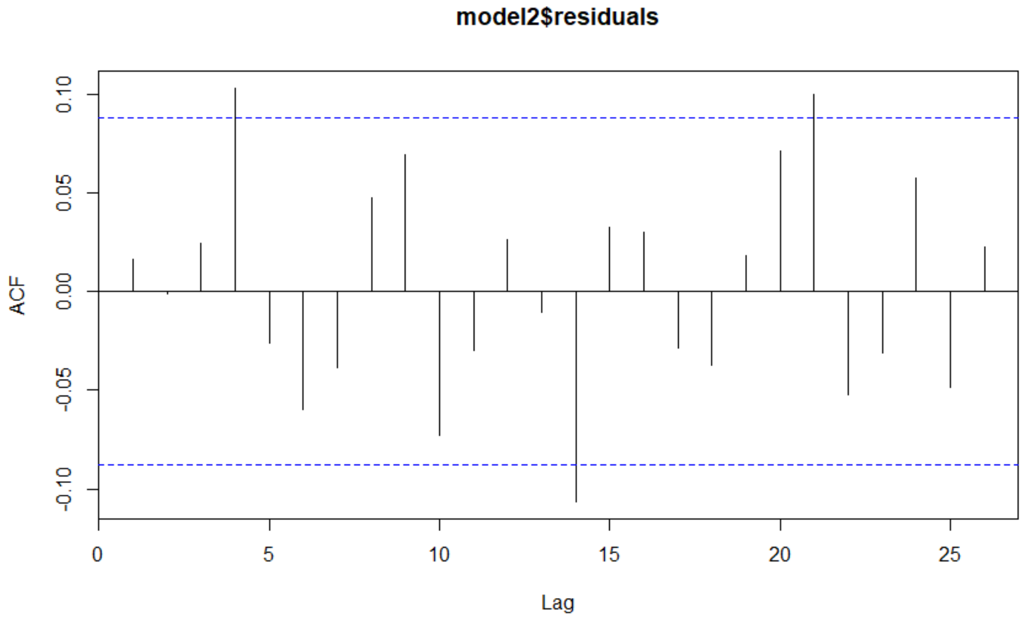


Figure 3.2: ACF model 2



The ACF shows for both models that the residuals may not be autocorrelated which is furthermore confirmed by the Ljung-Box test. The Ljung-Box test for both models computed a p-value above 0.05, which indicates that the residuals may not be autocorrelated. The ARCH-LM test for both models computed a p-value of above 0.05, which means that we do not reject our null hypothesis, which says that there are no arch effect. This means that the ARCH-LM test confirms with the other test that the squared residuals of the tests are probably white noise, so the residuals are probably homoscedastic. The Breusch-Godfrey test for both computed a p-value above 0.05, which shows evidence that there may not be serial correlation in the residuals, whereby a HAC-correction is not required.

Question 3.8 How can you check if you need more factors? Explain.

The method would be to make use of the idea that if m is the appropriate number of common factors, then there should be no significant decrease in the cross-sectional variance of the asset specific error ϵ_{it} when the number of factors moves from m to $m+1$.
To take the method into practice, we have used this code in r to get a hint on that we do not need more factors.[3]

$$\mathrm{cov}(\epsilon_{i,t}, \epsilon_{j,t})$$

which leads to this,

```
#Check if the factor model needs more factors,
should be very close to 0,
which means that we dont need more
vec_errs <- cbind(model1$residuals, model2$residuals)
#get correlation matrix
cond <- cor(vec_errs)
print(cond)
```

This lead to the final result of,

$$\text{cov}(\epsilon_{i,t}, \epsilon_{j,t}) = -0.0106592$$

This hints that our model does not need more factors.

Question 3.9 Giving your results in 3.6, estimate the value-at-risk at 95% **probability level assuming a Gaussian distribution for the minimum variance portfolio with the two assets. Hint: calculate the weights on asset 1 and 2 by using the following formulas**

$$w_1 = \frac{(var[r_{2,t}] - cov[r_{1,t}, r_{2,t}])}{(var[r_{1,t}] + var[r_{2,t}] - 2cov[r_{1,t}, r_{2,t}]}$$

$$w_1 = \frac{(var[0.61] - cov[0.0811])}{(var[0.707] + var[0.614] - 2cov[0.0811]} = 46\%$$

$$w_2 = 1 - 46\% = 54\%$$

The value-at-risk, is calculated in r,

```
# Estimate VaR
# level of significance of var
p <-0.05
# get the (1-p) quantile from standard normal distribution
quantile_n <- qnorm((1-p), mean = 0, sd = 1)
# capture betas (factor loadings)
loading1 <-model1_data$coefficients[2]
loading2 <-model2_data$coefficients[2]
# capture variance of shocks (specific factor)
var_shock1 <- model1_data$sigma^2
var_shock2 <- model2_data$sigma^2
# mean and variance factor
var_f <- var(ft)
mu_f <- mean(ft)
# weights portfolio
w1 <- (var_r2 - cov12) / ((var_r1 + var_r2 - (2*cov12)))
```

```
w2 <- 1- w1
# mean assets
mean_r1 <- loading1*mu_f
mean_r2 <- loading2*mu_f
# mean of portfolio
mu_p <- w1* mean_r1 + w2*mean_r2
# variance/covariance of assets
var_r1 <- (loading1^2)*var_f  + var_shock1
var_r2 <- (loading2^2)*var_f  + var_shock2
cov12 <- loading1*loading2*var_f
#variance of portfolio
varp <- (w1^2)*var_r1 + (w2^2)*var_r2 + 2*w1*w2*cov12
# var estimation
var_est <- mu_p + quantile_n*sqrt(varp)
print(var_est)
```

$$\text{VaR} = W(\mu + z_{1-\alpha}\sigma_{t+1})$$

$$\text{VaR} = W(0.0634 + 1.6448 \cdot \sqrt{0.3689}) = 1.062606$$

The value at risk is computed to be 106%, which means that there is a 5% probability the investor may lose the whole investment and a 95% probability of losing less. Note, that it is not possible to lose more than 100% of the investment, since there is no short positions.

Chapter 4

Cointegration and Price Spread

Let $p_{1,t}$ be the log price of asset 1 and $p_{2,t}$ be the log price of asset 2.
Question 4.1 How an equilibrium relationship between $p_{1,t}$ and $p_{2,t}$ can be expressed?
An equilibrium can be expressed through the use of regression models to write a long-term relationship between the log-prices. The expression is:

$$p_{1,t} = \beta_0 + \beta_1 p_{2,t} + \epsilon_t$$

ϵ_t are the deviation from the equilibrium, this can also be called the scaled spread.

Question 4.2 What are the conditions for your answer in 4.1 to be a long-term relationship.

The conditions for the long-term relationship are the following: $p_{1,t}$ and $p_{2,t}$ must follow random walk models which means that the log-prices of each assets moves in unpredictable ways. furthermore ϵ_t, the scaled price spread, must be stationary and mean-reverting, tho it must not necessarily be white noise.

Question 4.3 How can you identify if asset 2 is overvalued in relation to asset 1?
To identify if one of the assets is overvalued in relation to the other the following expression can be estimated:

$$p_{2,t} = \beta_0 + \beta_1 p_{1,t} + \epsilon_t$$

Then the following expression will be applied to identify if asset 2 is overvalued in relation to asset 1

$$\epsilon_t = p_{2,t} - \beta_0 - \beta_1 p_{1,t}$$

If $\epsilon_t = 0$ then asset 2 is exactly on its fundamental value explained by asset 2 and therefore the prices are aligned with the fundamentals. If $\epsilon_t < 0$ the asset 2 is overvalued since it is above the expected price.

The excel file "cointegration.xlsx" contains time series representing two asset daily log-prices, a total of 500 observations. The following questions refer to this dataset.
Question 4.4 Estimate a cointegration relationship

$$p_{1,t} = \beta_0 + \beta_{1p2,t} + \epsilon_t$$

Analyze the results. Is this regression valid?
The code used in r to validate the regression was,

```
# Test data
test_ret1<-ur.df(p1, type = "drift", lags = 10,
selectlags = "AIC")
summary(test_ret1)
test_ret2<-ur.df(p2, type = "drift", lags = 10,
selectlags = "AIC")
summary(test_ret2)

# Estimate the cointegration relationship
eqm <- lm(p1 ~ p2)
summary(eqm)
ACF1 <- eqm$residuals
plot(ACF1, type='l')
Acf(ACF1, lag.max = 26)
test_ret3<-ur.df(ACF1, type = "none", lags = 10,
selectlags = "AIC")
summary(test_ret3)
```

Figure 4.1: Cointegration relationship

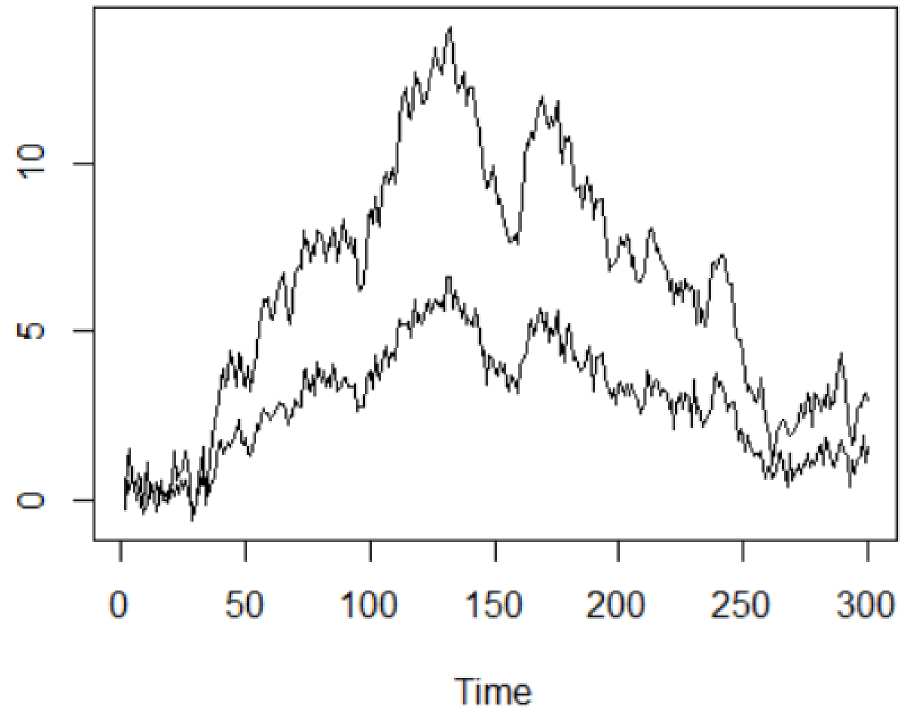
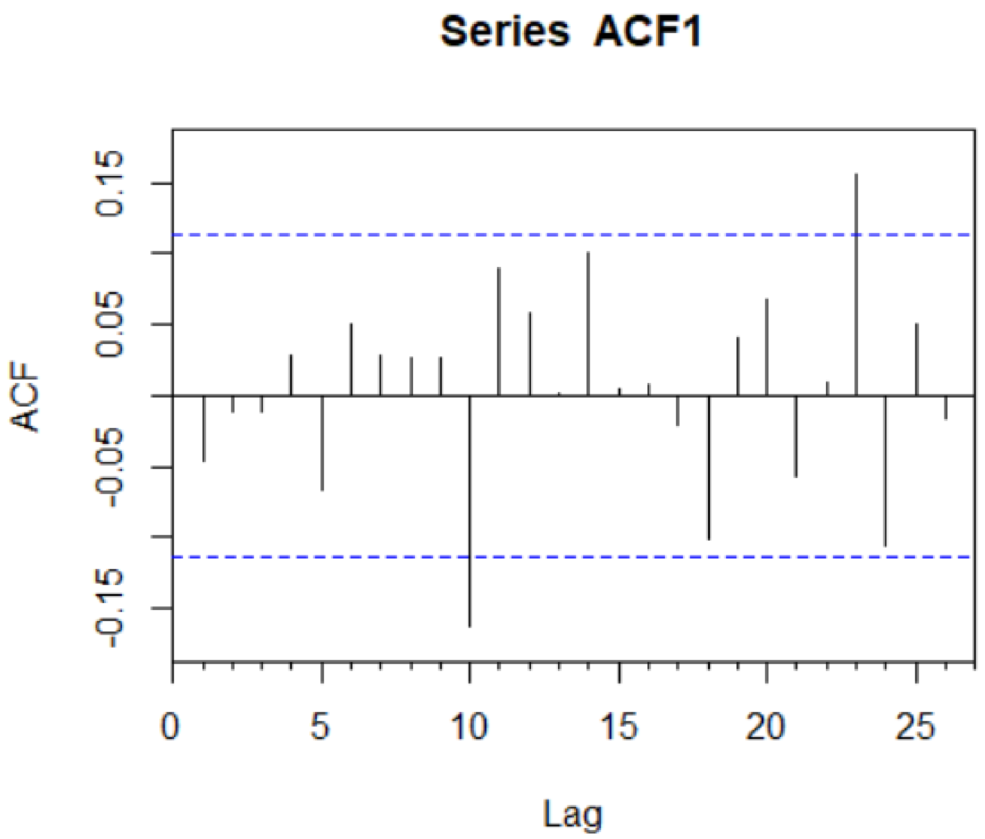


Figure 4.2: ACF



The Augmented Dickey-Fuller test on P_1 provided a non-significant p-value, while it is significant regarding P_2. This means that for p_1 there is an unit root present, since we do not reject the null hypothesis, while for p_2, we reject the null, which means that p_2 do not follow a random walk.
The cointegration coefficient is significant with a p-value below 0.05. Despite, that the cointegration coefficient is significant, the regression is valid, because the ADF test rejects our null regarding our error terms being stationary, we cannot estimate a cointegration relationship, because P_2 does not follow a random walk.

Question 4.5 Giving results in 4.4, is asset 1 overvalued or undervalued in relation to asset 2?
Based on the last observation of log-prices in the dataset and the output for β_1 and β_0 the disequilibrium/equilibrium is calculated with the following formula:

$$\epsilon_t = p_{1,t} - \beta_0 - \beta_1 p_{2,t}$$

$$-0.5872487 = 2.991531 - 0.23696 - 2.13987 \cdot 1.5617097$$

The function above provides a negative number of -0.5872847. This means asset 1 is overvalued, because $\epsilon_t < 0$, and thereby asset 1 is above its fundamental value explained by asset 2.

Question 4.6 Fix $\lambda = \sigma$, *i.e.*you trade when the price spread deviates 1 standard deviation from its mean. Giving the results in 4.4, is there a trade decision?
The following code has been used

```
delta <- sqrt(var(ect))
current_price_spread <- ACF1[length(ect)]
trade_decision1 <-   current_price_spread + delta
trade_decision1
# if negative asset 1 is undervalued
and asset 2 overvalued
trade_decision2 <-  current_price_spread - delta
trade_decision2
# if positive asset 1 is overvalued
and asset 2 is undervalued.
# If either, then no trading decision.
# we should buy asset 1 as it is undervalued
and short asset 2
```

Trade decision 1 provided a negative number, which means asset one is undervalued and asset 2 is overvalued. You trade when the price spread deviates 1 standard deviation from its mean. Giving the results in 4.4, is there a trade decision?
If there exist a cointegration relationship then yes, we should go long asset 1 as it is undervalued and short asset 2, which is overvalued. But since there is not a cointegration relationship, this decision is not evidence based and therefore the trading decision would be no.

Bibliography

[1] E.J. Elton, M.J. Gruber, and S.J. Brown. *Modern Portfolio Theory and Investment Analysis, 9th Edition: Ninth Edition*. Wiley Global Education, 2013. ISBN: 9781118805800. URL: `https://books.google.dk/books?id=aIFbAgAAQBAJ`.

[2] Xue-Zhong He, Kai Li, and Chuncheng Wang. "Volatility clustering: A nonlinear theoretical approach". In: *Journal of Economic Behavior Organization* 130 (2016), pp. 274–297. ISSN: 0167-2681. DOI: `https://doi.org/10.1016/j.jebo.2016.07.020`. URL: `https://www.sciencedirect.com/science/article/pii/S016726811630155X`.

[3] R.S. Tsay. *Analysis of Financial Time Series: Third Edition*. Aug. 2010. DOI: `10.1002/9780470644560`.