Time Series Analysis

Assignment 3

Authors

Elli Georgiou (s223408)
Dimitris Voukatas (s230148)
Dimitris Sousounis (s230118)

April 13, 2023

Part 1 - Univariate models

Question 4.1: Presenting the data

As it can be clearly seen in the average house sales prices in Denmark, an upwards trend can be observed as well as a few outliers around the middle of the time period and towards the end. These two observations indicate that the time series is not stationary and that the variance is non constant. Such particularities will most likely affect the selection of the integrated component of the model and auto-regressive, moving average respectively and potentially the seasonal components too. These issues can be tackled by first order difference and logarithmic or square root transformations.

Similarly, the interest rate demonstrate a downward trend and a non constant variance as well as several zero values towards the end of the time period. Meaning first order difference and a square root transformation will be necessary.

Lastly, the inflation rate illustrate no increasing or decreasing pattern until the end of the time period where a sharp rise is depicted. Regarding the time dependence of the variance, the data fluctuates wildly but the jury is still out. No clear indications are evident from the plotted data about possible transformations. Further analysis is needed.

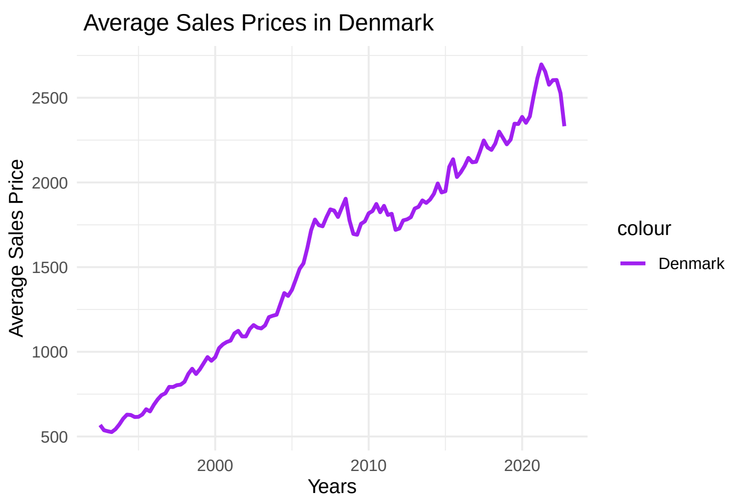


Figure 1: Average sales prices in Denmark

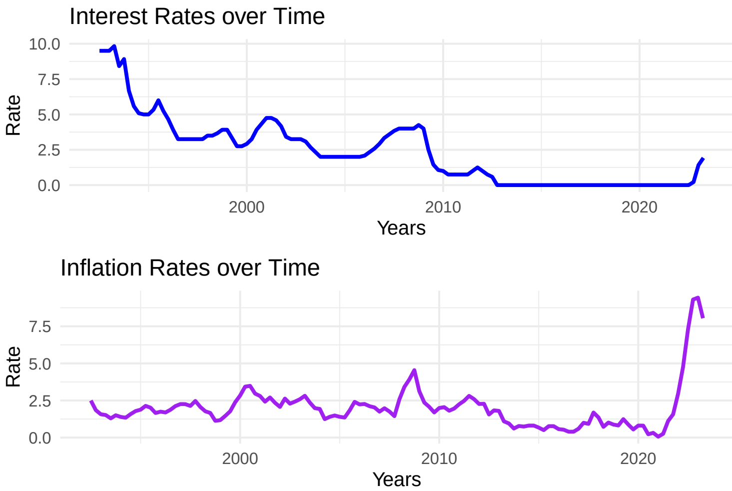
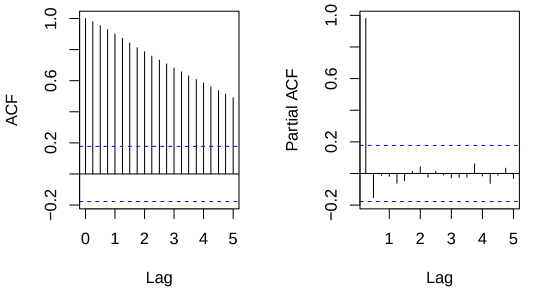


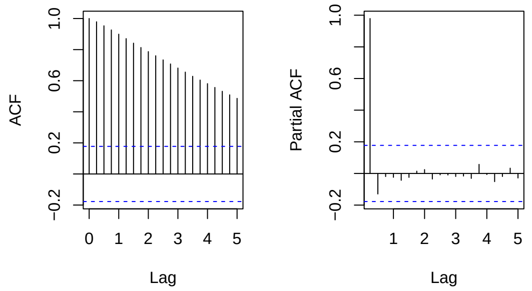
Figure 2: Average interest and inflation rate

Question 4.2: ACF and PACF

In figure 3a and 3b there is gradual decay of significant lags as expected due to the non stationarity of the time series. In figure 3 however, the data is much more interpretable. The ACF and PACF both demonstrate a sharp drop after lag 0 which indicates that autoregressive and moving average components are both 0. There is however a seasonal pattern of significant lags at ACF and the seasonal significant lag is within the C.I. after lag 4 which indicates that the seasonal autoregressive part is 1. So an initial guess for the model is an ARIMA model with order $(0,1,0)(1,0,0)_4$.



(a) ACF-PACF of original Denmark house prices



(b) ACF-PACF of square root transformed Denmark house prices

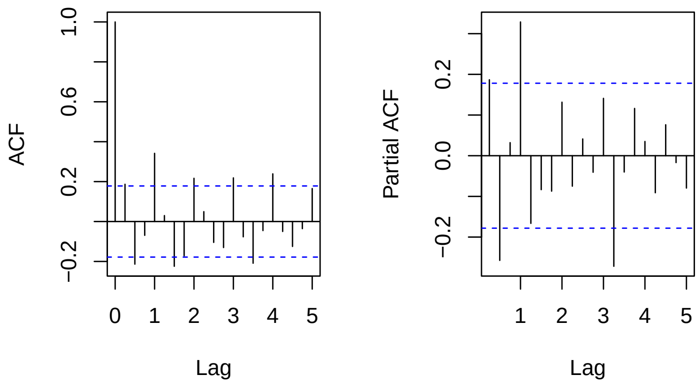


Figure 3: ACF-PACF of first order difference of square root transformed Denmark house sales prices

Figure 4: ACF-PACF plots

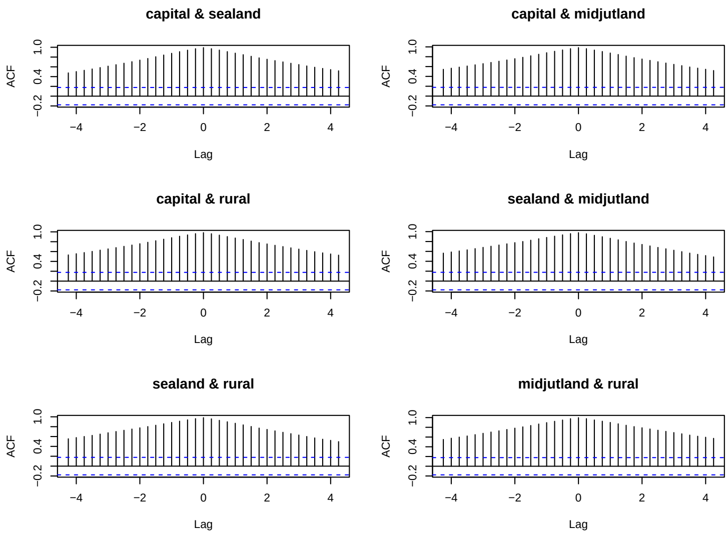
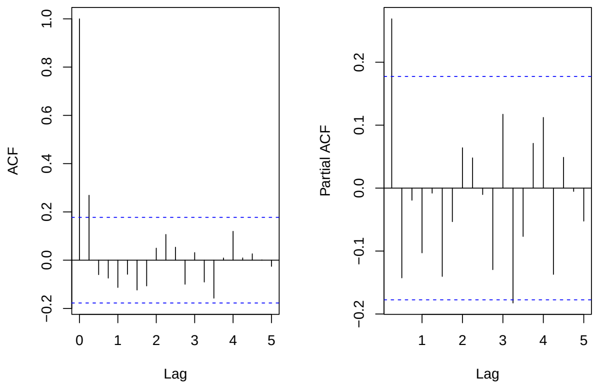


Figure 5: Cross correlation of the 4 regions

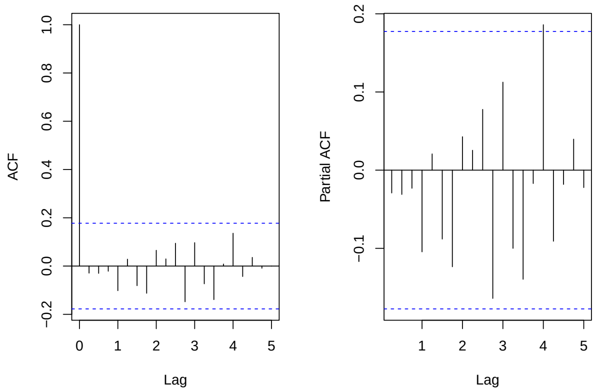
(NOTE: The CCF plot of the original time series of the 4 regions where provided as asked.)

Question 4.3: Univariate model selection

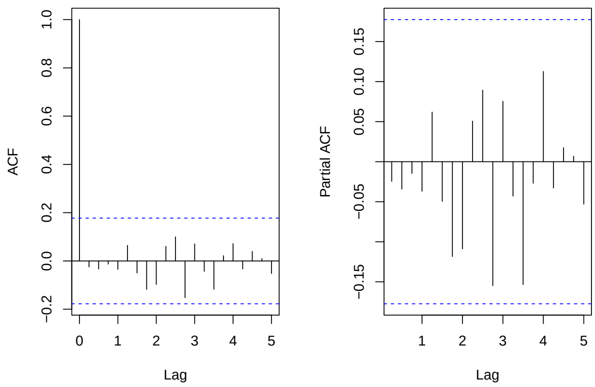
First off, since first order difference was applied the integrated component is 1, d=1. From the ACF and PACF plots of the residuals in figure 6a, a significant lag 1 is observed so maybe a moving average component is needed so q=1. In figure 6b, a seasonal significant lag at lag 4 remains at PACF so maybe we should increase the seasonal autoregressive component to 2, P=2. Now, in figure 6c, there is no correlation in the residuals thus the final model order is $(0,1,1)(2,0,0)_4$.



(a) Caption 1



(b) Caption 2

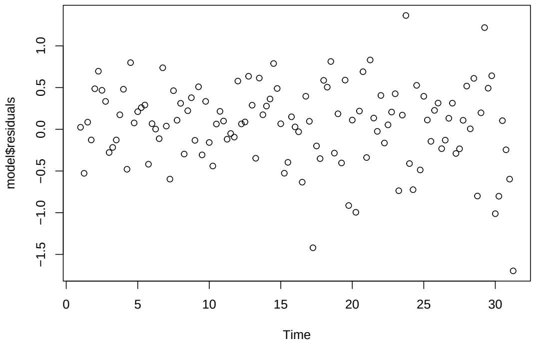


(c) Caption 3

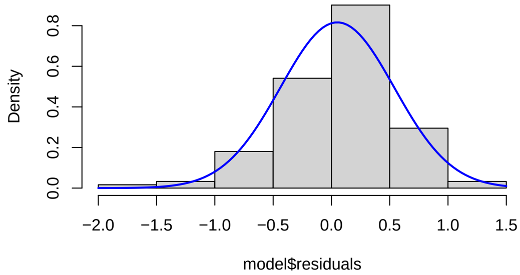
Figure 6: ACF-PACF

Question 4.4: Residual diagnostics

As depicted in figure 6c, there is no correlation in the residuals. Also the p-value of the t.test for the null hypothesis $\mu = 0$ is 0.2561>0.05. The p-value of the Ljung-Box test for testing if there is autocorrelation is also greater than 0.05<0.7833. Lastly, the number of sign changes of the residuals is $63 \in B(122-1, 0.5)$ times which also satisfies the sign test. Thus, from the aforementioned results and figures 7a and 8b, the residuals analysis shows that the residuals are white noise. Furthermore, from figures 7b,8a we can not assume that the residuals are normal distributed with absolute certainty. Although the residuals fit nicely the Q-Q plot, the histogram is a bit skewed.

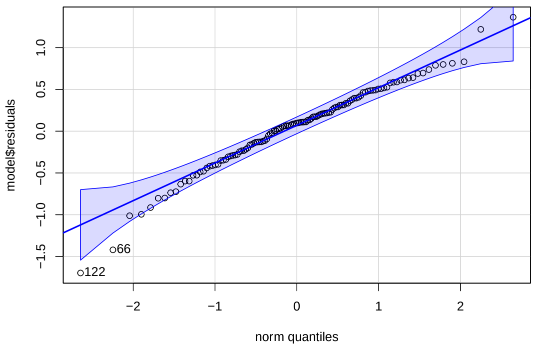


(a) Plot of the residuals

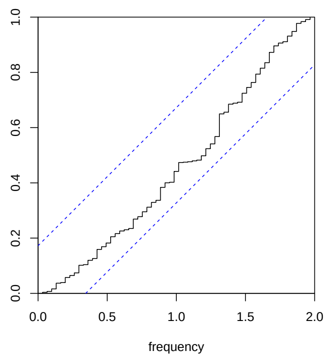


(b) Histogram of the residuals of the final model

Figure 7: Residual plots



(a) Q-Q plot of the residuals of the final model



(b) Cumulative Periodogram

Figure 8: Residual plots

Question 4.5: Forecasting the future house prices - I

The ARIMA model was used to predict real estate values with 95% prediction intervals for the following six quarters. With an expected increase from the first quarter of 2022 to the first quarter of 2024, the predicted values indicate to a moderate rise in housing prices. Furthermore, there is some uncertainty in our model's predictions, as demonstrated by the broad prediction intervals.

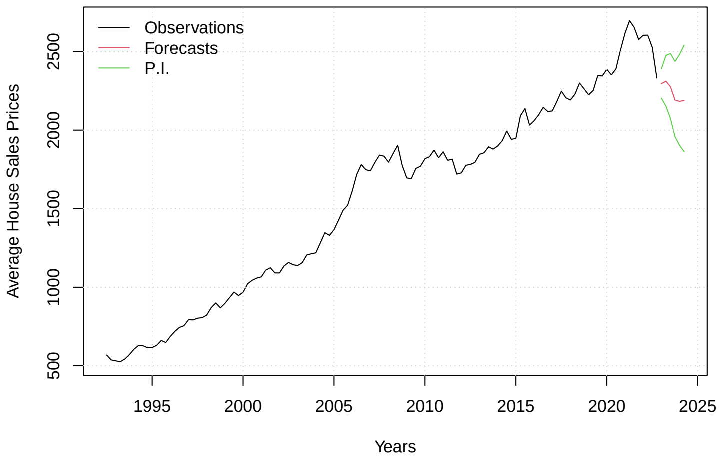


Figure 9: Forecasts of future house prices I-without external inputs

Yearly quarters	Predictions	Lower Bound	Upper Bound
2022Q1	2296.261	2203.780	2390.642
2023Q1	2311.597	2152.729	2476.121
2023Q2	2276.027	2073.768	2487.696
2023Q3	2191.238	1957.627	2438.013
2023Q4	2183.265	1903.394	2482.324
2024Q1	2189.052	1862.953	2541.438

Table 1: Forecasts of future house prices I-without external inputs

Question 4.6: External inputs

First off all, to extend the ARIMA model into an ARIMAX model we have to assume that the external inputs, namely the interest and inflation rates, are independent of the output time series, the average house sales prices in Denmark.

Whether it is worthwhile to keep either inputs in the model depends on a number of reasons like any theoretical basis (which they most likely do from an economics standpoint one would assume), high correlation with other variables, missing values and whether they are stationary or not.

In this case, correlation and non stationarity is not an issue since we can convert the input into white noise either with pre-whitening or apply transformations to achieve stationarity. The interest rate, however, has many zero values with a known pattern which may not affect the model since it does not introduce much variability. On the other hand, figures 10, 11 show that the two inputs are correlated, meaning collinearity is present in the model which makes it hard to come up with reliable results.

For the sake of a thorough analysis, we will examine all three cases: both inputs included, only the interest rate is included and only the inflation rate is included. To produce more accurate results, square root transformations followed by first order difference were applied to make the time series stationary, remove outliers and to scale the regressors the same way as the output.

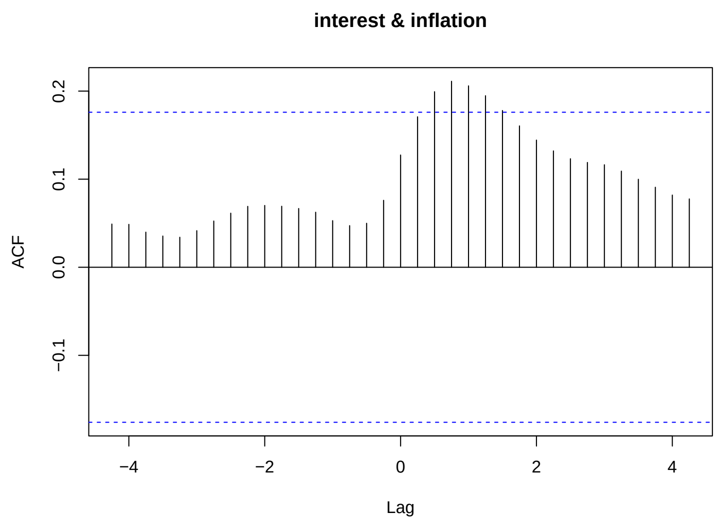


Figure 10: Cross-correlation of interest rate and inflation rate pre transformation

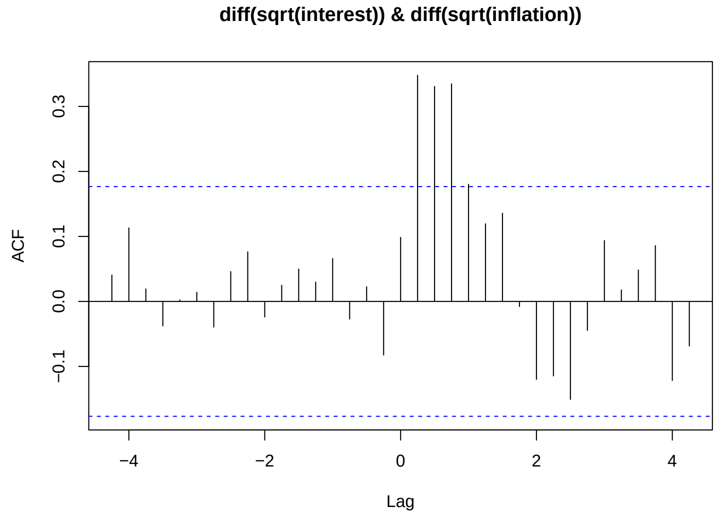


Figure 11: Cross-correlation of interest rate and inflation rate after transformation

Question 4.7: Forecasting the future house prices - II

According to figures 12, 13 and tables 2, 3 there is minuscule difference whether the inflation rate is included in the model or not. It can also be observed from table 4 that the sole inclusion of the inflation rate as an external input produces near same predictions as the ARIMA model (table 1). Further evidence to the significance of the interest rate over the inflation rate (despite the zero values) is their coefficients , -0.8461298 and 0.0004427142 respectively.

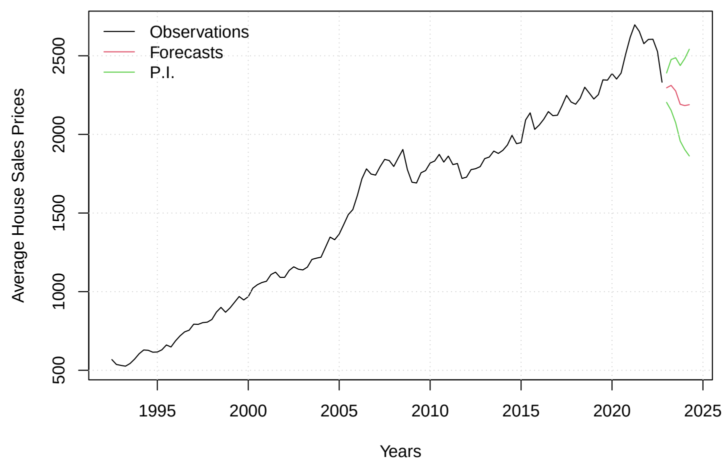


Figure 12: Forecasts of future house prices II - both external inputs with square roots

Yearly quarters	Predictions	Lower Bound	Upper Bound
2022Q1	2255.210	2165.172	2347.081
2023Q1	2252.473	2099.790	2410.514
2023Q2	2215.239	2021.310	2418.049
2023Q3	2143.070	1918.682	2379.864
2023Q4	2140.778	1870.929	2428.800
2024Q1	2145.825	1831.066	2485.536

Table 2: Forecasts of future house prices II-with external inputs with square roots

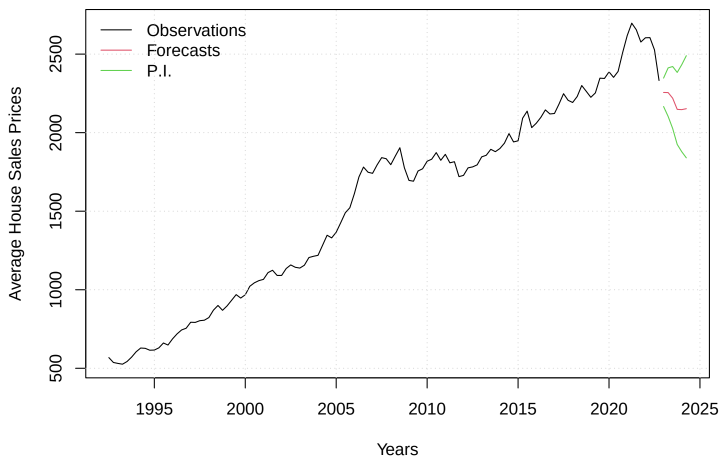


Figure 13: Forecasts of future house prices II - external input sqrt(interest)

Yearly quarters	Predictions	Lower Bound	Upper Bound
2022Q1	2255.909	2166.214	2347.424
2023Q1	2255.179	2103.341	2412.308
2023Q2	2219.438	2026.596	2421.044
2023Q3	2148.150	1925.000	2383.534
2023Q4	2146.693	1878.398	2432.892
2024Q1	2152.542	1839.712	2489.923

Table 3: Forecasts of future house prices II - external input sqrt(interest)

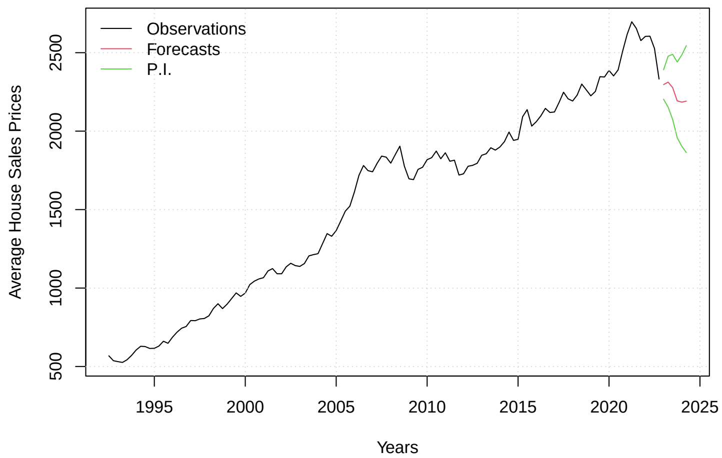


Figure 14: Forecasts of future house prices II - external input sqrt(inflation)

Yearly quarters	Predictions	Lower Bound	Upper Bound
2022Q1	2296.544	2203.667	2391.338
2023Q1	2312.433	2152.937	2477.629
2023Q2	2277.233	2074.177	2489.770
2023Q3	2192.698	1958.163	2440.495
2023Q4	2184.974	1904.013	2485.262
2024Q1	2190.962	1863.624	2544.771

Table 4: Forecasts of future house prices II - external input sqrt(inflation)

It is worth mentioning if the interest rate should be transformed via square root or not. As we can see from figures 15, 16 and tables 5, 6 the predictions are slightly bigger but more importantly the ACF and PACF plots of the non-transformed input display less significant lags as depicted in figures 17, 18. However, the model with the lowest AIC values is the one where the interest rate was square root transformed.

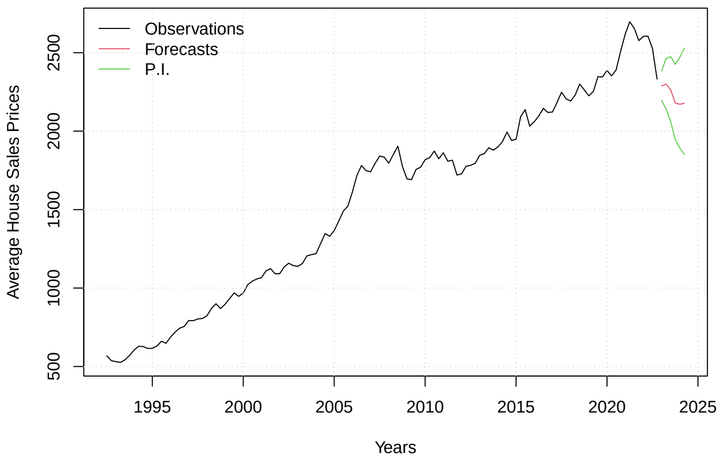


Figure 15: Forecasts of future house prices II - external inputs interest, sqrt(inflation)

Yearly quarters	Predictions	Lower Bound	Upper Bound
2022Q1	2288.477	2195.539	2383.342
2023Q1	2299.017	2140.371	2463.334
2023Q2	2263.510	2061.758	2474.677
2023Q3	2179.399	1946.471	2425.486
2023Q4	2171.528	1892.186	2470.094
2024Q1	2176.961	1851.386	2528.892

Table 5: Forecasts of future house prices II - external inputs interest, sqrt(inflation)

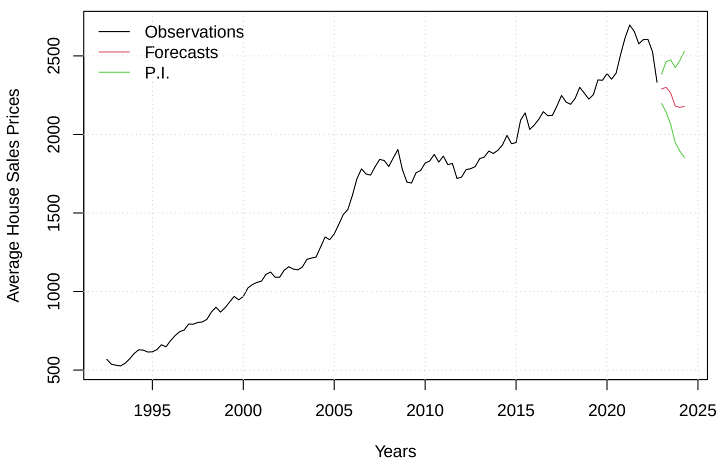


Figure 16: Forecasts of future house prices II - external input interest

Yearly quarters	Predictions	Lower Bound	Upper Bound
2022Q1	2289.183	2196.624	2383.652
2023Q1	2299.876	2141.891	2463.482
2023Q2	2264.506	2063.588	2474.755
2023Q3	2180.481	1948.507	2425.497
2023Q4	2172.704	1894.504	2469.956
2024Q1	2178.214	1853.970	2528.575

Table 6: Forecasts of future house prices II - external input interest

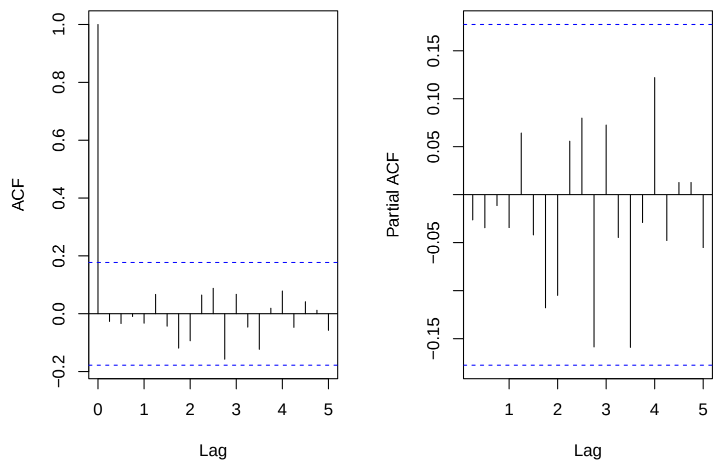


Figure 17: ACF, PACF of interest

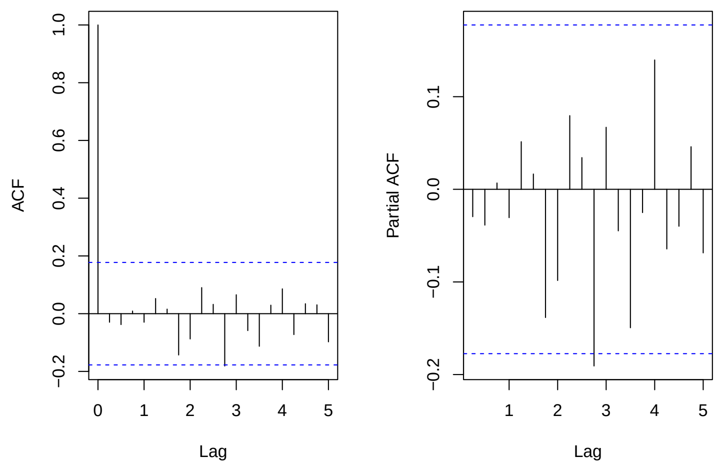


Figure 18: ACF, PACF of sqrt(interest)

Regressors	AIC
sqrt(interest),sqrt(inflation)	178.49
interest,sqrt(inflation)	184.34
sqrt(inflation)	182.84
interest	182.34
sqrt(interest)	176.55

Table 7: AIC values for different regressors

Question 4.8: Conclusions - I

From a purely statistical point of view, we deem the estimated model not trustworthy. If the white noise is not normal distributed that means that the model fails to capture all of the variability of the data. Hence the large width of the prediction intervals. Furthermore, the core assumption for the inclusion of the interest rate in the model is that it is independent of the house sales prices which is not true from an economics point of view (propably).

From an economics point of view, since the interest rate is rising to match the inflation rate, it would be wiser to wait for the sales prices to drop. So maybe wait a couple of years before buying a house until the prices reach a local minimum, around the end of 2023. Any later than that and according to the model the prices are going to rise again.

Part 2 - Multivariate models

Question 4.9: Re-presenting the data

As in the univariate case in part 1, we can observe an increasing trend in all four regions along some outliers at all four regions. More specifically, the data of the capital region exhibits the greatest groth rate and outliers, followed by the data from Sealand and middle Jutland. The outliers from the rural areas are not as evident as the rest of the regions. Thus, for starters, in order for the four time series to achieve stationarity first order difference is needed.

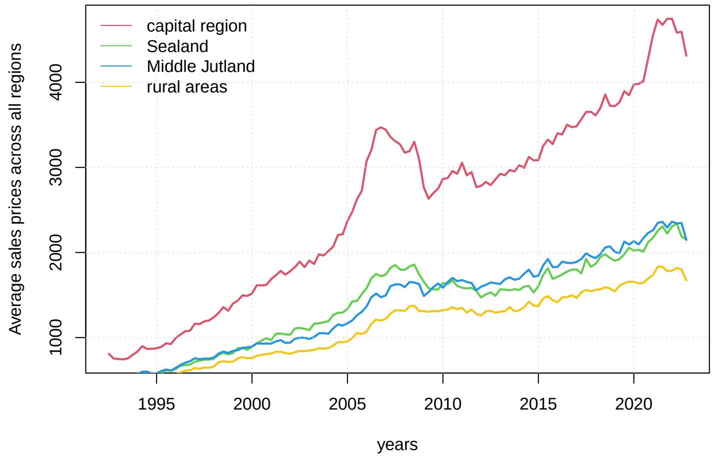


Figure 19: Average sales prices in all 4 regions

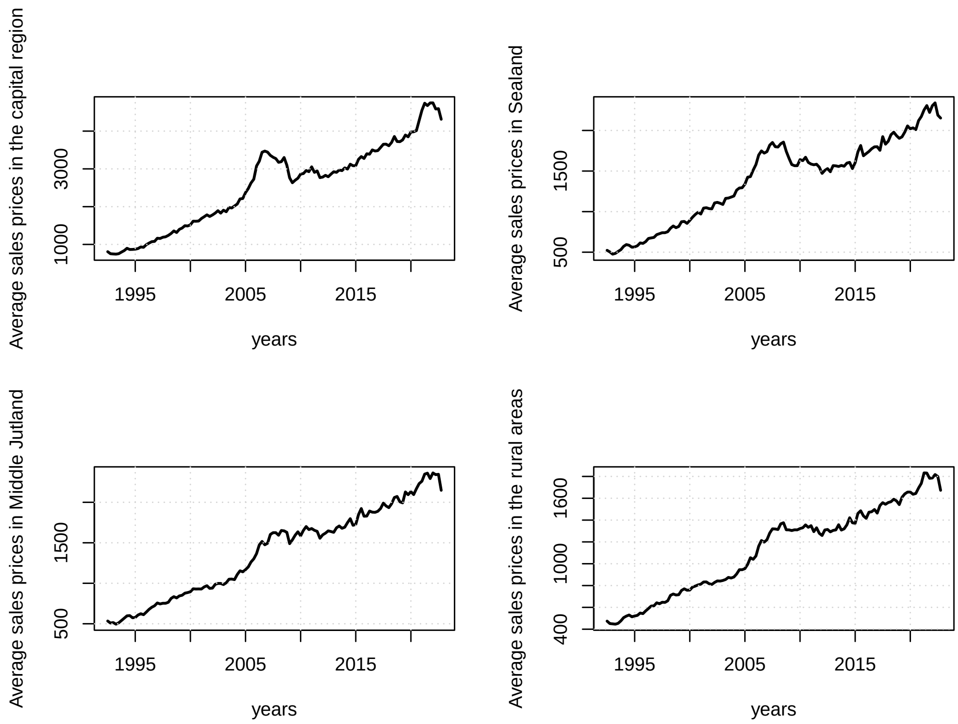


Figure 20

The table 9 shows that only the first two time series achieved stationarity through first order difference. This problem can be cleared by applying logarithm transformation and then first order difference. However, half of the time series are already stationary. If we transform them too we run the risk of inaccurate results by introducing new patterns in the data that was not present originally. Even if we do not transform them we run the same risk as the relationships between the variables may change.

Another way to pass this obstacle is to take the first order difference of the two last time series again in which case stationarity is achieved as shown in table 9. However the residuals of the 4 time series are not normal distributed without the logarithm transformation.

Since we are not sure which transformed variables will make the most accurate predictions, we will examine both cases and let the residuals analysis inform our final verdict.

Region	p-value of 1st O.D	p-value of 2nd O.D
Capital	0.02585	-
Sealand	0.02466	-
Mid. Jutland	0.09106	$<$0.01
Rural Areas	0.1218	$<$0.01
log(Capital)	0.01492	-
log(Sealand)	$<$0.01	-
log(Mid. Jutland)	$<$0.01	-
log(Rural)	$<$0.01	-

Table 8: Results of Augmented Dickey-Fuller test on the vector time series (Rstudio is not able to print the actual p-value)

Question 4.10: ACF and PACF

Case 1: log transformation was applied before first order difference to all time series

Based on the same method of model identification as in part 1, the model structure of each time series was estimated. We will not show all plots but the approach is this: make a first guess of the model order using the ACF and PACF plots, check the ACF and PACF plots of the residuals for any remaining significant lags, alter the model order accordingly and repeat the process until the residuals are white noise and normal distributed. This process includes the regressors as well. Using this method and the figures 21,22 the model orders of 4 regions were found (2,1,0), $(0,1,0)(2,0,0)_4$, $(0,1,0)(2,0,0)_4$, $(0,1,0)(2,0,0)_4$, (2,1,1) and $(2,1,1)(2,0,0)_4$. Although these results do not account for the cross-correlation between the variables, they will help us make an initial guess for the order of the model of the multivariate time series.

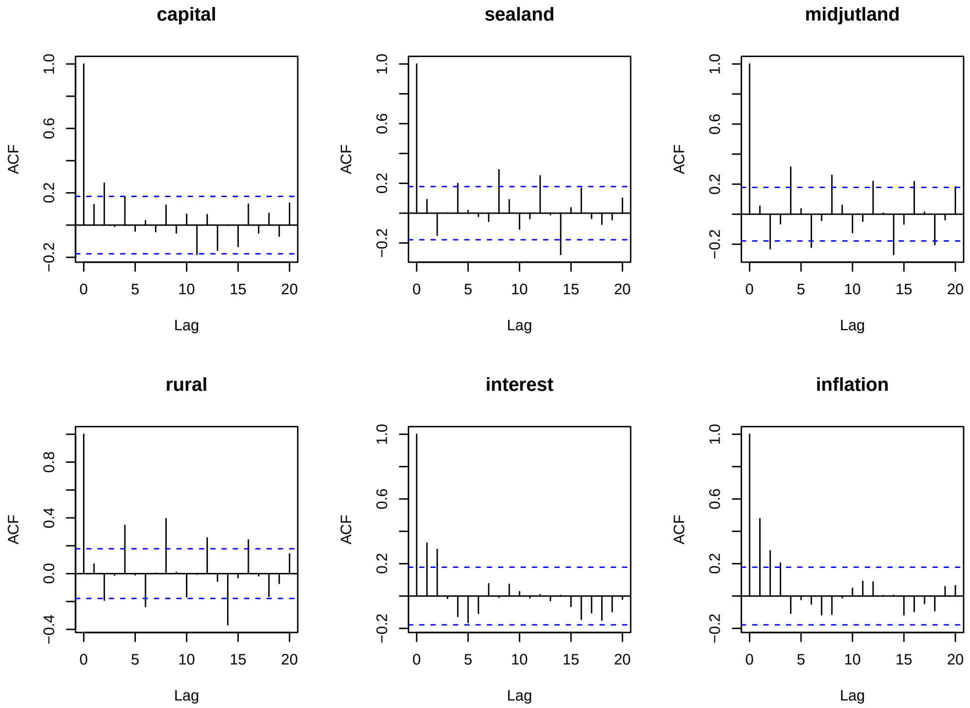


Figure 21: ACF for the prices in the four regions, log

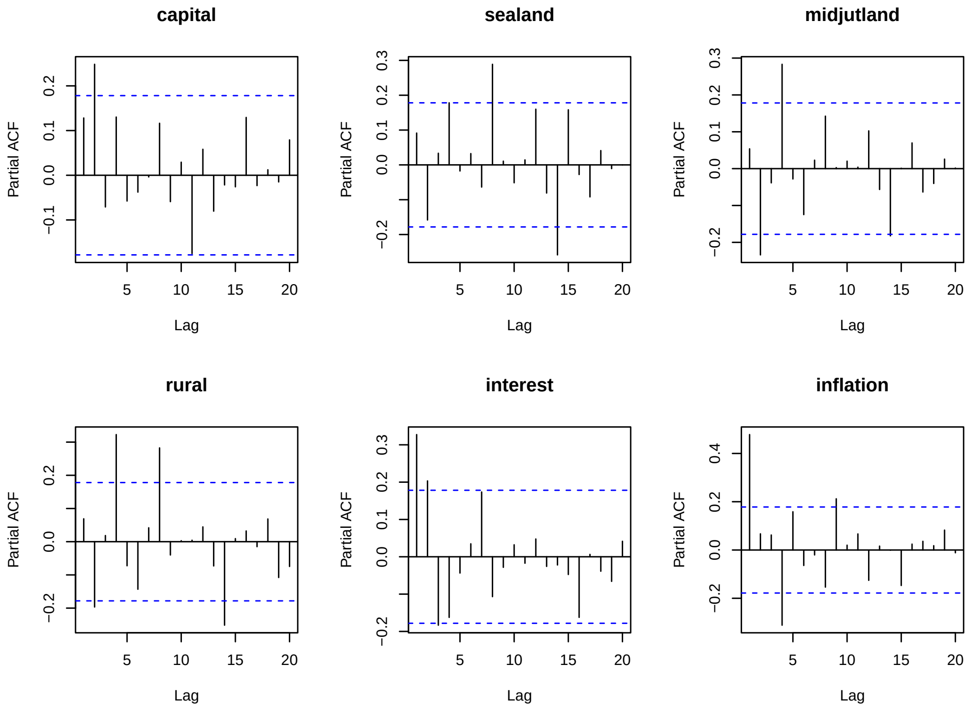


Figure 22: PACF for the prices in the four regions, log

Case 2: first order transformation was applied only to the time series of the capital and sealand region, while the middle jutland and rural areas were transformed via second order difference

In a similar manner as in the previous case by examining the ACF and PACF plots, the order of the individual univariate time series of the regions is (2,1,0), (0,1,0), (3,2,1), $(3,2,0)(1,0,0)_4$, (2,1,1) and $(2,1,1)(2,0,0)_4$ without accounting for cross-correlation. The approach attempts to achieve stationarity for all variables without introducing new patterns into the data.

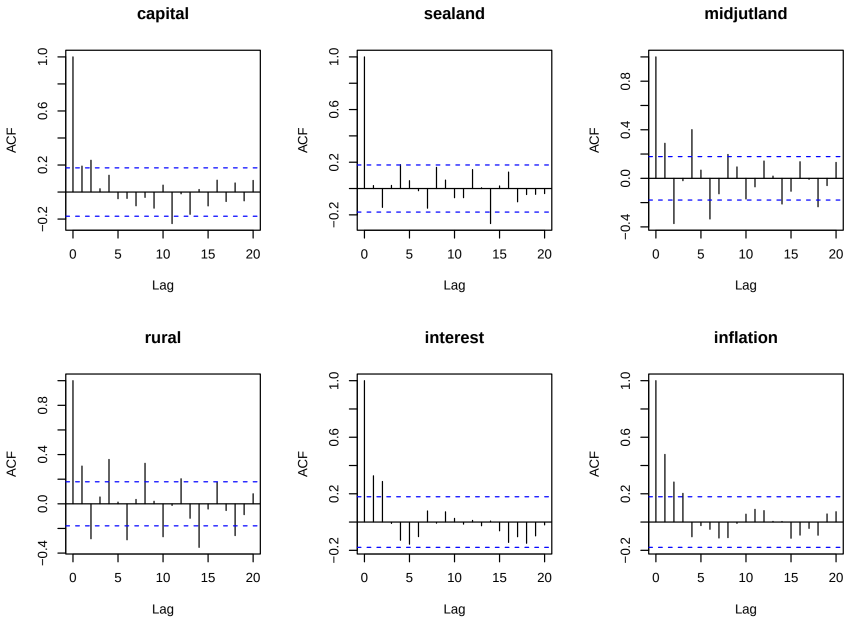


Figure 23: ACF for the prices in the four regions, no log

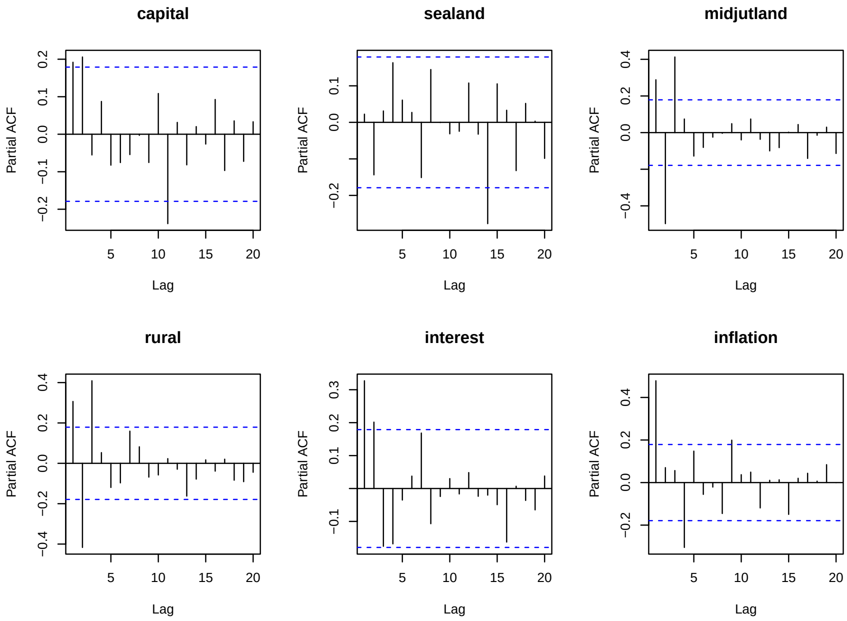


Figure 24: PACF for the prices in the four regions, no log

Question 4.11: Multivariate model selection

With the above results, an estimation of the order of the multivariate model can be made. Specifically, the maximum order of the auto-regressive components is the order of the auto-regressive component of the vector time series. So in case 1 the order of the auto-regressive component could 2, the order of the moving average component 1 and the order of the seasonal auto-regressive component 2. Similarly in case 2, the order of the auto-regressive component is 3, the moving average 1 and the seasonal auto-regressive 1. By plugging these values in the marima function in Rstudio we can see that both models converge nicely. There is however correlation left in both cases, in both the ACF and PACF plots.

Through trial and error, checking for correlation in the residuals and whether they are white noise and normal distributed, we found that no model order fits the data perfectly. Although the ACF plots shows little to no auto-correlation, cross-correlation at lag 0 is present in all pairings. The PACF shows more and bigger significant lags. In an effort to minimise correlation, we raised the order of the seasonal auto-regressive component to 3 since significant lags were observed in a few plots (namely speaking the ACF of middle Jutland and the CCF of capital and rural regions). Although the results are a bit better, the residuals of both models are still correlated. Based on the following figures and tables (and the fact that the residuals of the univariate time series in case 2 are not white noise) after model reduction/identification was applied, we decided the final model is that of case 1, $(2,1,0)(3,0,0)_4$.

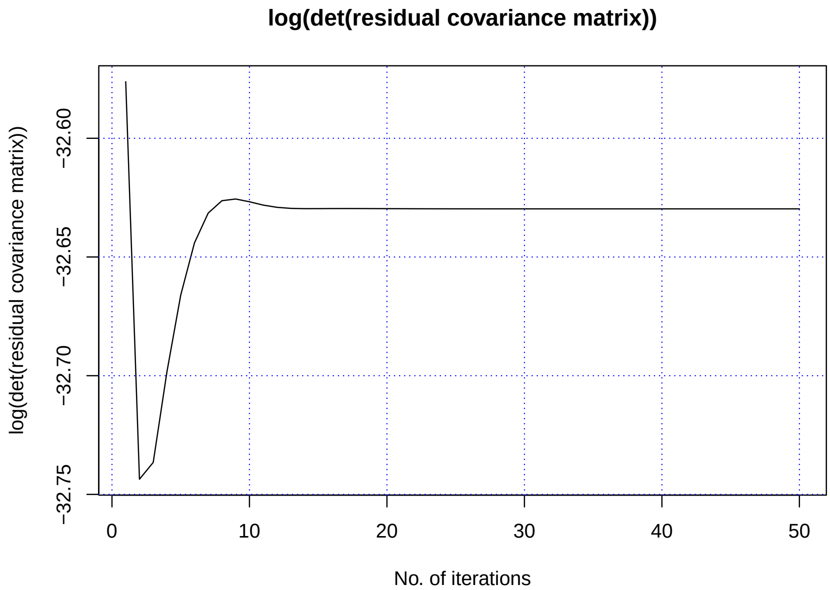


Figure 25: case 1

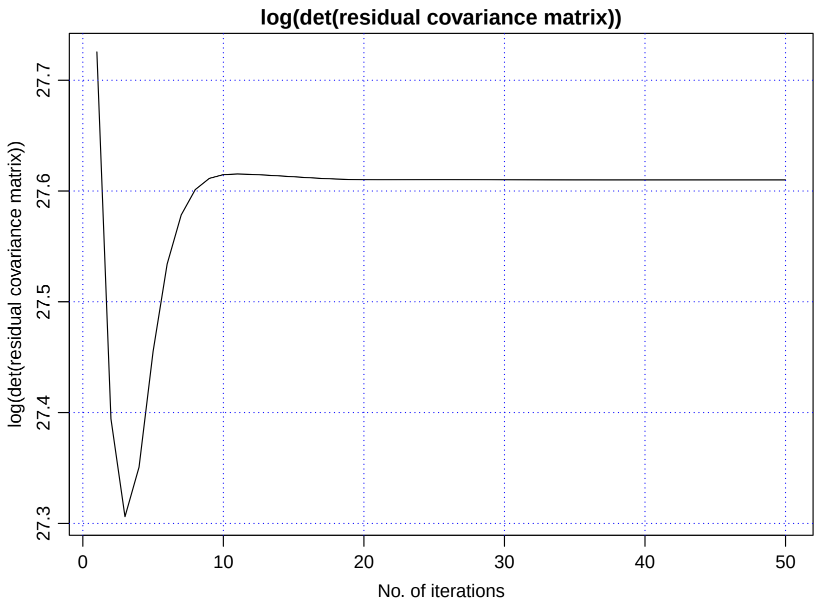


Figure 26: case 2

(NOTE: a 3rd model was tested where we performed square root and then first order difference to all 6 variables but the 4th time series, the rural region, was not stationary and its residuals where not normaly distributed so we decided not to include in the report since the residuals of the multivariate model were also correlated similarly with the other two models)

Another important question is whether the interest rate and inflation rate should be used as simple regression variables or independent variables. As we previously mentioned in the conclusion of part 1, the two variables are not independent off the house market (possibly). Furthermore, while they are complex, they do not have their own dynamics but they can be manipulated (to some extent). Thus we believe that they should be used as regressors.

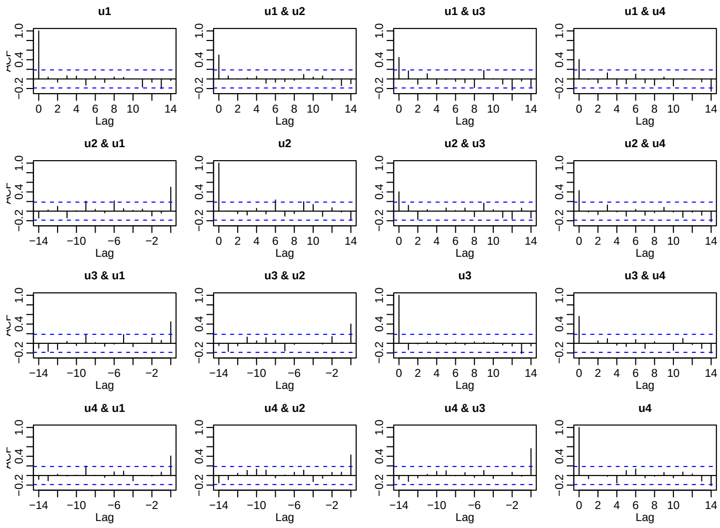


Figure 27: ACF of the residuals of the 4 regions, case 1. Although there are small significant lags, they are common in many plots (i.e. lag 6 at u2 & u1, u2, u4 indicating a strong relationship between the corresponding time series)

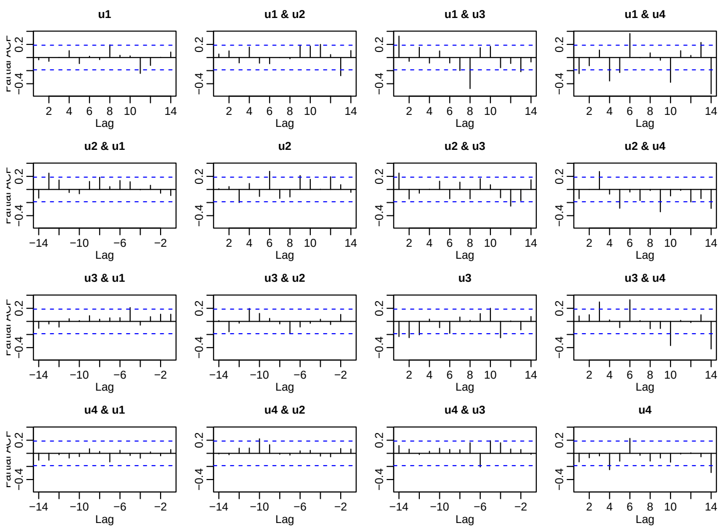


Figure 28: PACF of the residuals of the 4 regions,case 1

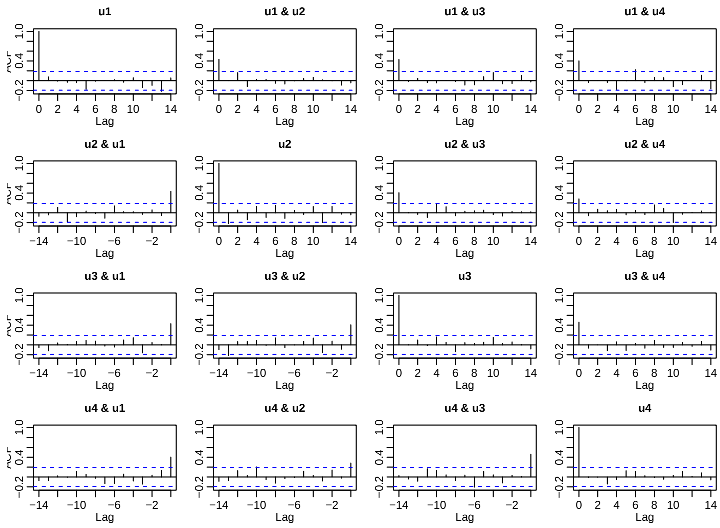


Figure 29: ACF of the residuals of the 4 regions, case 2

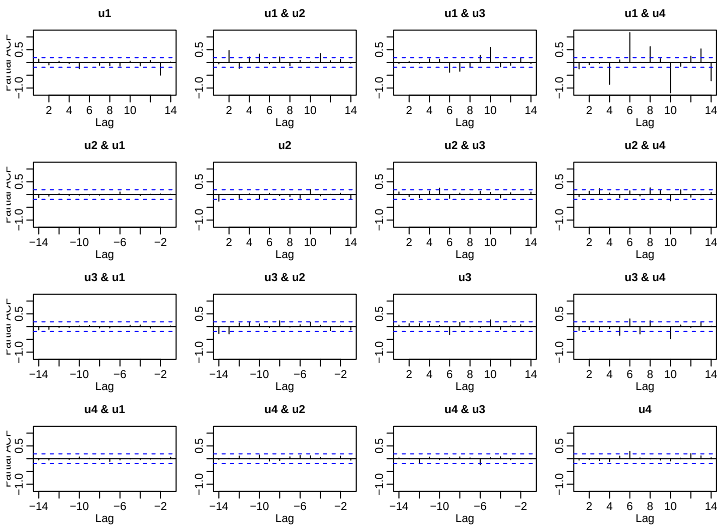


Figure 30: PACF of the residuals of the 4 regions, case 2

Region	t.test case 1	Ljung-Box test case 1	t.test case 2	Ljung-Box test case 2
Capital	0.9765	0.6857	0.9617	0.4092
Sealand	0.6188	0.9755	0.9121	0.02357
Mid. Jutland	0.7709	0.1671	0.9746	0.9819
Rural Areas	0.579	0.4845	0.4352	0.9366

Table 9: P-values of t.test and Ljung-box test for both cases. As we can see the hypotheses that the residuals of the 2nd time series are not independently distributed has been rejected, thus they are correlated.

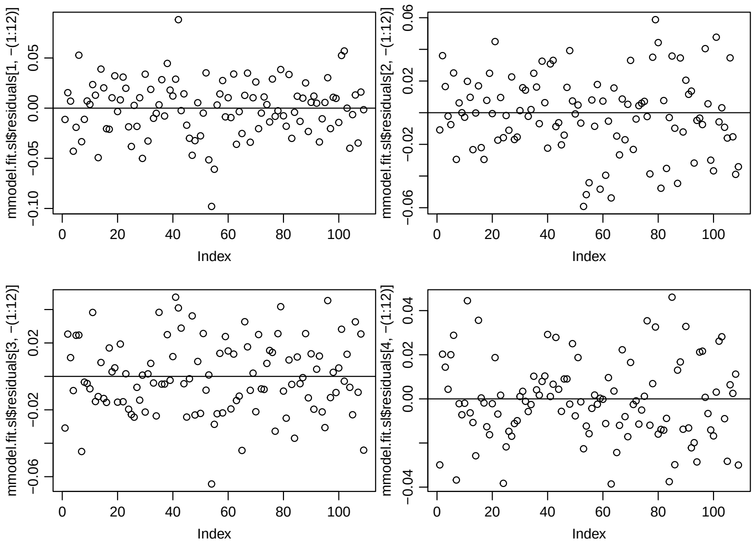


Figure 31: Residuals of the 4 regions, case 1

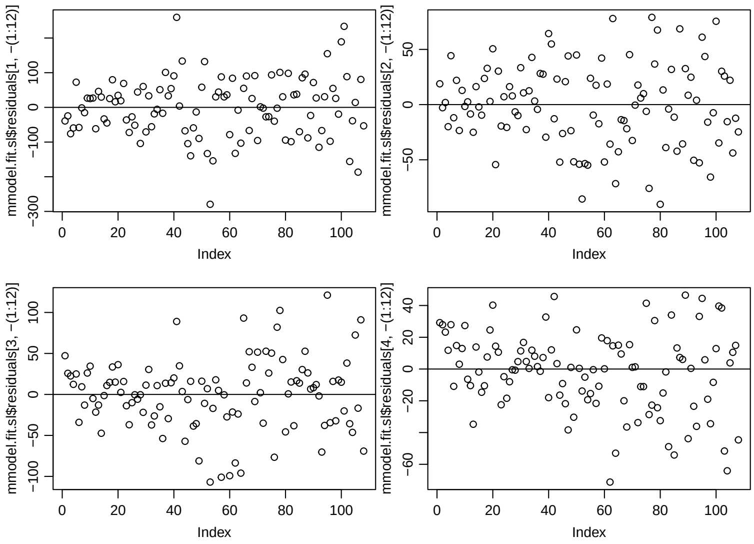


Figure 32: Residuals of the 4 regions, case 2

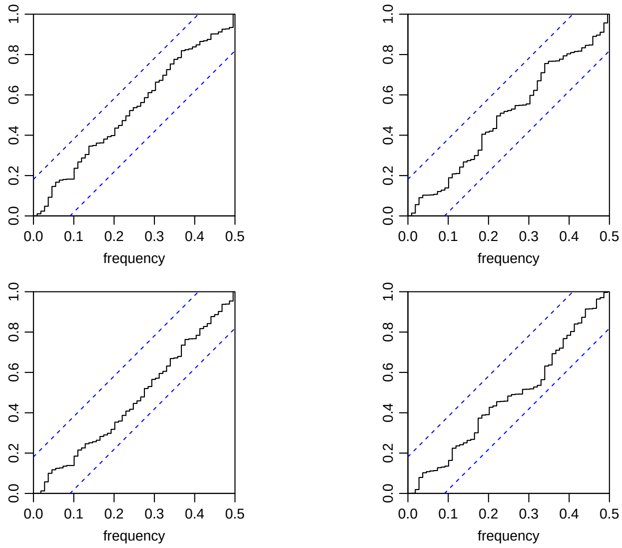


Figure 33: Cumulative periodogram of the 4 regions, case 1

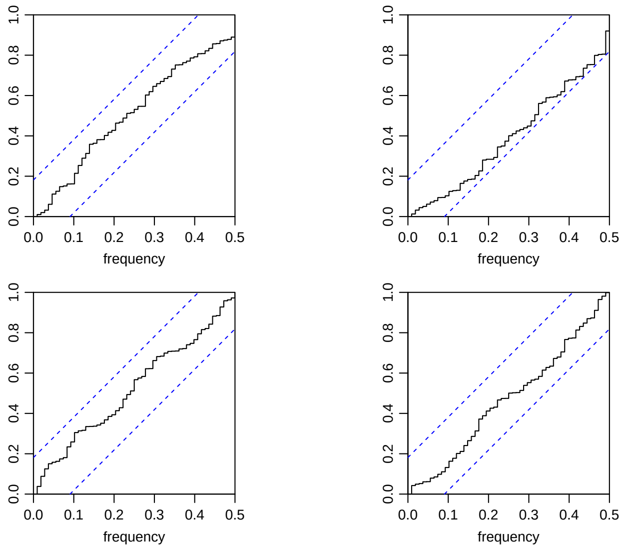


Figure 34: Cumulative periodogram of the 4 regions, case 2

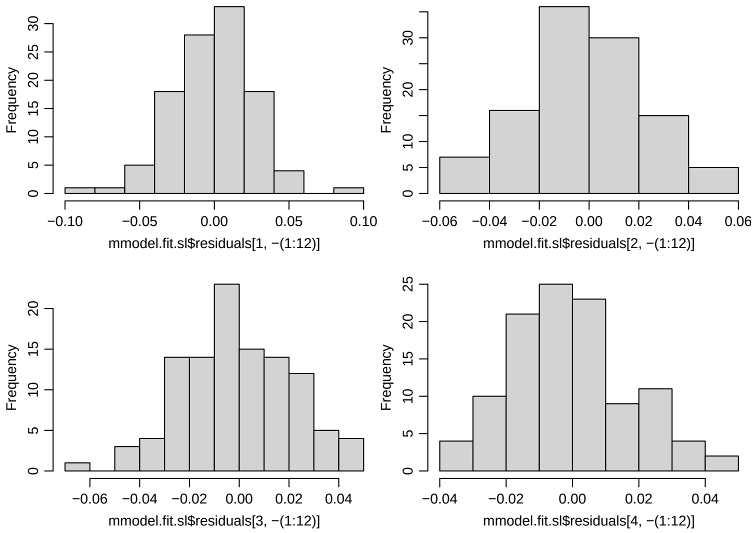


Figure 35: Histogram of the 4 regions, case 1

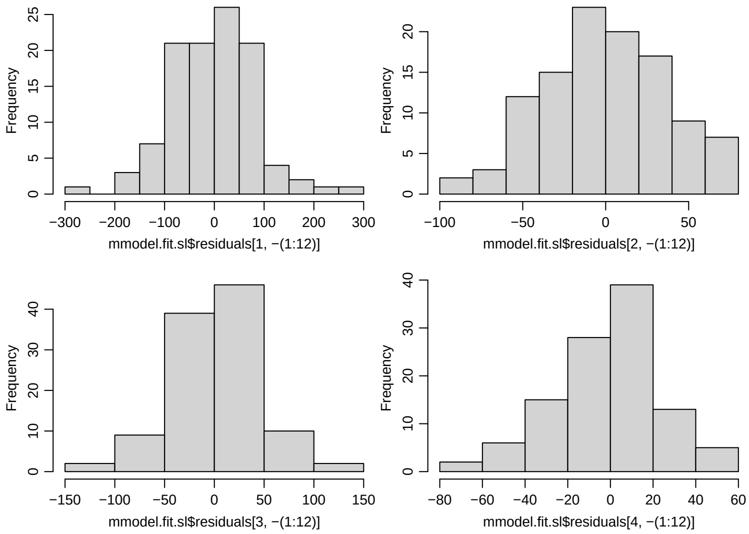


Figure 36: Histogram of the 4 regions, case 2

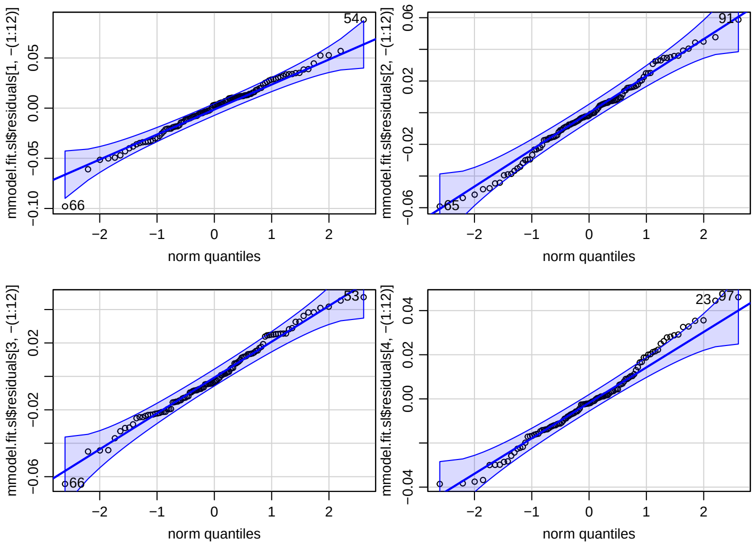


Figure 37: Q-Q plot of the 4 regions, case 1

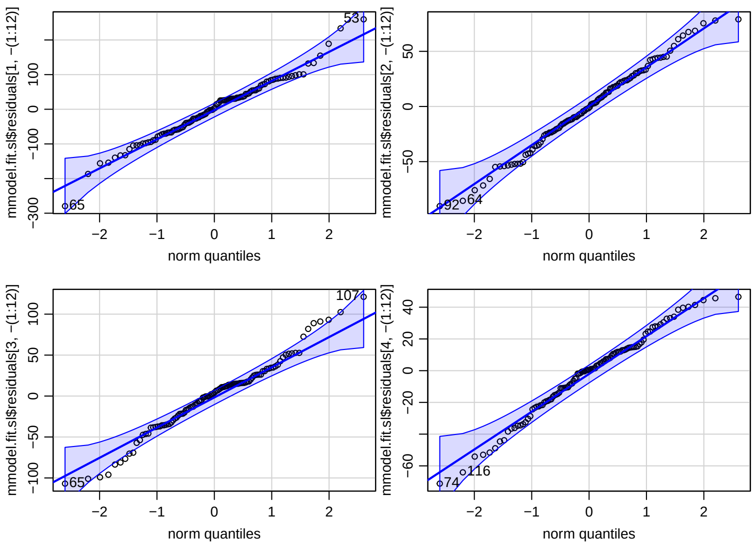


Figure 38: Q-Q plot of the 4 regions, case 2

Question 4.12: Multivariate explanation

A 4-dimensional time series can be written either as a equation where the coefficients are 4-by-4 matrices or 4 equations. The form of the model is the following:

$$
\begin{aligned}
&y_{1,t} - y_{1,t-1} - 0.3162997y_{1,t-2} + 0.3162997y_{1,t-3} \\
&\quad - 0.2220139y_{2,t-1} + 0.22201388y_{2,t-2} - 0.2373350y_{2,t-4} + 0.2373350y_{2,t-5} \\
&\quad - 0.2380360y_{4,t-8} + 0.2380360y_{4,t-9} \\
&\quad + 0.03255662x_{1,t-1} - 0.03255662x_{1,t-2} + 0.01331020x_{2,t-1} - 0.01331020x_{2,t-2} = \epsilon_{1,t}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
&y_{2,t} - 0.6039153y_{1,t-1} + 0.2826180y_{1,t-2} + 0.3212973y_{1,t-3} \\
&\quad - 0.6975826y_{2,t-1} - 0.04733322y_{2,t-2} - 0.2550842y_{2,t-3} - 0.2011211y_{2,t-8} + 0.2011211y_{2,t-9} \\
&\quad + 0.31241965y_{4,t-2} - 0.3124196y_{4,t-3} - 0.347127y_{4,t-12} + 0.347127y_{4,t-13} = \epsilon_{2,t}
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
&y_{3,t} - 0.2255328y_{1,t-4} + 0.2255328y_{1,t-5} \\
&\quad - 0.2364705y_{2,t-1} + 0.23647046y_{2,t-2} - 0.1430684y_{2,t-12} + 0.1430684y_{2,t-13} \\
&\quad - y_{3,t-1} \\
&\quad + 0.25002007y_{4,t-2} - 0.2500201y_{4,t-3} - 0.3117709y_{4,t-8} + 0.3117709y_{4,t-9} \\
&\quad + 0.02185611x_{1,t-1} - 0.02185611x_{1,t-2} + 0.01754350x_{2,t-1} - 0.01754350x_{2,t-2} = \epsilon_{3,t}
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
&y_{4,t} - 0.1955486y_{1,t-2} + 0.1955486y_{1,t-3} \\
&\quad - 0.1652308y_{2,t-1} + 0.16523078y_{2,t-2} - 0.1671005y_{2,t-4} + 0.1671005y_{4,t-5} \\
&\quad - 0.329006y_{3,t-1} + 0.329006y_{3,t-2} \\
&\quad - 0.6033709y_{4,t-1} - 0.06149601y_{4,t-2} - 0.3351331y_{4,t-3} - 0.3740119y_{4,t-8} + 0.3740119y_{4,t-9} \\
&\quad + 0.01303342x_{2,t-1} - 0.01303342x_{2,t-2} = \epsilon_{4,t}
\end{aligned} \tag{4}
$$

The MARIMAX model looks similar with 4 univariate models with the exception that the observation of each variable at the current time unit is dependent of the past observations of the other 3 variables. The parameter estimates dictate the influence of past observations on the current ones. The closer the absolute value of the parameter is to 1, the more influential it is and of course the sign defines whether the relationship between the past and present value is positive or negative (a positive parameter increases the current value while a negative decreases it).

It is noticeable that most parameters of the same univariate time series are opposite to the ones in the previous time period. That means that while the predictor variable has the same impact, the relationship between predictor and response variable has changed. This could be either due to the influence of the interest and inflation rates or due to the fact that residual correlation is present in the model.

Based on the above equations we can see that the sales prices in the capital region mostly affect (greatly and negatively) themselves since the sum of the estimated parameters is very close to 0 with exception of equation (1). Furthermore, its past values within the year impact all other regions but not sealand. The prices there are affected by older prices in the capital.

The sales prices in the Sealand region are also negatively affected by themselves by past observations up to 3 time periods back. Similarly to the capital region, the sales prices at sealand region within the year impact all other regions with the exception the sales prices in middle Jutland where the sales prices are affected by the prices in sealand 4 years back.

One the other hand, the prices in middle Jutland only affect themselves and the prices in the rural areas and only the recent ones within the year.

Lastly, the prices in the rural areas affect all the regions. The prices in the capital region are affected by the prices in the rural areas 2 to 3 years back, the same goes for the sealand prices with the inclusion of the more recent prices and the oldest (1 and 4 years old prices). The middle Jutland and rural prices are affected similarly from 1 and 3 year old sales.

Question 4.13: Forecasting the future house prices - III

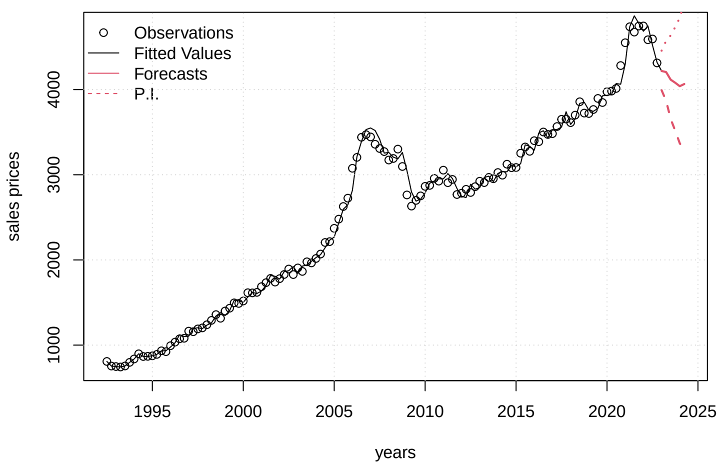


Figure 39: Forecasts of average housing sales prices in the capital region

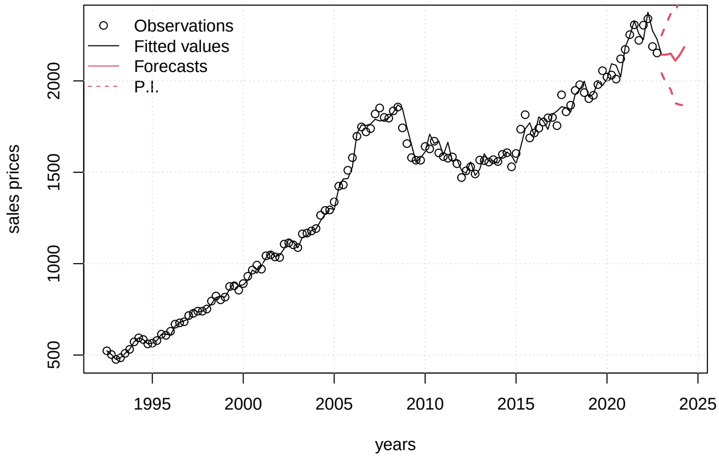


Figure 40: Forecasts of average housing sales prices in the Sealand region

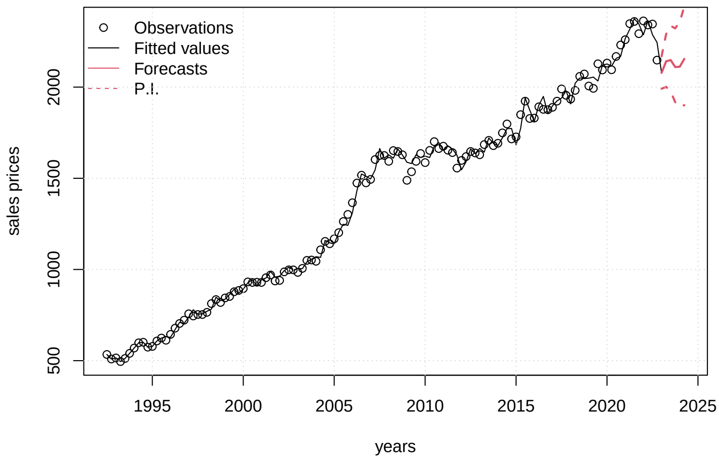


Figure 41: Forecasts of average housing sales prices in the middle Jutland region

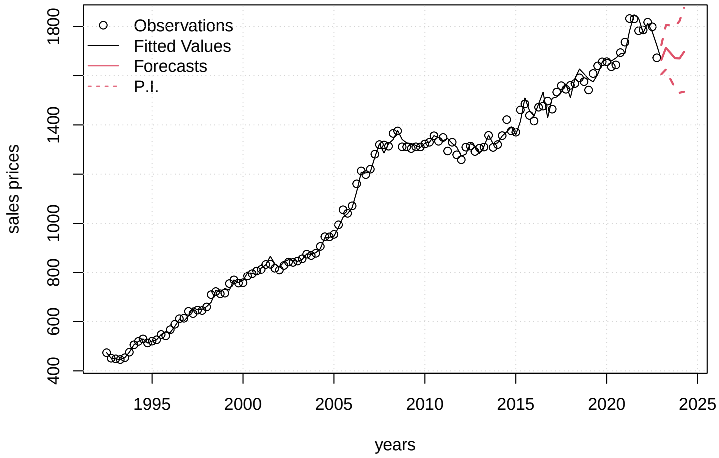
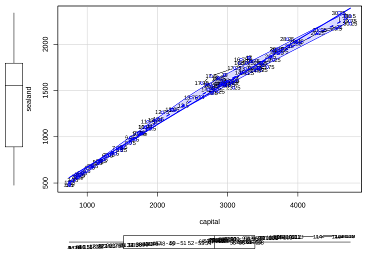


Figure 42: Forecasts of average housing sales prices in the rural areas

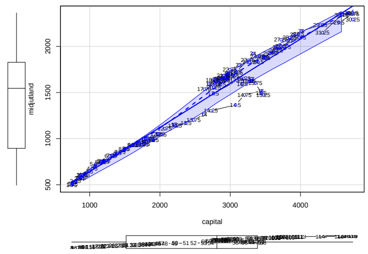
Question 4.8: Conclusions - II

To deem whether the predictions of the model are accurate several assumptions were made. First off, the individuals time series (all 6) have to be stationary, in this case they are. The second and third assumptions (no auto-correlation and no cross-correlation of the residuals) however are not kept as we can see from figures 27, 28. We also need to assume that the system that generates the stochastic process is linear but the following scatterplots show that the non transformed variables are almost linear with each other but no longer stationary and the transformed variables are stationary but not linear. Furthermore, while the Q-Q plots in figure 37 looks adequate, the histograms in 35 do not. While correlation is present between the external regressors, the low significant lags at figures 10, 11 show that collinearity is probably not an issue. Lastly, the figure 31 indicates that the mean value of the residuals is not 0, thus it is not white noise despite the results of the t.test 9.

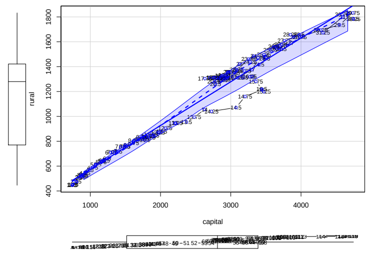
Overall, the model fails to completely capture all the data, as is evident from the figures 39, 40, 41, 42. However, if it were to be trusted and someone wants to buy a house in Denmark, we would advise them, in the case of the capital region, to wait a couple of years before buying a house, as prices are expected to fall and reach a local minimum after a few quarters. But since the prices in the rest of the regions is lower and increasing, maybe they could buy a house there if they are in a hurry.



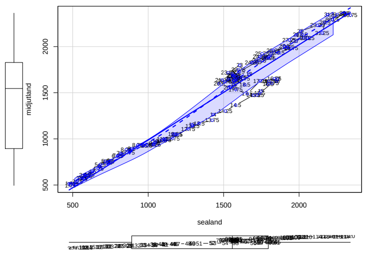
(a) Capital-Sealand



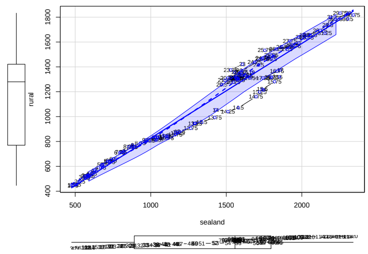
(b) Capital-Middle Jutland



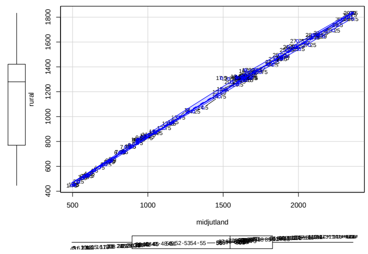
(c) Capital-Rural



(d) Sealand-MIddle Jutland

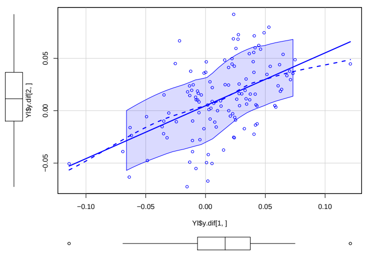


(e) Sealand-Rural

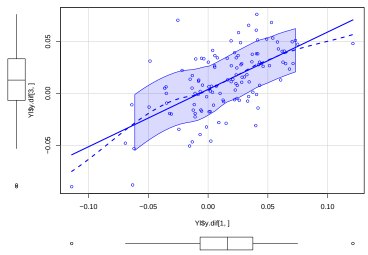


(f) Middle Jutland-Rural

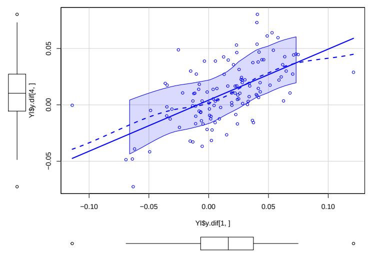
Figure 43: Scatter plots of pre transformed time series



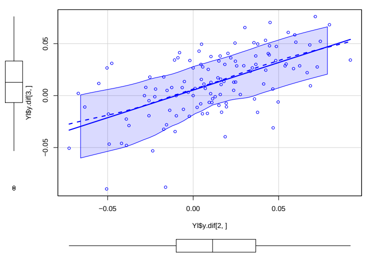
(a) Capital-Sealand



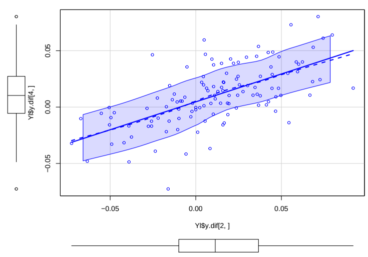
(b) Capital-Middle Jutland



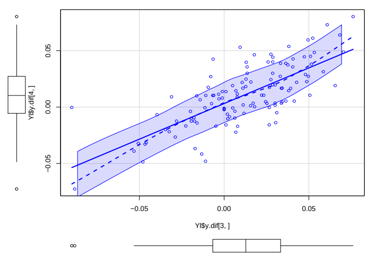
(c) Capital-Rural



(d) Sealand-MIddle Jutland



(e) Sealand-Rural



(f) Middle Jutland-Rural

Figure 44: Scatter plots of transformed time series

R Notebook

```r
rm(list=ls())
set.seed(123)
library(forecast)
library(tseries)
library(car)
library(marima)
library(ggplot2)
library(gridExtra)

data <- read.csv("A3Data.csv")
ts <- ts(data, frequency = 4)

denmark <- na.omit(ts[,2])
capital <- na.omit(ts[,3])
sealand <- na.omit(ts[,4])
midjutland <- na.omit(ts[,5])
rural <- na.omit(ts[,6])
interest <- na.omit(ts[,7])
inflation <- na.omit(ts[,8])

years <- seq(from=1992.5, to=2024.25, by=0.25)

################################### Part 1 ###################################

##### Presenting the Data #####

ggplot(df, aes(x = seq(1992.5, to=2022.75, by=0.25))) +
  geom_line(aes(y = Denmark, color = "Denmark"), size = 1) +
  scale_color_manual(values = c( "Denmark" = "purple")) +
  labs(x = "Years", y = "Average Sales Price", title = " Average Sales Prices in Denmark") +
  theme_minimal()

df <- data.frame(
  Quarter = time(capital),
  Capital = as.numeric(capital),
  Sealand = as.numeric(sealand),
  Midjutland = as.numeric(midjutland),
  Rural = as.numeric(rural),
  Denmark=as.numeric(denmark)
)

# Plotting time series for interest rates
df_interest <- data.frame(
  Quarter = time(interest),
  Interest = as.numeric(interest)
)

interest_plot <- ggplot(df_interest, aes(x = seq(from=1992.5, to=2023.25, by=0.25), y = Interest)) +
```

```r
  geom_line(color = "blue", size = 1) +
  labs(x = "Years", y = "Rate", title = "Interest Rates over Time") +
  theme_minimal()

# Plotting time series for inflation rates
df_inflation <- data.frame(
  Quarter = time(inflation),
  Inflation = as.numeric(inflation)
)

inflation_plot <- ggplot(df_inflation, aes(x = seq(from=1992.5, to=2023.25, by=0.25), y
↪  = Inflation)) +
  geom_line(color = "purple", size = 1) +
  labs(x = "Years", y = "Rate", title = "Inflation Rates over Time") +
  theme_minimal()

# Arrange the two plots side by side
grid.arrange(interest_plot, inflation_plot, nrow=2)



##### ACF & PACF & CCF #####

### Guesstimate Denmark ###

# Transformations #

ds <- sqrt(denmark)
adf.test(ds)

dsf1 <- diff(ds)
adf.test(dsf1) # stationary

# Plots #

plot(denmark)
plot(ds)
plot(dsf1)

par(mfrow=c(1,2))
acf(denmark) # as expected, first order difference is needed thus d=1
pacf(denmark)

par(mfrow=c(1,2))
acf(ds)
pacf(ds)

par(mfcol=c(1,2))
acf(dsf1)
pacf(dsf1)
# explanations follows a similar method to Rune's code
# (0,1,0)(0,0,0) remove seasonal significant lags
# (0,1,0)(1,0,0) lag 1 is significant
```

```r
# (0,1,1)(1,0,0) some seasonal correlation still exists
# (0,1,1)(2,0,0) *thumbs up*

### Cross-Covariances ###

par(mfrow = c(3,2))
ccf(capital, sealand)
ccf(capital, midjutland)
ccf(capital, rural)
ccf(sealand, midjutland)
ccf(sealand, rural)
ccf(midjutland, rural)

##### Find a suitable (and SIMPLE) ARIMA model #####

model <- Arima(ds, order = c(0,1,0), seasonal = list(order =c(1,0,0),
                                                     period = 4))
par(mfrow=c(1,2))
acf(model$residuals, main="")
pacf(model$residuals, main="")
tsdiag(model)

model <- Arima(ds, order = c(0,1,1), seasonal = list(order =c(1,0,0),
                                                     period = 4))

par(mfrow=c(1,2))
acf(model$residuals, main="")
pacf(model$residuals, main="")
tsdiag(model)

model <- Arima(ds, order = c(0,1,1), seasonal = list(order =c(2,0,0),
                                                     period = 4))

par(mfrow=c(1,2))
acf(model$residuals, main="")
pacf(model$residuals, main="")
tsdiag(model)

##### Residuals Analysis #####

## check if the noise is white ##

# plot #
par(mfrow=c(1,1))
plot(model$residuals, type = "p")
acf(model$residuals)
pacf(model$residuals)
# check if the mean value is 0 #
t.test(model$residuals, mu=0)
# check if there is no autocorrelation #
Box.test(model$residuals, type="Ljung-Box")
# check the sign change #
rbinom(10, length(model$residuals)-1, 0.5)
```

```
sum(diff(sign(model$residuals)) != 0)
# Test in the cumulated periodogram
cpgram(model$residuals, main="")
# check if it is normal distributed #
hist(model$residuals, main="")
qqPlot(model$residuals)

##### Forecasting future house prices 1 #####

predictions <- forecast(model, h=6, level = c(0.95))

plot(seq(from=1992.5,to=2024.25,by=0.25),ts[,2], type = "l", xlab = "Years",
     ylab = "Average House Sales Prices", col=1)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions$mean)^2, col=2)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions$lower)^2, col=3)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions$upper)^2, col=3)
legend("topleft",c("Observations", "Forecasts","P.I."), col=c(1,2,3),lty=1,
       bty='n')
grid()

##### External Inputs #####

xreg <- cbind(interest[1:(length(interest)-2)],
              inflation[1:(length(inflation)-2)])

plot(xreg[,1], type = "l")
acf(xreg[,1])
plot(diff(xreg[,1]), type = "l")
acf(diff(xreg[,1]))
plot(sqrt(xreg[,1]), type = "l")
acf(sqrt(xreg[,1]))
plot(diff(sqrt(xreg[,1])), type = "l")
acf(diff(sqrt(xreg[,1])))
adf.test(xreg[,1])
adf.test(diff(xreg[,1]))
adf.test(sqrt(xreg[,1]))
adf.test(diff(sqrt(xreg[,1])))
model_i1 <- arima(sqrt(xreg[,1]), order = c(2,1,2))
acf(model_i1$residuals)
pacf(model_i1$residuals)
hist(model_i1$residuals)
qqPlot(model_i1$residuals)

plot(xreg[,2], type = "l")
acf(xreg[,2])
plot(diff(xreg[,2]), type = "l")
acf(diff(xreg[,2]))
plot(sqrt(xreg[,2]), type = "l")
acf(sqrt(xreg[,2]))
plot(diff(sqrt(xreg[,2])), type = "l")
acf(diff(sqrt(xreg[,2])))
plot(log(xreg[,2]), type = "l")
acf(log(xreg[,2]))
```

```r
plot(diff(log(xreg[,2])))
acf(log(xreg[,2]))
adf.test(xreg[,2])
adf.test(diff(xreg[,2]))
adf.test(sqrt(xreg[,2]))
adf.test(diff(sqrt(xreg[,2])))
adf.test(log(xreg[,2]))
adf.test(diff(log(xreg[,2])))
model_i2 <- arima(sqrt(xreg[,2]), order = c(2,1,0), seasonal = list(order =c(2,0,0),
                                                                  period = 4))
acf(model_i2$residuals)
pacf(model_i2$residuals)
hist(model_i2$residuals)
qqPlot(model_i2$residuals)

par(mfrow=c(1,3))
ccf(denmark, interest)
ccf(denmark, inflation)
ccf(interest, inflation)

par(mfrow=c(1,3))
ccf(diff(sqrt(denmark)), diff(sqrt(interest)))
ccf(diff(sqrt(denmark)), diff(sqrt(inflation)))
ccf(diff(sqrt(interest)), diff(sqrt(inflation)))
par(mfrow=c(1,1))

# Only inflation has been square root transformed
xreg <- cbind(xreg[,1],sqrt(xreg[,2]))
model_x <- Arima(ds, order = c(0,1,1), seasonal = c(2,0,0), xreg = xreg)
summary(model_x)
tsdiag(model_x)

# both inputs have been square root transformed
xreg1 <- cbind(sqrt(xreg[,1]),sqrt(xreg[,2]))
model_xs <- Arima(ds, order = c(0,1,1), seasonal = c(2,0,0), xreg = xreg1)
summary(model_xs)
tsdiag(model_xs)

# only interest was included in the model WITHOUT square root transform
model_x1 <- Arima(ds, order = c(0,1,1), seasonal = c(2,0,0), xreg = xreg[,1])
summary(model_x1)
tsdiag(model_x1)

# only interest was included in the model WITH square root transform
model_x1s <- Arima(ds, order = c(0,1,1), seasonal = c(2,0,0), xreg = sqrt(xreg[,1]))
summary(model_x1s)
tsdiag(model_x1s)

# only inflation was included in the model with root transform
model_x2 <- Arima(ds, order = c(0,1,1), seasonal = c(2,0,0), xreg = xreg[,2])
summary(model_x2)
tsdiag(model_x2)
```

```r
par(mfrow=c(2,2)) # Only inflation has been square root transformed
acf(model_x$residuals)
pacf(model_x$residuals)
hist(model_x$residuals)
qqPlot(model_x$residuals)

par(mfrow=c(2,2)) # both inputs have been square root transformed
acf(model_xs$residuals)
pacf(model_xs$residuals)
hist(model_xs$residuals)
qqPlot(model_xs$residuals)

par(mfrow=c(2,2)) # only interest was included in the model WITHOUT square root
↪  transform
acf(model_x1$residuals, main="")
pacf(model_x1$residuals, main="")
hist(model_x1$residuals)
qqPlot(model_x1$residuals)

par(mfrow=c(2,2)) # only interest was included in the model WITH square root transform
acf(model_x1s$residuals,main="")
pacf(model_x1s$residuals,main="")
hist(model_x1s$residuals)
qqPlot(model_x1s$residuals)

par(mfrow=c(2,2)) # only inflation was included in the model with root transform
acf(model_x2$residuals)
pacf(model_x2$residuals)
hist(model_x2$residuals)
qqPlot(model_x2$residuals)

##### Forecasting future house prices 2 #####

newxreg <- cbind(interest[(length(interest)-1):length(interest)],
                 inflation[(length(inflation)-1):length(inflation)])
newxreg <- rbind(newxreg, newxreg[2,], newxreg[2,], newxreg[2,], newxreg[2,])
par(mfrow=c(1,1))
# Only inflation has been square root transformed
newxreg <- cbind(newxreg[,1],sqrt(newxreg[,2]))
predictions_x <- forecast(model_x, xreg = newxreg, level = 95)
plot(seq(from=1992.5,to=2024.25,by=0.25), ts[,2], type="l", xlab = "Years",
     ylab = "Average House Sales Prices", col=1)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x$mean)^2, col=2)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x$lower)^2, col=3)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x$upper)^2, col=3)
legend("topleft",c("Observations", "Forecasts","P.I."), col=c(1,2,3),lty=1,
       bty='n')
grid()

# both inputs have been square root transformed
newxreg1 <- cbind(sqrt(newxreg[,1]),sqrt(newxreg[,2]))
predictions_xs <- forecast(model_xs, xreg = newxreg1, level = 95)
plot(seq(from=1992.5,to=2024.25,by=0.25), ts[,2], type="l", xlab = "Years",
```

```r
     ylab = "Average House Sales Prices", col=1)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x$mean)^2, col=2)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x$lower)^2, col=3)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x$upper)^2, col=3)
legend("topleft",c("Observations", "Forecasts","P.I."), col=c(1,2,3),lty=1,
       bty='n')
grid()

# only interest was included in the model WITHOUT square root transform
predictions_x1 <- forecast(model_x1, xreg = newxreg[,1], level = 95)
plot(seq(from=1992.5,to=2024.25,by=0.25), ts[,2], type="l", xlab = "Years",
     ylab = "Average House Sales Prices", col=1)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x1$mean)^2, col=2)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x1$lower)^2, col=3)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x1$upper)^2, col=3)
legend("topleft",c("Observations", "Forecasts","P.I."), col=c(1,2,3),lty=1,
       bty='n')
grid()

# only interest was included in the model WITH square root transform
predictions_x1s <- forecast(model_x1s, xreg = sqrt(newxreg[,1]), level = 95)
plot(seq(from=1992.5,to=2024.25,by=0.25), ts[,2], type="l", xlab = "Years",
     ylab = "Average House Sales Prices", col=1)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x1s$mean)^2, col=2)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x1s$lower)^2, col=3)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x1s$upper)^2, col=3)
legend("topleft",c("Observations", "Forecasts","P.I."), col=c(1,2,3),lty=1,
       bty='n')
grid()

# only inflation was included in the model with root transform
predictions_x2 <- forecast(model_x2, xreg = newxreg[,2], level = 95)
plot(seq(from=1992.5,to=2024.25,by=0.25), ts[,2], type="l", xlab = "Years",
     ylab = "Average House Sales Prices", col=1)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x2$mean)^2, col=2)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x2$lower)^2, col=3)
lines(seq(from=2023.0,to=2024.25,by=0.25),(predictions_x2$upper)^2, col=3)
legend("topleft",c("Observations", "Forecasts","P.I."), col=c(1,2,3),lty=1,
       bty='n')
grid()

cbind(predictions$mean^2,predictions_x$mean^2,predictions_xs$mean^2,predictions_x1$mean^2,predictions_
cbind(predictions$lower^2,predictions_x$lower^2,
↪ predictions_x1$lower^2,predictions_x2$lower^2)
cbind(predictions$upper^2,predictions_x$upper^2,
↪ predictions_x1$upper^2,predictions_x2$upper^2)

################################### Part 2 ###################################

### Plots ###
par(mfrow=c(1,1))
plot(years[1:122], capital, col=2, type = "l", lwd=2,
     ylab = "Average sales prices across all regions", xlab = "years")
```

```
lines(years[1:122], sealand, col=3, lwd=2)
lines(years[1:122], midjutland, col=4, lwd=2)
lines(years[1:122], rural, col=7, lwd=2)
legend("topleft", c("capital region", "Sealand", "Middle Jutland", "rural areas"),
       col = c(2,3,4,7), lty = c(1,1,1,1), bty = "n")
grid()

par(mfrow = c(2,2))
plot(years[1:122], capital, col=1, type = "l", lwd=2,
     ylab = "Average sales prices in the capital region", xlab = "years")
grid()
plot(years[1:122], sealand, col=1, type = "l", lwd=2,
     ylab = "Average sales prices in Sealand", xlab = "years")
grid()
plot(years[1:122], midjutland, col=1, type = "l", lwd=2,
     ylab = "Average sales prices in Middle Jutland", xlab = "years")
grid()
plot(years[1:122], rural, col=1, type = "l", lwd=2,
     ylab = "Average sales prices in the rural areas", xlab = "years")
grid()

par(mfrow=c(1,1))
plot(years[1:121], diff(capital), col=2, type = "l", lwd=2,
     ylab = "Average sales prices across all regions", xlab = "years")
lines(years[1:121], diff(sealand), col=3, lwd=2)
lines(years[1:121], diff(midjutland), col=4, lwd=2)
lines(years[1:121], diff(rural), col=7, lwd=2)
legend("topleft", c("capital region", "Sealand", "Middle Jutland", "rural areas"),
       col = c(2,3,4,7), lty = c(1,1,1,1), bty = "n")
grid()

par(mfrow = c(2,2))
plot(years[1:121], diff(capital), col=1, type = "l", lwd=2,
     ylab = "Average sales prices in the capital region", xlab = "years")
grid()
plot(years[1:121], diff(sealand), col=1, type = "l", lwd=2,
     ylab = "Average sales prices in Sealand", xlab = "years")
grid()
plot(years[1:121], diff(midjutland), col=1, type = "l", lwd=2,
     ylab = "Average sales prices in Middle Jutland", xlab = "years")
grid()
plot(years[1:121], diff(rural), col=1, type = "l", lwd=2,
     ylab = "Average sales prices in the rural areas", xlab = "years")
grid()

adf.test(capital) #nope
adf.test(sealand) #nope
adf.test(midjutland) #nope
adf.test(rural) #nope

adf.test(diff(capital)) #yep
adf.test(diff(sealand)) #yep
adf.test(diff(midjutland)) #nope
```

```r
adf.test(diff(rural)) #nope

adf.test(diff(sqrt(capital))) #yep
adf.test(diff(sqrt(sealand))) #yep
adf.test(diff(sqrt(midjutland))) #yep
adf.test(diff(sqrt(rural))) #nope

adf.test(diff(log(capital))) #yep
adf.test(diff(log(sealand))) #yep
adf.test(diff(log(midjutland))) #yep
adf.test(diff(log(rural))) #yep

adf.test(diff(capital)) #yep
adf.test(diff(sealand)) #yep
adf.test(diff(diff(midjutland))) #yep
adf.test(diff(diff(rural))) #yep

##########################################

# sqrt
ys <- cbind(sqrt(capital[1:122]), sqrt(sealand[1:122]), sqrt(midjutland[1:122]),
            sqrt(rural[1:122]), sqrt(interest[1:122]), sqrt(inflation[1:122]))

differences <- matrix(c(1,1,2,1,3,1,4,1,5,1,6,1), nrow = 2)
Ys <- define.dif(ys, difference=differences)

par(mfrow=c(1,2))
# Capital
acf(Ys$y.dif[1,])
pacf(Ys$y.dif[1,])
csd1 <- arima(Ys$y.dif[1,], order = c(2,0,0), seasonal = list(order =c(0,0,0),
                                                               period = 4))
acf(csd1$residuals)
pacf(csd1$residuals)
hist(csd1$residuals)
qqPlot(csd1$residuals)
tsdiag(csd1)

# Sealand
acf(Ys$y.dif[2,])
pacf(Ys$y.dif[2,])
ssd1 <- arima(Ys$y.dif[2,], order = c(0,0,0), seasonal = list(order =c(1,0,0),
                                                               period = 4))
acf(ssd1$residuals)
pacf(ssd1$residuals)
hist(ssd1$residuals)
qqPlot(ssd1$residuals)
tsdiag(ssd1)

# Middle Jutland
acf(Ys$y.dif[3,])
pacf(Ys$y.dif[3,])
msd1 <- arima(Ys$y.dif[3,], order = c(0,0,0), seasonal = list(order =c(2,0,0),
```

```
                                                              period = 4))
acf(msd1$residuals)
pacf(msd1$residuals)
hist(msd1$residuals)
qqPlot(msd1$residuals)
tsdiag(msd1)

# Rural
acf(Ys$y.dif[4,])
pacf(Ys$y.dif[4,])
rsd1 <- arima(Ys$y.dif[4,], order = c(0,0,0), seasonal = list(order =c(2,0,0),
                                                              period = 4))
acf(rsd1$residuals)
pacf(rsd1$residuals)
hist(rsd1$residuals)
qqPlot(rsd1$residuals)
tsdiag(rsd1)

# Interest
acf(Ys$y.dif[5,])
pacf(Ys$y.dif[5,])
i1sd1 <- arima(Ys$y.dif[5,], order = c(1,0,0), seasonal = list(order =c(0,0,0),
                                                              period = 4))
acf(i1sd1$residuals)
pacf(i1sd1$residuals)
hist(i1sd1$residuals)
qqPlot(i1sd1$residuals)
tsdiag(i1sd1)

# Inflation
acf(Ys$y.dif[6,])
pacf(Ys$y.dif[6,])
i2sd1 <- arima(Ys$y.dif[6,], order = c(2,0,1), seasonal = list(order =c(2,0,0),
                                                              period = 4))
acf(i2sd1$residuals)
pacf(i2sd1$residuals)
hist(i2sd1$residuals)
qqPlot(i2sd1$residuals)
tsdiag(i2sd1)

# log
yl <- cbind(log(capital[1:122]), log(sealand[1:122]), log(midjutland[1:122]),
           log(rural[1:122]), (interest[1:122]), (inflation[1:122]))

difference <- matrix(c(1,1,2,1,3,1,4,1,5,1,6,1), nrow = 2)
Yl <- define.dif(yl, difference=difference)

par(mfrow=c(1,2))
# Capital
acf(Yl$y.dif[1,])
pacf(Yl$y.dif[1,])
cld1 <- arima(Yl$y.dif[1,], order = c(2,0,0), seasonal = list(order =c(0,0,0),
                                                             period = 4))
```

```
acf(cld1$residuals)
pacf(cld1$residuals)
hist(cld1$residuals)
qqPlot(cld1$residuals)
tsdiag(cld1)

# Sealand
acf(Yl$y.dif[2,])
pacf(Yl$y.dif[2,])
sld1 <- arima(Yl$y.dif[2,], order = c(0,0,0), seasonal = list(order =c(2,0,0),
                                                                  period = 4))
acf(sld1$residuals)
pacf(sld1$residuals)
hist(sld1$residuals)
qqPlot(sld1$residuals)
tsdiag(sld1)

# Middle Jutland
acf(Yl$y.dif[3,])
pacf(Yl$y.dif[3,])
mld1 <- arima(Yl$y.dif[3,], order = c(0,0,0), seasonal = list(order =c(2,0,0),
                                                             period = 4))
acf(mld1$residuals)
pacf(mld1$residuals)
hist(mld1$residuals)
qqPlot(mld1$residuals)
tsdiag(mld1)

# Rural
acf(Yl$y.dif[4,])
pacf(Yl$y.dif[4,])
rld1 <- arima(Yl$y.dif[4,], order = c(0,0,0), seasonal = list(order =c(2,0,0),
                                                             period = 4))
acf(rld1$residuals)
pacf(rld1$residuals)
hist(rld1$residuals)
qqPlot(rld1$residuals)
tsdiag(rld1)

# Interest
acf(Yl$y.dif[5,])
pacf(Yl$y.dif[5,])
i1d1 <- arima(Yl$y.dif[5,], order = c(2,0,1), seasonal = list(order =c(0,0,0),
                                                              period = 4))
acf(i1d1$residuals)
pacf(i1d1$residuals)
hist(i1d1$residuals)
qqPlot(i1d1$residuals)
tsdiag(i1d1)

# Inflation
acf(Yl$y.dif[6,])
pacf(Yl$y.dif[6,])
```

```
i2d1 <- arima(Yl$y.dif[6,], order = c(2,0,1), seasonal = list(order =c(2,0,0),
                                                              period = 4))
acf(i2d1$residuals)
pacf(i2d1$residuals)
hist(i2d1$residuals)
qqPlot(i2d1$residuals)
tsdiag(i2d1)

# diff
yd <- cbind((capital[1:122]), (sealand[1:122]), (midjutland[1:122]),
            (rural[1:122]), (interest[1:122]), (inflation[1:122]))

differenced <- matrix(c(1,1,2,1,3,1,3,1,4,1,4,1,5,1,6,1), nrow = 2)
Yd <- define.dif(yd, difference=differenced)

# Capital
acf(Yd$y.dif[1,])
pacf(Yd$y.dif[1,])
cd <- arima(Yd$y.dif[1,], order = c(2,0,0), seasonal = list(order =c(0,0,0),
                                                             period = 4))
acf(cd$residuals)
pacf(cd$residuals)
hist(cd$residuals)
qqPlot(cd$residuals)
tsdiag(cd)

# Sealand
acf(Yd$y.dif[2,])
pacf(Yd$y.dif[2,])
sd <- arima(Yd$y.dif[2,], order = c(0,0,0), seasonal = list(order =c(0,0,0),
                                                             period = 4))
acf(sd$residuals)
pacf(sd$residuals)
hist(sd$residuals)
qqPlot(sd$residuals)
tsdiag(sd)

# Middle Jutland
acf(Yd$y.dif[3,])
pacf(Yd$y.dif[3,])
md <- arima(Yd$y.dif[3,], order = c(3,0,1), seasonal = list(order =c(0,0,0),
                                                             period = 4))
acf(md$residuals)
pacf(md$residuals)
hist(md$residuals)
qqPlot(md$residuals)
tsdiag(md)

# Rural
acf(Yd$y.dif[4,])
pacf(Yd$y.dif[4,])
rd <- arima(Yd$y.dif[4,], order = c(3,0,0), seasonal = list(order =c(1,0,0),
                                                             period = 4))
```

```r
acf(rd$residuals)
pacf(rd$residuals)
hist(rd$residuals)
qqPlot(rd$residuals)
tsdiag(rd)

# Interest
acf(Yd$y.dif[5,])
pacf(Yd$y.dif[5,])
i1d <- arima(Yd$y.dif[5,], order = c(2,0,1), seasonal = list(order =c(0,0,0),
                                                              period = 4))
acf(i1d$residuals)
pacf(i1d$residuals)
tsdiag(i1d)

# Inflation
acf(Yd$y.dif[6,])
pacf(Yd$y.dif[6,])
i2d <- arima(Yd$y.dif[6,], order = c(2,0,1), seasonal = list(order =c(2,0,0),
                                                              period = 4))
acf(i2d$residuals)
pacf(i2d$residuals)
tsdiag(i2d)

##############################################################################

res <- cbind(cld1$residuals,sld1$residuals,mld1$residuals,rld1$residuals)
acf(res)
pacf(res)

##################### ACF & PACF ###########################################

par(mfrow=c(2,3))
acf(Yl$y.dif[1,],main="capital")
acf(Yl$y.dif[2,],main="sealand")
acf(Yl$y.dif[3,],main="midjutland")
acf(Yl$y.dif[4,],main="rural")
acf(Yl$y.dif[5,],main="interest")
acf(Yl$y.dif[6,],main="inflation")
par(mfrow=c(2,3))
pacf(Yl$y.dif[1,],main="capital")
pacf(Yl$y.dif[2,],main="sealand")
pacf(Yl$y.dif[3,],main="midjutland")
pacf(Yl$y.dif[4,],main="rural")
pacf(Yl$y.dif[5,],main="interest")
pacf(Yl$y.dif[6,],main="inflation")
par(mfrow=c(2,3))
acf(Yd$y.dif[1,],main="capital")
acf(Yd$y.dif[2,],main="sealand")
acf(Yd$y.dif[3,],main="midjutland")
acf(Yd$y.dif[4,],main="rural")
acf(Yd$y.dif[5,],main="interest")
acf(Yd$y.dif[6,],main="inflation")
```

```r
par(mfrow=c(2,3))
pacf(Yd$y.dif[1,],main="capital")
pacf(Yd$y.dif[2,],main="sealand")
pacf(Yd$y.dif[3,],main="midjutland")
pacf(Yd$y.dif[4,],main="rural")
pacf(Yd$y.dif[5,],main="interest")
pacf(Yd$y.dif[6,],main="inflation")

par(mfrow=c(1,1))
acf(t(Yl$y.dif[1:6,]))
pacf(t(Yl$y.dif[1:6,]))
acf(t(Yd$y.dif[1:6,]))
pacf(t(Yd$y.dif[1:6,]))

### Guesstimate MARIMAX order ###

# 1st option, Yl log: ar=1,2,4,8,12 ,ma=1
# 2nd option, Yd 2-diff: ar=1,2,4,8,12 ,ma=1
ar=c(1,2,4,8)
ma=c(1)
mmodel <- define.model(kvar = 6, ar = ar, ma = ma, reg.var = c(5,6))
mmodel.fit <- marima(Yl$y.dif, means=1,
                     ar.pattern = mmodel$ar.pattern,
                     ma.pattern=mmodel$ma.pattern,
                     Check=TRUE, Plot='log.det',
                     penalty=0.0)

acf(t(mmodel.fit$residuals[c(1:4),-(1:max(ar))]))
pacf(t(mmodel.fit$residuals[c(1:4),-(1:max(ar))]))

### Residuals Analysis ###

par(mfrow=c(2,2))
# plot #
plot(mmodel.fit$residuals[1,-(1:max(ar))])
abline(h=0)
plot(mmodel.fit$residuals[2,-(1:max(ar))])
abline(h=0)
plot(mmodel.fit$residuals[3,-(1:max(ar))])
abline(h=0)
plot(mmodel.fit$residuals[4,-(1:max(ar))])
abline(h=0)
# check if the mean value is 0 #
t.test(mmodel.fit$residuals[1,-(1:max(ar))], mu=0)
t.test(mmodel.fit$residuals[2,-(1:max(ar))], mu=0)
t.test(mmodel.fit$residuals[3,-(1:max(ar))], mu=0)
t.test(mmodel.fit$residuals[4,-(1:max(ar))], mu=0)
# check if there is no autocorrelation #
Box.test(mmodel.fit$residuals[1,-(1:max(ar))], type="Ljung-Box")
Box.test(mmodel.fit$residuals[2,-(1:max(ar))], type="Ljung-Box")
Box.test(mmodel.fit$residuals[3,-(1:max(ar))], type="Ljung-Box")
Box.test(mmodel.fit$residuals[4,-(1:max(ar))], type="Ljung-Box")
# check the sign change #
```

```r
rbinom(10, length(mmodel.fit$residuals[1,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit$residuals[1,-(1:max(ar))])) != 0)
rbinom(10, length(mmodel.fit$residuals[2,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit$residuals[2,-(1:max(ar))])) != 0)
rbinom(10, length(mmodel.fit$residuals[3,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit$residuals[3,-(1:max(ar))])) != 0)
rbinom(10, length(mmodel.fit$residuals[4,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit$residuals[4,-(1:max(ar))])) != 0)
# Test in the cumulated periodogram
cpgram(mmodel.fit$residuals[1,-(1:max(ar))], main="")
cpgram(mmodel.fit$residuals[2,-(1:max(ar))], main="")
cpgram(mmodel.fit$residuals[3,-(1:max(ar))], main="")
cpgram(mmodel.fit$residuals[4,-(1:max(ar))], main="")
# check if it is normal distributed #
hist(mmodel.fit$residuals[1,-(1:max(ar))], main="")
hist(mmodel.fit$residuals[2,-(1:max(ar))], main="")
hist(mmodel.fit$residuals[3,-(1:max(ar))], main="")
hist(mmodel.fit$residuals[4,-(1:max(ar))], main="")
qqPlot(mmodel.fit$residuals[1,-(1:max(ar))])
qqPlot(mmodel.fit$residuals[2,-(1:max(ar))])
qqPlot(mmodel.fit$residuals[3,-(1:max(ar))])
qqPlot(mmodel.fit$residuals[4,-(1:max(ar))])

source("step.slow.marima_2017.R")
source("step.slow.p.marima_2017.R")

mmodel.fit.sl <- step.slow.p(mmodel.fit, data=Yl$y.dif)
mmodel.fit.sl

acf(t(mmodel.fit.sl$residuals[1:4,-(1:max(ar))]))
pacf(t(mmodel.fit.sl$residuals[1:4,-(1:max(ar))]))

par(mfrow=c(2,2))
# plot #
plot(mmodel.fit.sl$residuals[1,-(1:max(ar))])
abline(h=0)
plot(mmodel.fit.sl$residuals[2,-(1:max(ar))])
abline(h=0)
plot(mmodel.fit.sl$residuals[3,-(1:max(ar))])
abline(h=0)
plot(mmodel.fit.sl$residuals[4,-(1:max(ar))])
abline(h=0)
# check if the mean value is 0 #
t.test(mmodel.fit.sl$residuals[1,-(1:max(ar))], mu=0)
t.test(mmodel.fit.sl$residuals[2,-(1:max(ar))], mu=0)
t.test(mmodel.fit.sl$residuals[3,-(1:max(ar))], mu=0)
t.test(mmodel.fit.sl$residuals[4,-(1:max(ar))], mu=0)
# check if there is no autocorrelation #
Box.test(mmodel.fit.sl$residuals[1,-(1:max(ar))], type="Ljung-Box")
Box.test(mmodel.fit.sl$residuals[2,-(1:max(ar))], type="Ljung-Box")
Box.test(mmodel.fit.sl$residuals[3,-(1:max(ar))], type="Ljung-Box")
Box.test(mmodel.fit.sl$residuals[4,-(1:max(ar))], type="Ljung-Box")
# check the sign change #
```

```
rbinom(10, length(mmodel.fit.sl$residuals[1,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit.sl$residuals[1,-(1:max(ar))])) != 0)
rbinom(10, length(mmodel.fit.sl$residuals[2,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit.sl$residuals[2,-(1:max(ar))])) != 0)
rbinom(10, length(mmodel.fit.sl$residuals[3,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit.sl$residuals[3,-(1:max(ar))])) != 0)
rbinom(10, length(mmodel.fit.sl$residuals[4,-(1:max(ar))])-1, 0.5)
sum(diff(sign(mmodel.fit.sl$residuals[4,-(1:max(ar))])) != 0)
# Test in the cumulated periodogram
cpgram(mmodel.fit.sl$residuals[1,-(1:max(ar))], main="")
cpgram(mmodel.fit.sl$residuals[2,-(1:max(ar))], main="")
cpgram(mmodel.fit.sl$residuals[3,-(1:max(ar))], main="")
cpgram(mmodel.fit.sl$residuals[4,-(1:max(ar))], main="")
# check if it is normal distributed #
hist(mmodel.fit.sl$residuals[1,-(1:max(ar))], main="")
hist(mmodel.fit.sl$residuals[2,-(1:max(ar))], main="")
hist(mmodel.fit.sl$residuals[3,-(1:max(ar))], main="")
hist(mmodel.fit.sl$residuals[4,-(1:max(ar))], main="")
qqPlot(mmodel.fit.sl$residuals[1,-(1:max(ar))])
qqPlot(mmodel.fit.sl$residuals[2,-(1:max(ar))])
qqPlot(mmodel.fit.sl$residuals[3,-(1:max(ar))])
qqPlot(mmodel.fit.sl$residuals[4,-(1:max(ar))])

###

ar.aggregated <- pol.mul( mmodel.fit.sl$ar.estimate, Yl$dif.poly,
                          L = ( max(ar) + dim( Yl$dif.poly)[3]))

### Forecasts ###

pred.data <- t(yl)

pred.data <- cbind(pred.data
                   , c(NA, NA, NA, NA, (ts[123,7:8]))
                   , c(NA, NA, NA, NA, (ts[124,7:8]))
                   , c(NA, NA, NA, NA, (ts[124,7:8]))
                   , c(NA, NA, NA, NA, ts[124,7:8])
                   , c(NA, NA, NA, NA, ts[124,7:8])
                   , c(NA, NA, NA, NA, ts[124,7:8]))

pred <-  arma.forecast(series=pred.data, nstart=122, nstep=6,
                     dif.poly = Yl$dif.poly, marima=mmodel.fit.sl, check = TRUE)
par(mfrow=c(1,1))

plot(years, exp(pred.data[1,]), ylab="sales prices")
lines(years, exp(pred$forecasts[1,]))
pred.int <- pred$forecasts[1,123:128] +
                 cbind(rep(0, 6), -1, 1)*qnorm(0.975)*sqrt(pred$pred.var[1,1,])
matlines(years[123:128], exp(pred.int), lty=c(1,2,3), col=2, lwd=2 )
legend("topleft",c("Observations", "Fitted Values", "Forecasts","P.I."),
       col=c(1,1,2,2), pch = c(1, NA, NA, NA),lty=c(NA,1, 1, 2),
       bty='n')
grid()
```

```
plot(years, exp(pred.data[2,]), ylab="sales prices")
lines(years, exp(pred$forecasts[2,]))
pred.int <- pred$forecasts[2,123:128] +
                  cbind(rep(0, 6), -1, 1)*qnorm(0.975)*sqrt(pred$pred.var[2,2,])
matlines(years[123:128], exp(pred.int), lty=c(1,2,2), col=2, lwd=2 )
legend("topleft",c("Observations", "Fitted values", "Forecasts","P.I."),
       col=c(1,1,2,2), pch = c(1, NA, NA, NA),lty=c(NA, 1, 1, 2),
       bty='n')
grid()

plot(years, exp(pred.data[3,]), ylab="sales prices")
lines(years, exp(pred$forecasts[3,]))
pred.int <- pred$forecasts[3,123:128] +
                  cbind(rep(0, 6), -1, 1)*qnorm(0.975)*sqrt(pred$pred.var[3,3,])
matlines(years[123:128], exp(pred.int), lty=c(1,2,2), col=2, lwd=2 )
legend("topleft",c("Observations", "Fitted values", "Forecasts","P.I."),
       col=c(1,1,2,2), pch = c(1, NA, NA, NA),lty=c(NA, 1, 1, 2),
       bty='n')
grid()

plot(years, exp(pred.data[4,]), ylab="sales prices")
lines(years, exp(pred$forecasts[4,]))
pred.int <- pred$forecasts[4,123:128] +
                  cbind(rep(0, 6), -1, 1)*qnorm(0.975)*sqrt(pred$pred.var[4,4,])
matlines(years[123:128], exp(pred.int), lty=c(1,2,2), col=2, lwd=2 )
legend("topleft",c("Observations", "Fitted Values", "Forecasts","P.I."),
       col=c(1,1,2,2), pch = c(1, NA, NA, NA),lty=c(NA, 1, 1, 2),
       bty='n')
grid()

##Conclusion

scatterplot(capital, sealand)
scatterplot(capital, midjutland)
scatterplot(capital, rural)
scatterplot(sealand, midjutland)
scatterplot(sealand, rural)
scatterplot(midjutland, rural)

scatterplot(Yl$y.dif[1,], Yl$y.dif[2,])
scatterplot(Yl$y.dif[1,], Yl$y.dif[3,])
scatterplot(Yl$y.dif[1,], Yl$y.dif[4,])
scatterplot(Yl$y.dif[2,], Yl$y.dif[3,])
scatterplot(Yl$y.dif[2,], Yl$y.dif[4,])
scatterplot(Yl$y.dif[3,], Yl$y.dif[4,])
```