Problem 1

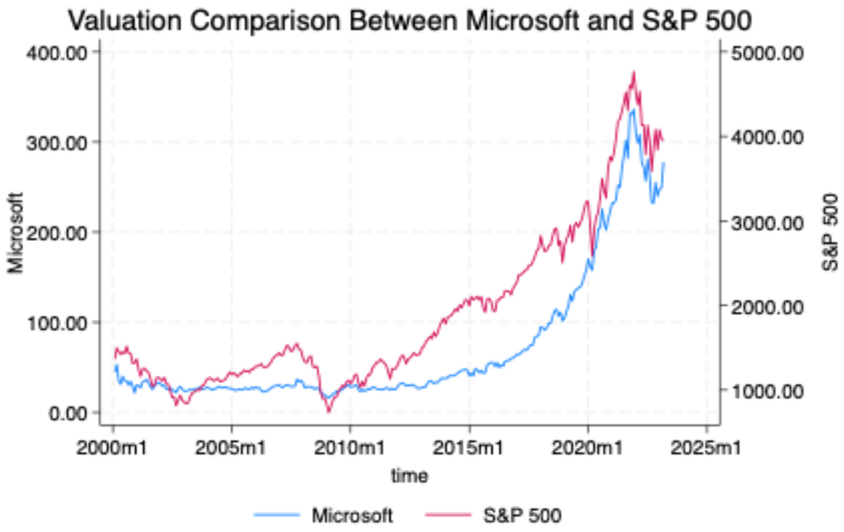

Figure 1: Valuation comparison

Tabel 1: Simple returns

Variable	Obs	Mean	Std. dev.	Min	Max
r_sp500	277	.0048435	.0446258	-.1694245	.1268441
r_mstf	277	.0098833	.0812261	-.3435294	.407781

Tabel 2: Log returns

Variable	Obs	Mean	Std. dev.	Min	Max
lnr_sp500	277	.003831	.0450407	-.1856365	.1194209
lnr_mstf	277	.0065946	.0809689	-.4208774	.3420147

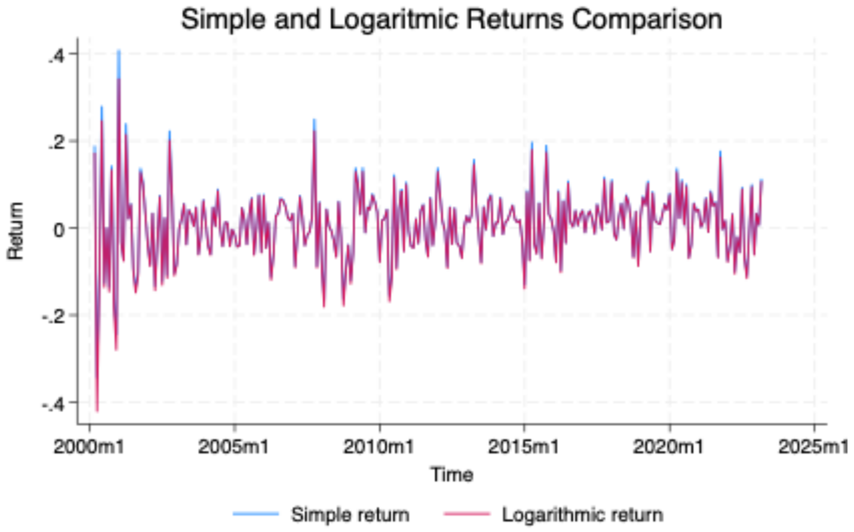

Figure 2:

Figure 2 shows the comparison between estimations for the simple return and the logarithmic return. As can be seen in the figure both estimates follow each other closely and the most notable difference are the local peaks for the simple return is slightly higher than that of the logarithmic return.

Table 3: Monthly and Annually Expected Log-Returns estimates for Microsoft and S&P 500.

	Montly	Annualy
Microsoft	.0065946	.07913517
S&P 500	.00383104	.04597244

In table 3 we have the estimated logarithmic returns for Microsoft and the S&P 500 index.

Table 4: Volatility

	Montly	Annualy
Microsoft	.08096893	.28048459
S&P 500	.04504072	.15602562

In table 4 we have the volatility of the estimated logarithmic returns for Microsoft and the S&P 500 index.

In Stata we got the standard deviation monthly and to annualize it we used the formula "Annual volatility= monthly volatility*12^0.5". This gives us the results shown in table 4. The volatility in both Microsoft and S&P 500 are larger annually which was expected since the price of the asset has more time to change.

Table: 5: Confidence interval

	Mean	Std. err.	[95% conf. interval]	
lnr_sp500	.003831	.0027062	-.0014964	.0091585
lnr_mstf	.0065946	.004865	-.0029825	.0161717

In table 5 we have the 95% confidence interval for the estimated logarithmic returns for Microsoft and the S&P 500 on the right side of the table. It can be seen that for both estimates zero is captured by the intervals. This can be interpreted as that the null hypothesis that the estimated mean values are equal to zero can't be rejected at the 5% significance level. By looking at figure 1 a clear upward trend can however be seen for both Microsoft and the S&P 500 Index and compared with the confidence interval this indicates that the data can suffer from high volatility or noise.

Table 6: Skewness and kurtosis

	Skewness	Kurtosis
Microsoft	-.3677255	6.721815
S&P 500	-.6675007	4.10026

Table 6 provides the skewness and kortosis of our estimated distribution of the estimated logarithmic return

The skewness of the data shows the symmetry of the data. If the skew is negative the histogram is left and the opposite if it is positive. This means that the majority of the observations are smaller than the median for both Microsoft and S&P 500. This is important because the standard deviation assumes a normal distribution and skewed data may lead to inaccurate predictions. The kurtosis measures the fatness of the tails with a normal distribution having a kurtosis of 3 or an excess kurtosis of 0. In this case both Microsoft and S&P 500 have a positive kurtosis with Microsoft having the largest, which means the tails of the histogram are fatter than a normal distribution implying more extreme values.

Table 7: JB test

				Joint test	
Variable	Obs	Pr(skewness)	Pr(kurtosis)	Adj chi2(2)	Prob>chi2
lnr_mstf	277	0.0128	0.0000	28.74	0.0000

7:

ac lag1 = **-0.1336**

CI lower for any lag	**-0.1178**
CI upper for any lag	**0.1178**

ac lag2 =	-0.0946785

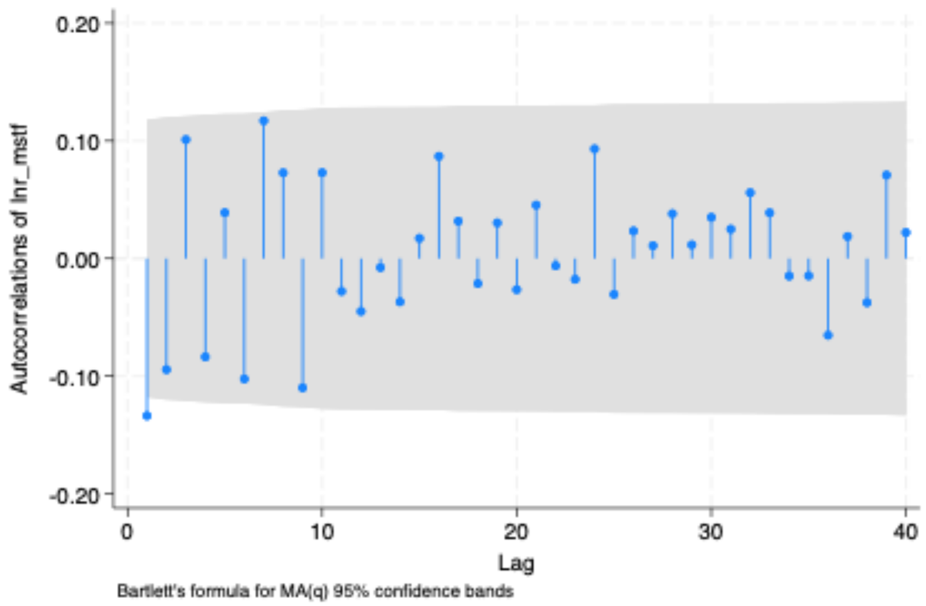

Figure: 3 correlogram

The JB test is conducted to test for normality and to see if it follows a normal distribution. Observing the results we can see that we can reject H0 at 1% level of significance meaning we reject the assumption that the data follows a normal distribution. The same is true for kurtosis which has a p-value approximately equal to 0. For the skewness we can reject H0 at the 5% significance level with a p-value of 0. 0128 and this Means that the data does not follow a normal distribution.

For the Autocorrelation calculations in excel we got the same results in Stata for lag 1 with a value of -0.1336 which is outside the confidence interval and -0.095 for lag 2 which is inside the confidence interval. This means it is statistically significance at the 95% level and suggests a small but significant negative autocorrelation meaning that the return of Microsoft tends to reverse direction between periods.

Table 8: regression with log returns of S&P 500 on time.

Source	SS	df	MS			
Model	.004354722	1	.004354722	Number of obs	=	277
Residual	.555557124	275	.002020208	F(1, 275)	=	2.16
				Prob > F	=	0.1432
Total	.559911846	276	.002028666	R-squared	=	0.0078
				Adj R-squared	=	0.0042
				Root MSE	=	.04495

lnr_sp500	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
t	.0000496	.0000338	1.47	0.143	-.0000169	.0001161
_cons	-.0031109	.0054451	-0.57	0.568	-.0138304	.0076085

Table 9: regression with log returns of Microsoft on time.

Source	SS	df	MS			
Model	.031451336	1	.031451336	Number of obs	=	277
Residual	1.77799556	275	.006465438	F(1, 275)	=	4.86
				Prob > F	=	0.0282
Total	1.8094469	276	.006555967	R-squared	=	0.0174
				Adj R-squared	=	0.0138
				Root MSE	=	.08041

lnr_mstf	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
t	.0001333	.0000604	2.21	0.028	.0000143	.0002522
_cons	-.0120615	.0097411	-1.24	0.217	-.0312382	.0071151

In table 8 and 9 we see the returns of both Microsoft and S&P 500 on time. Observing the p-values of table 8 we see that neither the variable time nor the intercept is significant at the 10% level and a R-squared equal to 0.0078 implying a very inaccurate model. This makes the model very inappropriate for predictions, leading to inaccurate forecasts. The same goes for the regression with Microsoft, with a R-squared of 0.017 and neither the intercept or the slope-coefficient being significant at the 10% level. Because of this, the conclusion that the model I too simple can be made and not accurate enough to make precise prediction.

Table:10 Regression of Log returns of Microsoft on S&P500

Source	SS	df	MS			
Model	.672237231	1	.672237231	Number of obs	=	277
Residual	1.13720967	275	.004135308	F(1, 275)	=	162.56
				Prob > F	=	0.0000
Total	1.8094469	276	.006555967	R-squared	=	0.3715
				Adj R-squared	=	0.3692
				Root MSE	=	.06431

lnr_mstf	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnr_sp500	1.095725	.0859398	12.75	0.000	.9265413	1.264908
_cons	.0023968	.0038778	0.62	0.537	-.0052371	.0100308

In table 10 we see a regression on log returns of Microsoft on the S&P500 which gave us a low R-squared equal to 0.37 and only the slope coefficient(S&P500) being significant. This implies that Microsoft returns are strongly correlated to the market and since the intercept is not significant it suggests that the return of Microsoft that is not predicted by the market is not different from 0. Since the R-squared is low it shows that the model in this case has a low explanatory power

Table:11 correlation estimate

covariance estimate	**0.0022**
correlation estimate	**0.6069**

Table:12 correlation between log returns of Microsoft and S&P500 and p-value.

	lnr_mstf	lnr_~500
lnr_mstf	1.0000	
lnr_sp500	0.6095 0.0000	1.0000

In Table 11 and 12 we see the correlation between log returns of Microsoft and S&P500 being 0.61 and a p-value of 0.000 which means that the correlation is statistically significant.

Problem 2

Part A

Figure:4 seasonally adjusted Industrial production

Figure 4 shows the seasonally adjusted industrial production for the US. The line seems to follow an upward exponential trajectory with cycles of both increasing and decreasing rates of production.

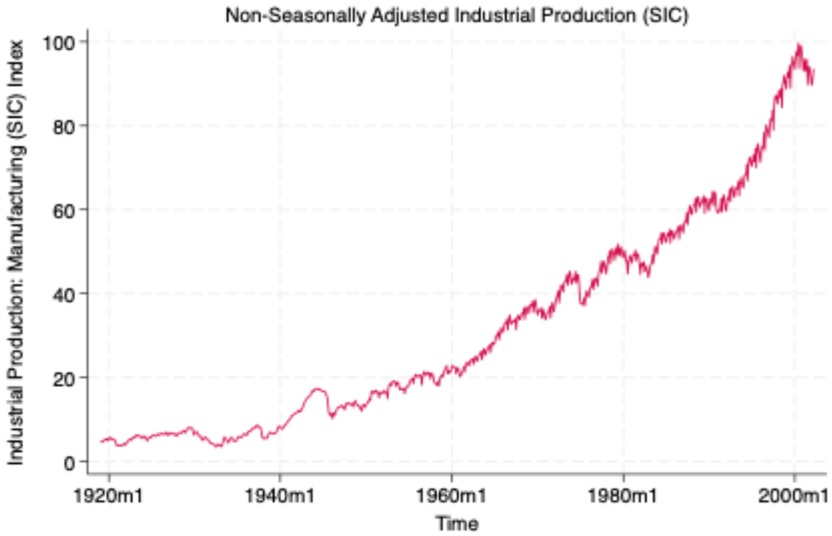

Figure: 5 non-seasonally adjusted production

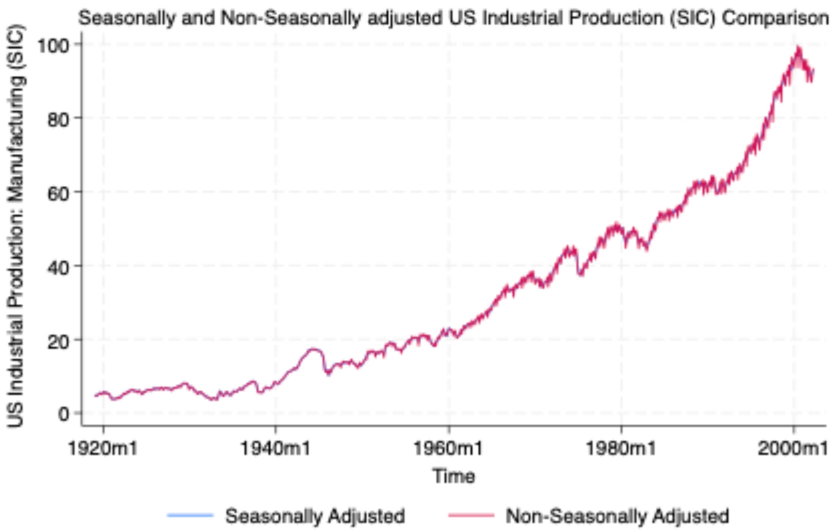

Figure 6: Comparison between non-seasonally and seasonally adjusted

Figure 6 compares the seasonally and non-seasonally adjusted industrial production. Unsurprisingly we see that both lines follow each other very closely but that the difference is that the non-seasonally adjusted line is much more volatile the closer we look. The seasonally adjusted line is, as the name suggests, adjusted to remove these spikes in the data.

Part B

Table 13: regression on adjusted industrial production

Source	SS	df	MS			
				Number of obs	=	1,000
				F(1, 998)	=	8898.79
Model	581251.886	1	581251.886	Prob > F	=	0.0000
Residual	65187.4464	998	65.3180825	R-squared	=	0.8992
				Adj R-squared	=	0.8991
Total	646439.332	999	647.086419	Root MSE	=	8.082

IPMANSICN	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
month	.0835166	.0008853	94.33	0.000	.0817793	.085254
_cons	31.09747	.2556603	121.64	0.000	30.59578	31.59917

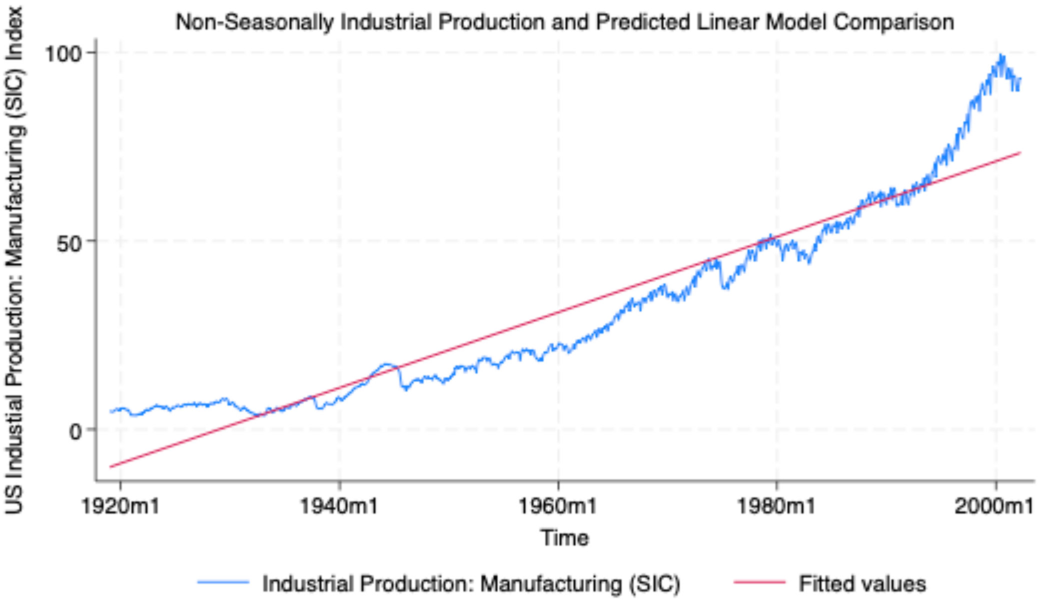

Figure: 7 Non-seasonally adjusted industrial production with estimated regression line.

Table: 14 Regression on log adjusted industrial production

Source	SS	df	MS			
				Number of obs	=	1,000
				F(1, 998)	=	27327.82
Model	861.99132	1	861.99132	Prob > F	=	0.0000
Residual	31.4795454	998	.031542631	R-squared	=	0.9648
				Adj R-squared	=	0.9647
Total	893.470866	999	.894365231	Root MSE	=	.1776

l_IPMANSICN	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
month	.0032162	.0000195	165.31	0.000	.003178	.0032544
_cons	3.044753	.0056182	541.95	0.000	3.033728	3.055778

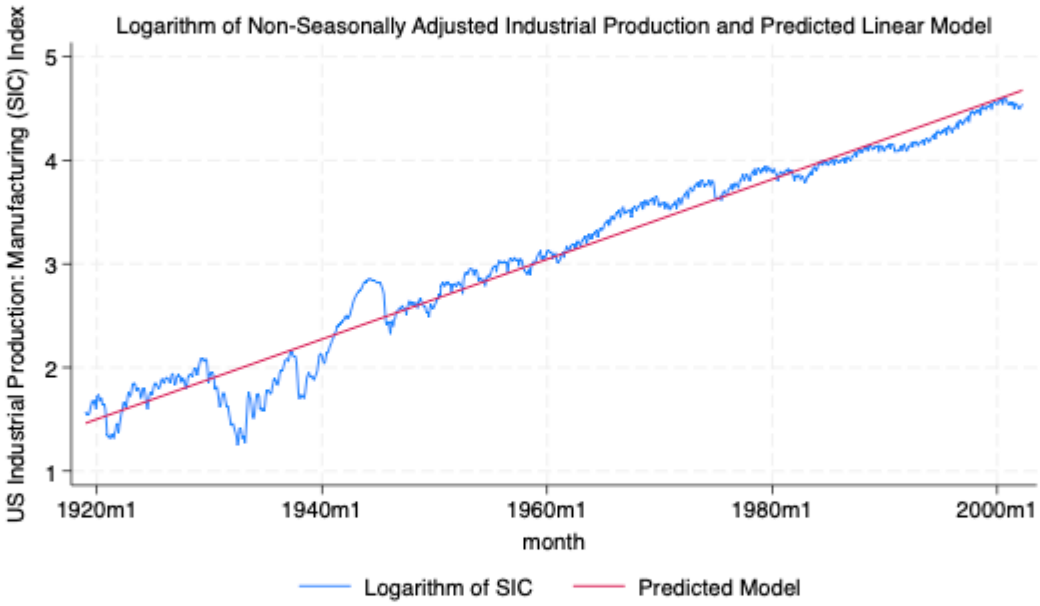

Figure: 8 log non-seasonally adjusted production and fitted line

The R^2 increases when regressing on the logarithmic value, from around 0.9 to 0.96, as can be seen in table 13 and 14. This is visualized from the figures where the predicted regression line seems to fit the logarithmic value of the SIC better. The logarithmic SIC seems to reasonably follow a straight line when looking at the period after 1950. In the period before the data suffers more from variations from the trend. This is the period where the world saw both the great depressions of the 1920s and WW2, two huge events that effecter the production of the US.

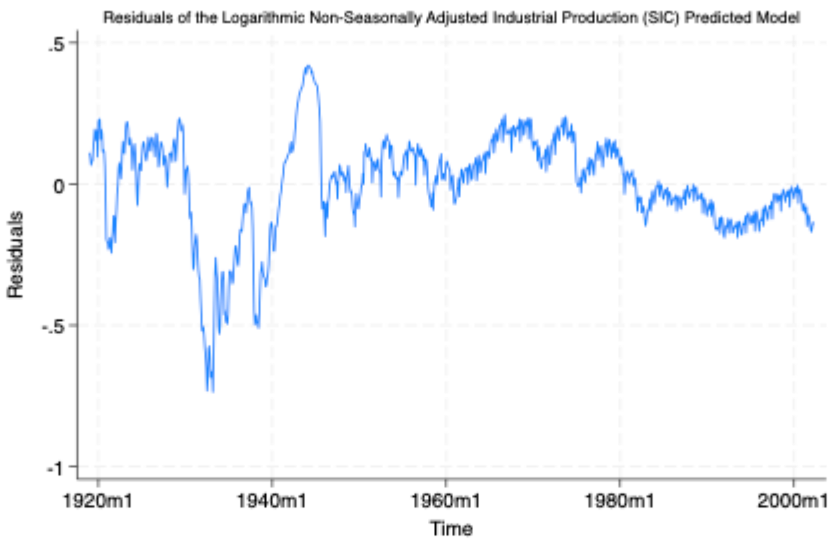

Figure: 9 residuals of non-seasonally adjusted production.

Figure 9 are the residuals from our estimated linear model of logarithmic industrial production. There seems to be some kind of trend in the residuals since it doesn't seem like the residuals are randomly distributed around the mean zero. For example, in the period 1930-1940 we see that the residuals are constantly below zero and periods following a decreasing residual is likely to also be decreasing. This is a sign that autocorrelation is present in the data.

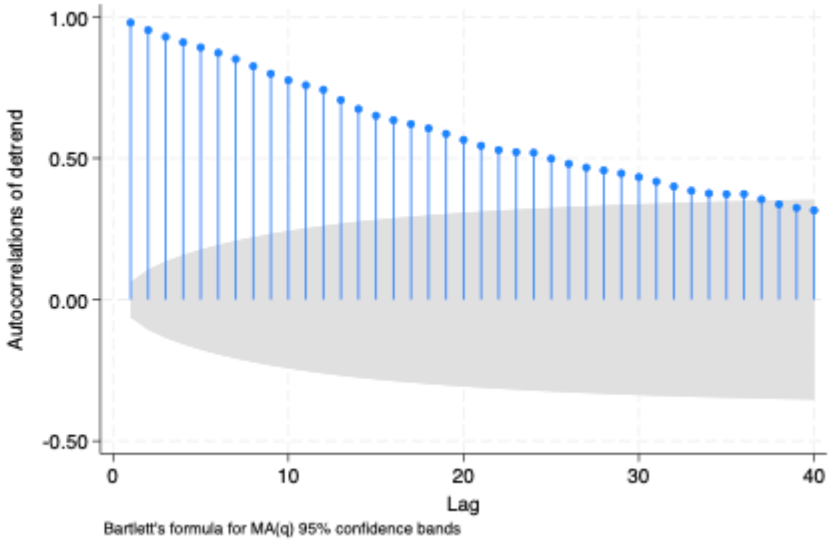

Figure: 10 Correlogram on autocorrelation

In figure 10 the autocorrelation theory is tested and as is shown we cannot reject the hypothesis that autocorrelation is present until a lag of 38. This not only indicates that autocorrelation is present, but it is very strong in the data. For the real world this means that a period with increasing growth is most likely followed by a period of increasing growth and vice versa. The reason that the autocorrelation is so strong is probably due to the fact that economic periods of growth often occur over many years and our data looks at the monthly values. Declines are often shorter but after a drop of productivity it takes time to catch up to the levels production where at before the drop resulting in multiple periods of below mean residuals in the data.

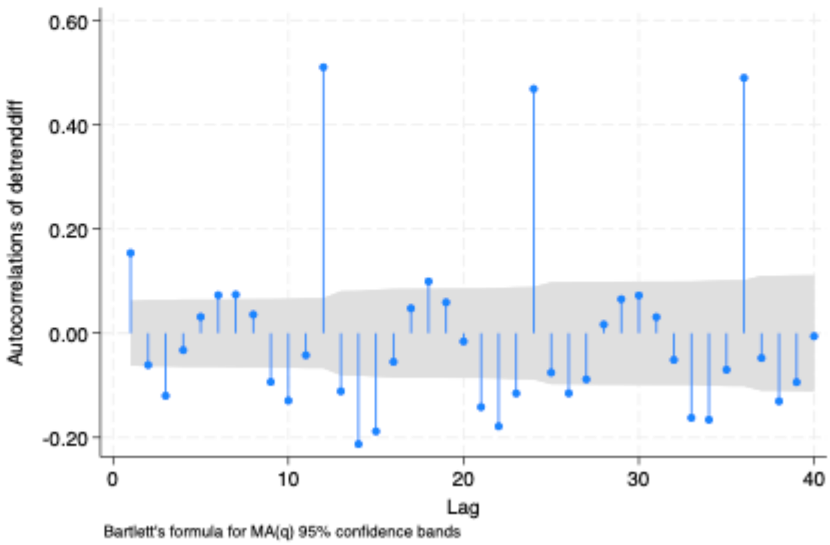

Figure: 11 Correlogram of the difference in residuals from one period to another.

Figure 11 is a correlogram of the residuals from period t minus the residuals from period t – 1. This helps us determine if any seasonal or cyclical component is present in the series. As can be seen in the figure the autocorrelation oscillates and has high points at intervals at twelve. Since the data ties into economic theory where cycles often can last years, and our series uses monthly data, what the correlogram captures is likely a seasonal component in the series. There can be many reasons why this is, for example household spendings increases around Christmas, but that is outside the scope of the assignment.

Table: 15 Regression of the estimated change in residuals for each month.

Source	SS	df	MS			
				Number of obs	=	999
				F(1, 997)	=	50.31
Model	.059980402	1	.059980402	Prob > F	=	0.0000
Residual	1.18874412	997	.001192321	R-squared	=	0.0480
				Adj R-squared	=	0.0471
Total	1.24872453	998	.001251227	Root MSE	=	.03453

detrenddiff	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
dummy	-.0022444	.0003164	-7.09	0.000	-.0028653	-.0016234
_cons	.0143115	.002326	6.15	0.000	.0097471	.018876

Table 15 is a regression on the difference of the residuals on its lagged with one month counterpart. The R^2 is only around 0.05 which is very low indicating that there is much white present in the residuals.

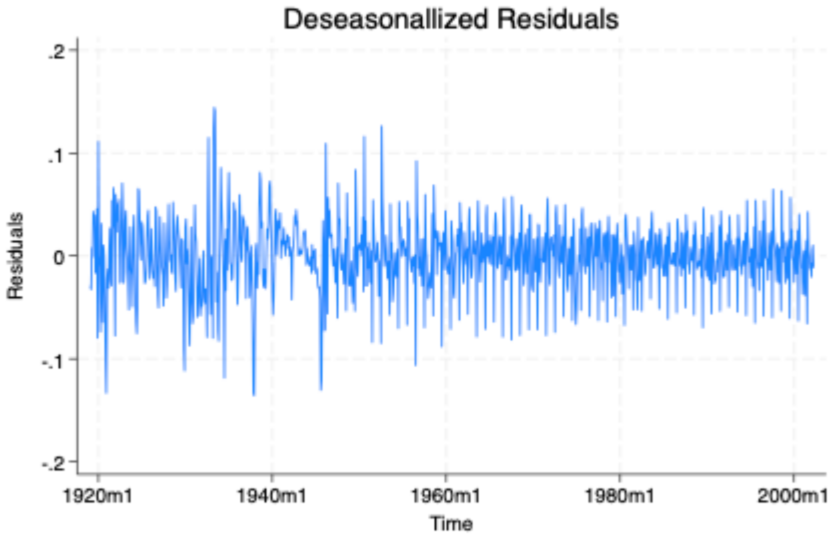

Figure: 12 Deseasonallized residuals.

Figure 12 is the deseasonalized residuals and after the 1960's the residuals seem to mostly indicate white noise since they oscillate around the mean in a what appears to be a random way. In the period before 1960 the residuals are more volatile in line with the events in the US and the world at that time. Some parts of this period can be interpreted as including a cycle component, for example in the years following 1940 there seems to be a downward trend in the residuals, but this can't be confirmed from the figure alone.

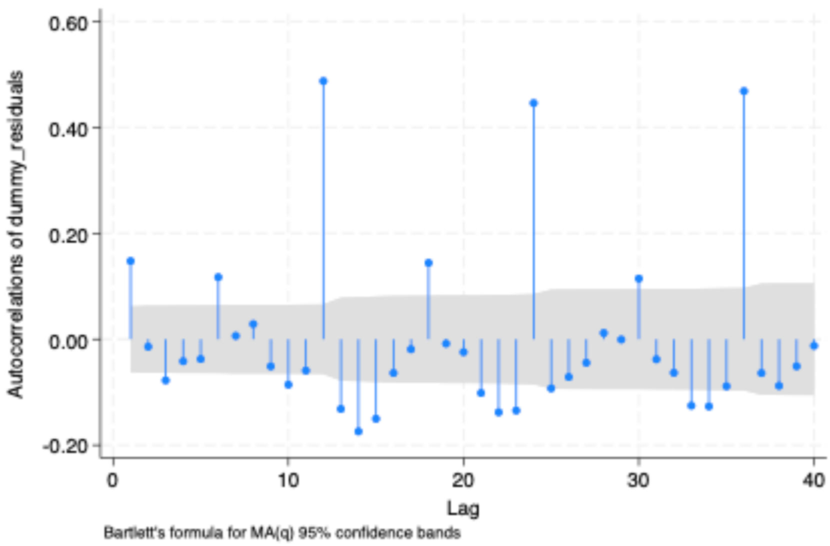

Figure 13: Correlogram of the residuals from previous regression.

Not very much has changed in figure 13 compared to figure 11. We see that autocorrelation is still present. More lagged values are now showing that the hypothesis that there is no autocorrelation can’t be rejected and that the value for negative correlation is on average less negative. Since the values of intervals of 12 still have very high values of autocorrelation indicate that we have not been able filter out all of the seasonal effect. Since the very low R^2 in the regression above this is not very surprising.