Computer Lab 1: Time series analysis and statistical properties of asset returns

Table of contents:

1. *Background*
2. *Results*
3. *Description of the estimation approach and purpose of the tests*
4. *Discussion*
5. *Appendix*

Victoria Lagerstedt & Kajsa Hansson Willis

Fall 2024
EXTQ30: Empirical Finance

1. Background

This computer lab is about time series analysis and statistical properties of asset returns. The first part analyses the log returns defined by the MA(3) process shown in equation 1 where e_t is a randomly generated series.

$$r_t = 0.1 + 0.3e_{t-1} - 0.2e_{t-2} + 0.1e_{t-3} + e_t \qquad \text{Equation 1}$$

The second part analyses the log returns u_t which is a randomly generated series from a different distribution than e_t.

When analyzing if a series is normally distributed, skewness and kurtosis are often used. Skewness measures how symmetrical the distribution is. If the skewness has a positive value this means the distribution is shifted to the right. The skewness of a normal distribution is zero. Kurtosis measures how much of the data is in the peak respectively the tails. If the kurtosis has a positive value it means that the peak is smaller. For a normal distribution, the kurtosis is 3,0.

To test if the data's skewness and kurtosis match a normal distribution a Jarque-Bera test is used, see equation 2. The Jarque-Bera value is compared to a chi-squared distribution.

$$BJ = \frac{T}{6}(skewness)^2 + \frac{T}{24}(excess\ kurtosis)^2 \qquad \text{Equation 2}$$

Testing the Random Walk hypothesis through methods like the Ljung-Box and Variance Ratio tests allows analysis of whether data behaves randomly or exhibits systematic patterns. If data follows the Random Walk hypothesis, data points only have small variations from their previous values and can be described by equation 3. This is synonymous with that the autocorrelation functions should be strong for a low lag and decrease when the lag increases.

$$x_t' = x_{t-1} + \varepsilon_t \qquad \text{Equation 3}$$

The Ljung-Box test is a test that checks if the ACFs are zero up to a chosen lag m, see equation 4. The value is compared to a chi-distribution to check the hypothesis.

$$Q_m' \equiv T(T+2)\sum_{k=1}^{m}\frac{\rho_k^2}{(T-k)} \qquad \text{Equation 4}$$

The variance ratio is the ratio of the variance of q-period overlapping returns times 1/q over the variance of one-period returns, see equation 5.

$$VR(q) = \frac{var[r_t(q)]}{q \cdot var[r_t]} \qquad \text{Equation 5}$$

A moving average process is a process of the following formula, see equation 6. The autocorrelation for a MA process of order q is zero for lags higher than q.

$$X_t = e_t + c_t e_{t-1} + ... + c_q e_{t-q} \qquad \text{Equation 6}$$

2. *Results*

2.1 Results for r_t

To test if r_t, which are the log returns, follow a normal distribution the z-value for skewness and excess kurtosis was calculated. The Z-value for the skewness is 0,0336 and the Z-value for the excess kurtosis is -1,14. A Jarque-Bera statistics test on the skewness and excess kurtosis was also performed. The Jarque-Bera statistic for the joint test of r_t is 1,41.

The p-value for normal distribution for both skewness and excess kurtosis was calculated as well as the p-value for chi-distribution. The p-value for normal distribution for skewness is 0,737 and the p-value for normal distribution for kurtosis is 0,255. The p-value for chi-distribution for BJ is 0,506, see table 1.

Table 1: Skewness, kurtosis and p-values for normal distribution and chi distribution for r_t

	Test for skewness:	Test for excess kurtosis:
	0,02606622	-0,176699498
	Positive means a little to the right	Negative
Z:	0,336008301	-1,138878187
Normal distribution	0,736864607	0,254753965
BJ:	1,409945	
Chi distribution	0,505878	

The sample ACFs of order 1 to 5 of r_t were calculated to be as shown in table 2.

Table 2: ACFs for order 1 to 5 of r_t and their p-value and confidence interval.

	1	2	3	4	5
The values of the different lags:					
	0,222860371 K=1	-0,110982 K=2	0,088198581 K=3	0,0206648 K=4	-0,003151 K=5
	0,223134675 K=1	-0,108953 K=2	0,087882485 K=3	0,0183729 K=4	
	0,223374589 K=1	-0,108962 K=2	0,087942478 K=3		
	0,223964106 K=1	-0,107219 K=2			
	0,223817525 K=1				
ACF:	0,223430253	-0,109029	0,088007848	0,0195188	-0,003151
p-value:	2,01239E-12	0,0005986	0,005597022	0,5389141	0,9209745
confidence interval:	0,161168801	-0,17129	0,025746396	-0,042743	-0,065413
	0,285691705	-0,046768	0,1502693	0,0817803	0,0591101

The Ljung-Box statistics test was performed and gave 69,6, see table 3. The p-value for the ljung-box statistics test is almost one, see table 4.

Table 3: Ljung-box statistics for the ACFs

Ljung-box statisics
69,55396356

Table 4: p-value for the Ljung-box statistics

Chi-squared value:
0,9999999999998730

It is observed that zero is within the confidence interval for the fourth and the fifth order but not for order 1-3, see figure 1.

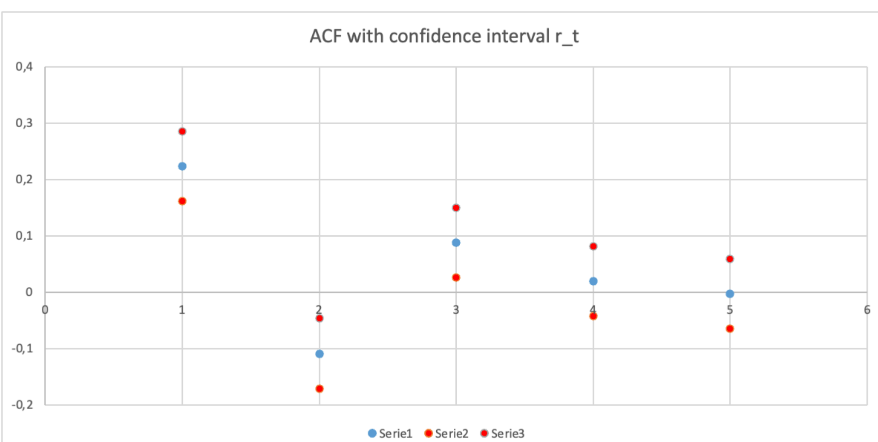

Figure 1: Correlogram for r_t with 95% confidence interval

Estimation of the partial autocorrelation function (PACF) was done using the regression tool in Excel and the table below, table 5, was calculated, showcasing the coefficients and the p-values for the variables in X.

Table 5: Coefficients and p-values for r_t from the regression tool in Excel

	Koefficienter	*p-värde*
Konstant	0,0839103	0,01161214
X-variabel 1	0,30464622	7,9006E-21
X-variabel 2	-0,23636403	1,9132E-12
X-variabel 3	0,19834588	3,8885E-09
X-variabel 4	-0,08896978	0,00736971
X-variabel 5	0,05211818	0,10171626

The analytic values of the ACFs given the parameters of the MA(3) process , calculated using equations 7, 8 and 9, are shown in table 6.

$$\rho_1 = \frac{\theta_1+\theta_1\theta_2+\theta_2\theta_3}{1+\theta_1^2+\theta_2^2+\theta_3^2} \qquad \textit{Equation 7}$$

$$\rho_2 = \frac{\theta_2+\theta_1\theta_3}{1+\theta_1^2+\theta_2^2+\theta_3^2} \qquad \textit{Equation 8}$$

$$\rho_3 = \frac{\theta_3}{1+\theta_1^2+\theta_2^2+\theta_3^2} \qquad \textit{Equation 9}$$

The values are 0,193, -0,149 and 0,088 respectively.

Table 6: Estimated ACFs through given coefficients

Estimating ACF through coefficients

Number	Coeff	ACF
0	1	0,192982
1	0,3	-0,149123
2	-0,2	0,087719
3	0,1	

The variance ratio for overlapping 5th differences of r_t is 1,31. This, as well as the test statistics, can be observed in table 7. The p-value of the test statistics is 0,00000723.

Table 7: Test statistics for the Random Walk hypothesis and the variance ratio for r_t

RW of r_t

average of one-period returns	antal (T)	m	q
0,107620322	993	4920,1007	5

Unconditional variance (a):	**variance ratio:**
1,174747219	1,31194144
	test statistics:
	0,31194144
Conditional variance ©:	**variance of test statistics**
1,54119956	0,00483384
p-value of test statistics:	
0,000007233	

2.2 Results for u_t

If the log return is assumed to be u_t, the distribution is portrayed in the histogram below, see figure 2. It can be observed that there is an outlier that is -6,78.

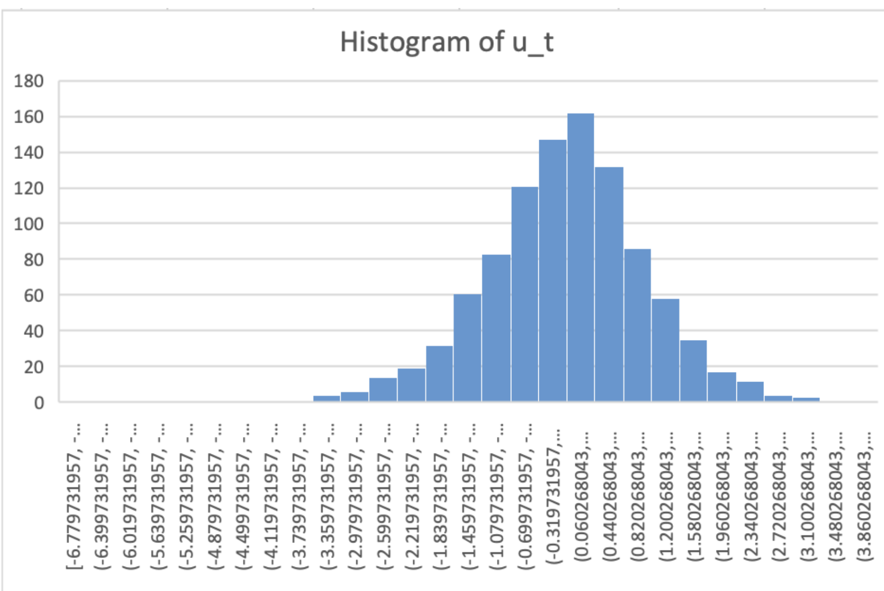

Figure 2: Histogram plot of u_t

To test if it follows a normal distribution, tests for skewness and excess kurtosis are done with Z-values and normal distribution calculated, as well as the Bera-Jarque statistics and its Chi-squared distribution. The results can be seen in table 8. The test for skewness is -0,257 and the test for excess kurtosis is 1,84. The Z-values are -3,31 and 11,9 respectively. The p-values for normal distribution are 0,000939 and a value that Excel rounds to zero. The Bera-Jarque statistic is 152 which gives a Chi-squared distribution significance value of 1,00.

Table 8: Skewness, kurtosis and p-values for normal distribution and chi distribution for u_t

	Test for skewness:		Test for excess kurtosis:
	-0,256646755		1,840881449
	Negative means a little to the left		Positive
Z:	-3,308321678		11,86500101
Normal distribution:	0,000938569		0
BJ:		151,723241	
Chi distribution:		1	

The sample ACFs were calculated according to and can be observed in table 9.

Table 9: ACFs for order 1 to 5 of u_t and their p-value and confidence interval

	1	2	3	4	5
The values of the different lags:	0,047144531 K=1	-0,0489764 K=2	0,012295751 K=3	-0,0278218 K=4	-0,0544025 K=5
	0,046897376 K=1	-0,0482364 K=2	0,012713959 K=3	-0,0278961 K=4	
	0,047391024 K=1	-0,0480081 K=2	0,012675305 K=3		
	0,046709523 K=1	-0,0479225 K=2			
	0,046781196 K=1				
ACF (averaged):	0,04698473	-0,0482859	0,012561672	-0,0278589	-0,0544025
p-value:	0,137926413	0,12734877	0,691633958	0,37904669	0,08583778
confidence interval:	-0,015089093	-0,1103597	-0,049512151	-0,0899327	-0,1164763
	0,109058553	0,01378797	0,074635494	0,0342149	0,00767136

It can be observed that the autocorrelation functions of all five orders have zero within the confidence intervals. This is illustrated in figure 3.

The Ljung-Box statistics test results in 8,45, see table 10. The p-value for the Ljung-Box statistics test is 0,867, see table 11.

Table 10: The Ljung-Box statistics of u_t

Ljung-box statisics
8,449596351

Table 11: The p-value from the Ljung-Box statistics test of u_t

Chi-squared value:
0,866863467

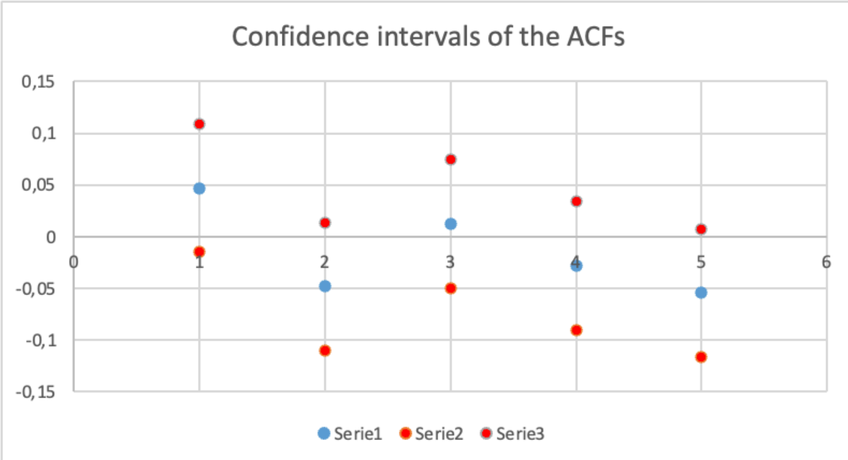

Figure 3: Correlogram with the confidence intervals of the ACFs of u_t

Estimation of the partial autocorrelation function was done using the regression tool in Excel and the table below, table 12, was calculated, showcasing the coefficients and the p-values for the variables in X.

Table 12: Coefficients, p-values and other information for u_t from the regression tool in Excel

	Koefficienter	*Standardfel*	*t-kvot*	*p-värde*	*Nedre 95%*	*Övre 95%*	*Nedre 95,0%*	*Övre 95,0%*
Konstant	0.06198611	0.03485535	1.77838119	0.07564842	-0.0064128	0.13038505	-0.0064128	0.13038505
X-variabel 1	0.04952084	0.03173395	1.56050018	0.11896164	-0.0127528	0.11179445	-0.0127528	0.11179445
X-variabel 2	-0.0523625	0.03175979	-1.6487051	0.09952565	-0.1146869	0.00996179	-0.1146869	0.00996179
X-variabel 3	0.01673407	0.03181761	0.52593732	0.59904966	-0.0457037	0.07917184	-0.0457037	0.07917184
X-variabel 4	-0.0295638	0.03178146	-0.9302211	0.35248363	-0.0919306	0.03280306	-0.0919306	0.03280306
X-variabel 5	-0.0501725	0.03175701	-1.5798866	0.11445259	-0.1124913	0.01214639	-0.1124913	0.01214639

The variance ratio using an overlap of 5 is 1,03. This value and the test statistics for the Random Walk hypothesis can be observed in table 13. The p-value for the test statistics is 0,665.

Table 13: Test statistics for the Random Walk hypothesis and the variance ratio for u_t

RW of u_t

average of one-period returns	antal (T)	m	q
0,057732457	993	4920,1007	5

Unconditional variance (a):	**variance ratio:**
1,193837564	1,0300703
	test statistics:
	0,0300703
Conditional variance ©:	**variance of test statistics**
1,229736587	0,0048338
p-value:	
0,6653743	

3. *Description of the estimation approach*

The estimation approach has been in accordance with the theories that have been used. These are the normal distribution, the autocorrelation function as well as the partial autocorrelation function, the moving average model, the Random Walk hypothesis, the variance ratio, the Ljung-Box test statistics and the variance of the test statistics.

All correlations for the autocorrelation function were calculated separately for the specific time lags. The average of the different values for the particular time lag was then calculated and used. Since the number of values differed between the time lags, it was deemed sufficient to use the same for each calculation after consultation with the TA.

There were fewer data points for the variance ratio and testing the Random Walk Hypothesis, so a lower value of T was used. Other than this, the formulas were applied, and the results are interpreted in the discussion below.

4. *Discussion*

4.1 Discussion about the r_t distribution

The calculated log returns' skewness, rt, was positive, which means the possibility of values higher than the mean is slightly higher than the values lower. The excess kurtosis of the log returns was negative, which means the peak of the distribution is lower and wider than that of a normal distribution.

For the normal distribution, the p-value test can help determine if a collection of data points follows it, or rather if it does not follow it. Considering that the null hypothesis is that the distribution follows the normal distribution if the p-value is small, it indicates a very low likelihood that the data follows a normal distribution. For instance, a p-value of 0,05 or lower indicates that the distribution is outside the 95% confidence interval of identifying a normal distribution, and thus, it is very likely that the data does not follow a normal distribution. A p-value of 0,74, which is the value for the skewness of r_t, is significantly higher, which indicates that it follows a normal distribution. Regarding the excess kurtosis of r_t, the p-value is 0,25, which is significantly lower, but it is still very reasonable that the data would follow a normal distribution. As mentioned in the results, the Bera-Jarque statistic for the r_t distribution is 1,41 and the Chi-squared significance value is 0,506. This value can be compared to 0,004 which would be the Chi-squared value that indicates that the distribution is outside the 95% confidence interval of following a normal distribution. Therefore, this test also shows that it is quite reasonable that the distribution of r_t is normal. These values are quite aligned with the possibility that r_t follows a normal distribution. With this in mind, it is not possible to discard the hypothesis that r_t follows the normal distribution. It is still not possible to say that the distribution follows a normal distribution for certain, but it is likely.

It was observed in the results that 0 is within the confidence interval for the fourth and fifth order of the autocorrelation functions of r_t. This is in line with the fact that it is a MA(3) process since the MA(3) process only has three coefficients. Therefore, the fourth and fifth-order coefficients would have to be zero. The first through third-order coefficients do not have zero within their confidence intervals, which strengthens this. The Ljung-Box statistics test results in 69,9 with a p-value that rounds to 1. Comparing this to the previously presented value of the 95% confidence interval of a Chi-squared distribution with two degrees of freedom, it can be concluded that 1 is higher, which means that it is not possible to discard the null hypothesis, and it is not unreasonable that the ACFs up to lag 5 are 0, which is the null hypothesis. If a probability of 95% that the null hypothesis is true was desired, the Chi-squared value would have to be 5,99, which is significantly higher than the value obtained.

The partial autocorrelation functions (PACFs) have the coefficients 0,305, -0,236, 0,198, -0,0890, and 0,0521 respectively. For an MA process, the PACF function should be damped exponential or sine functions. Every other value of the calculated PACFs is negative which means that it may follow a sine function, which would confirm the expectations. Most of the p-values, which can be observed in table 5, are quite small. Only the p-value of the fifth-order coefficient is above a significance level of 5%, which means that the null hypothesis that there is no correlation between r_t and r_{t-5} cannot be discarded, but it can be for the other dependent variables. This is in line with what was expected, however, with an MA(3) process, the fourth-order coefficient would also be expected to be discarded. That value is also quite low, so it is likely outside the significance level due to pure test variance.

The analytical values of the ACFs are observed in table 6. Comparing this to the sample ACFs in table 2, it can be observed that they are quite similar. The difference is 0,03 for order 1, 0,04 for order 2, and 0,0003 for order 3. These are quite percentually small differences, especially for the third order.

The significance level for the test statistics was determined to be 0,00000723, which is very low. This indicates that the Random Walk hypothesis can be discarded and that there are factors that affect the behavior of the distribution. This can be compared to the results from testing the sample ACFs, as the ACFs should be very strong for a lag of one and then decrease for each increasing lag if the distribution were to follow the Random Walk hypothesis. If the absolute value of the autocorrelation functions are observed, the value for the first lag is about 0,22, the second lag is about 0,11, the third lag is about 0,09 and the fourth lag is 0,019 and the fifth lag is 0,003. There is a clear pattern that the ACF decreases with increasing lags. This is in line with the Random Walk theory.

4.2 Discussion about the u_t distribution

Moving on to the scenario where u_t replaces r_t as the log-normal distribution, the skewness has a negative value which means it is shifted to the left which is also possible to see from the histogram. The kurtosis is smaller than 3 which means it has a wider peak. The p-values for the skewness and the kurtosis are 0,000939 and rounded to zero, respectively. These are small values, less than 0,05, that indicate that the distribution does not follow a normal distribution. A factor to take into consideration when looking at whether the process follows the normal distribution or not is that there is an outlier with a value below -6 that sticks out a lot. This can influence the tests to a large extent, and it is quite possible that the tests would be different if the outlier were to be taken out from the sample. However, the Chi-squared significance value rounds to 1, which means it is not possible to discard the null hypothesis. The null hypothesis says that the function is following a normal distribution. This is not in line with previous tests but a value of 1 in this test is still not high enough to confirm that the null hypothesis is very likely.

The sample ACFs for the u_t distribution can be observed in table 9. Zero is within the confidence interval for all of the orders of ACFs. This makes it very difficult to determine the order of an AR/MA/ARMA process to describe the distribution. The Ljung-Box statistics test gives a p-value of 0,867 which is high. This strengthens the observation that the ACFs up to lag 5 are zero.

The partial autocorrelation functions for u_t have coefficients that are closer to zero than the ones for r_t. Zero is in the 95% confidence intervals for the coefficients of all the dependent variables which means all the variables are statistically insignificant. This is also noticeable in the p-values; all the p-values for the variables are larger than 0,05 which means that the null hypothesis of no correlation between u_t and all the dependent variables cannot be discarded and there is probably no correlation.

The Random Walk hypothesis test statistics is about 0,03 which gives a p-value of 0,665 which is quite high. This indicates that the null hypothesis that the distribution follows a Random Walk should not be discarded. This can be compared to the results from testing the sample ACFs, and as previously mentioned the ACFs should be very strong for a lag of one and then decrease for each increasing lag if the distribution were to follow the Random Walk hypothesis. If the absolute value of the autocorrelation functions are observed for u_t, the value for the first lag is about 0,047, the second lag is about 0,048, the third lag is about 0,013 and the fourth lag is 0,028 and the fifth lag is 0,005. In this case, there is not a clear pattern that the ACF decreases with increasing lags. This does not indicate that the distribution follows the Random Walk hypothesis.

5. Appendix

The excel sheet "DataLab12024" is attached in Canvas.

5.1 Use of AI-based tools

ChatGPT has been used in a very limited amount. This AI-based tool has only been used to further develop our understanding of financial and statistical concepts, such as random walk. The information generated by ChatGPT was compared against the course literature to check its validity. This use of ChatGPT was very helpful since it simplified some concepts and helped us understand them better, which made the analysis in this report better.