Final Report for the Course Predictive Analytics

Predicting Housing Prices in Copenhagen

Author: Konrad Schulte (149872)

Examiner: Andrea Tafuro

Copenhagen, 15.06.2022

Pages: 10

Characters including Spaces: 20,614

Table of Contents

1. Introduction

While inflation has been rising all over Europe in recent months and new highs are being announced continuously, the ECB is still sticking to its lax monetary policy with negative interest rates (Forbes, 2022). For the traditional saver, this means that he or she loses money every day just by leaving it in the account. The Corona pandemic and the war in Ukraine have also shown that the stock markets are becoming increasingly detached from reality. This uncertainty discourages many people from investing their assets in stocks or other securities. Another, more conservative option for investing assets lies in the real estate market. However, in order to invest in real estate, the investor must first ascertain how the value of this property will develop over the next few years.

This paper will use ten years of monthly data to try to predict this development in value for the Copenhagen apartment and house market. More precisely, it will try to predict square meter prices in Copenhagen. For this purpose, different ETS and ARIMA forecasting models are developed and compared with respect to various statistical criteria. The study shows that among all models, the ETS (A,Ad,A) model performs best. Overall, however, it appears to be very difficult to produce a truly reliable forecast for housing prices based on time trends and the methods used.

2. Data Origin & Preparation

The dataset used for this project was created by web scraping data from the danish housing portal boliga.dk. Boliga.dk is used for buying or selling apartments, houses, and other property types in Denmark. Moreover, it lists historical property transactions with information about property size, price, location and many more. Especially the included prices per square meter in the data have made Boliga.dk a very valuable data source. Most available datasets about housing prices do not include property sizes or prices per square meter which are important for proper comparison. In total, the raw dataset contained 24 columns and 1.7 million observations with transactions all over Denmark ranging from 1992 to 2022. In a first preprocessing step the dataset was cleaned for strong outliers regarding prices and size. Subsequently, the dataset was reduced to the last ten full years, namely 2012 to 2021, and filtered only for apartments and houses. As the aim of this paper is to forecast housing prices in Copenhagen, the dataset was also filtered for properties only in Copenhagen (including Frederiksberg). The dataset with daily information was then aggregated to monthly data using the median. I decided for the median and against the average to prevent the data from being biased by strong outliers. After this preparation and a last filtering for the relevant two columns, I was left with a dataset consisting of those two columns ("Price" and "Month") and 120 observations. The final dataset was then transformed to a time series using the "as_tsibble" command from the "tsibble" package in R.

3. Exploratory Data Analysis & Data Preprocessing

The time series starts in January 2012 and ends in December 2021 (given in the *%Y %b* format). Prices per square meter are given in DKK and range from 22,902.23 (02/2012) to 49,814.25 (07/2021) with a median of 35,567 and a slightly lower mean of 35,386. The forecasting models should later be configured to capture all important qualitative properties of the data. Therefore, in the following section, the time series is examined for various characteristics and, depending on their presence, a transformation of the data is performed.

3.1 Deflation

Since the prices are financial data, the dataset first must be adjusted for inflation. To accomplish this, I downloaded the list of the monthly consumer price index for Denmark from the Danish governmental organization "Statistics Denmark" and deflate all 120 prices respectively (unfortunately, the CPI available in R packages is only in annual format and only extended to 2017, which is insufficient in my case) (Statistics Denmark, 2022).

3.2 Nonlinearity

Figure 1 shows the time series after deflation. At first glance, this appears to be relatively linear data. However, especially the price development from 2018 onwards is characterized by a new dynamic, which includes a steeper development until mid-2021. In addition, the fluctuations in prices seem to vary over the years. All this lead me to apply a logarithmic transformation to the data. The necessity for a logarithmic transformation can also be confirmed by a calculated lambda_guerrero = -0.258 (i.e. close to zero). However, regarding a logarithmic transformation it must also be considered that this entails a certain loss of information, and the data must be transformed back later. A comparison of the time series before and after logarithmic transformation can be seen in appendix 1.

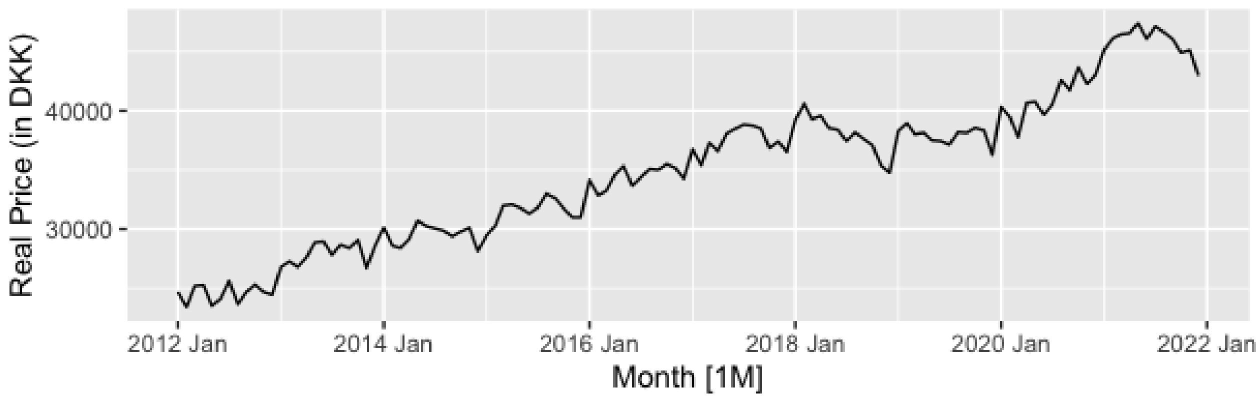


Figure 1: Time Series of (deflated) Housing Prices in Copenhagen

3.3 Trend

Another look at figure 1 shows that, despite minor fluctuations, prices have been rising steadily since 2012 until they experience a slight slump in 2018. From 2019 onwards, however, prices start to rise again. All in all, the trend is increasing from start to end. This trend can also be clearly seen in the decomposition plot in figure 2. The correlogram in figure 3 shows (with the number of lags decreasing though still) significant autocorrelations for all depicted lags and is constantly positive. This also clearly indicates a trend. In addition to just identifying a trend, non-stationarity of the data can also be concluded from this. This finding will be discussed in more detail later in section 3.4.

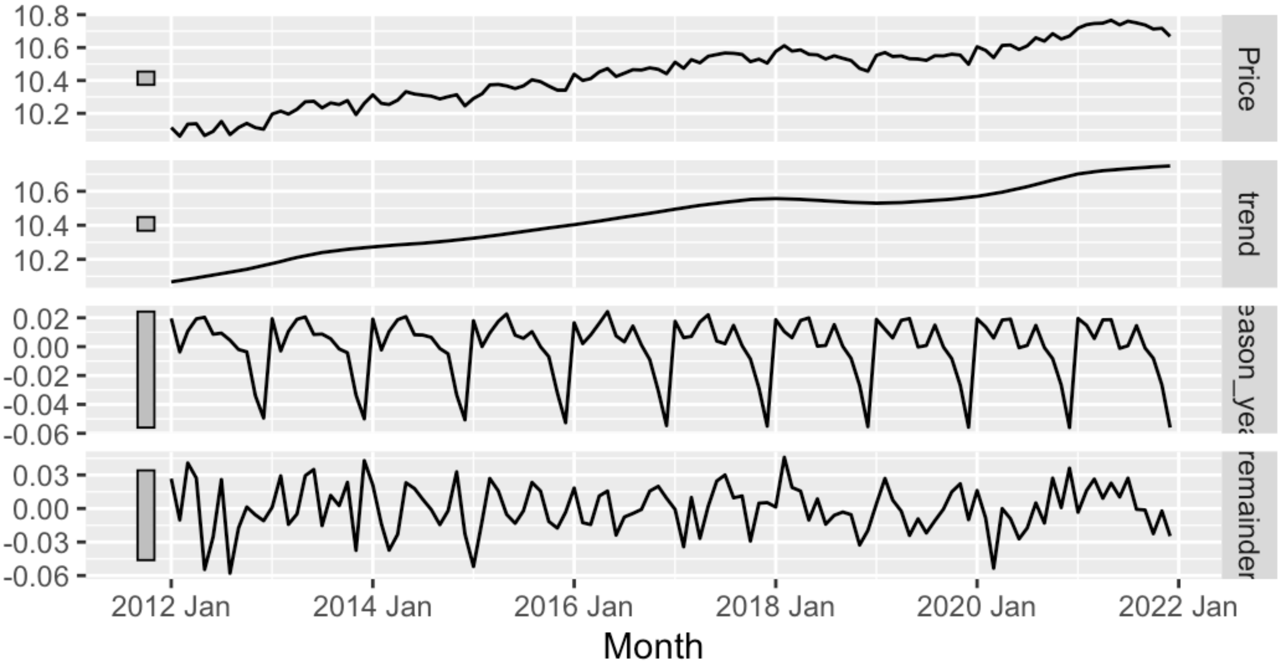


Figure 2: Classical Additive Decomposition of Time Series

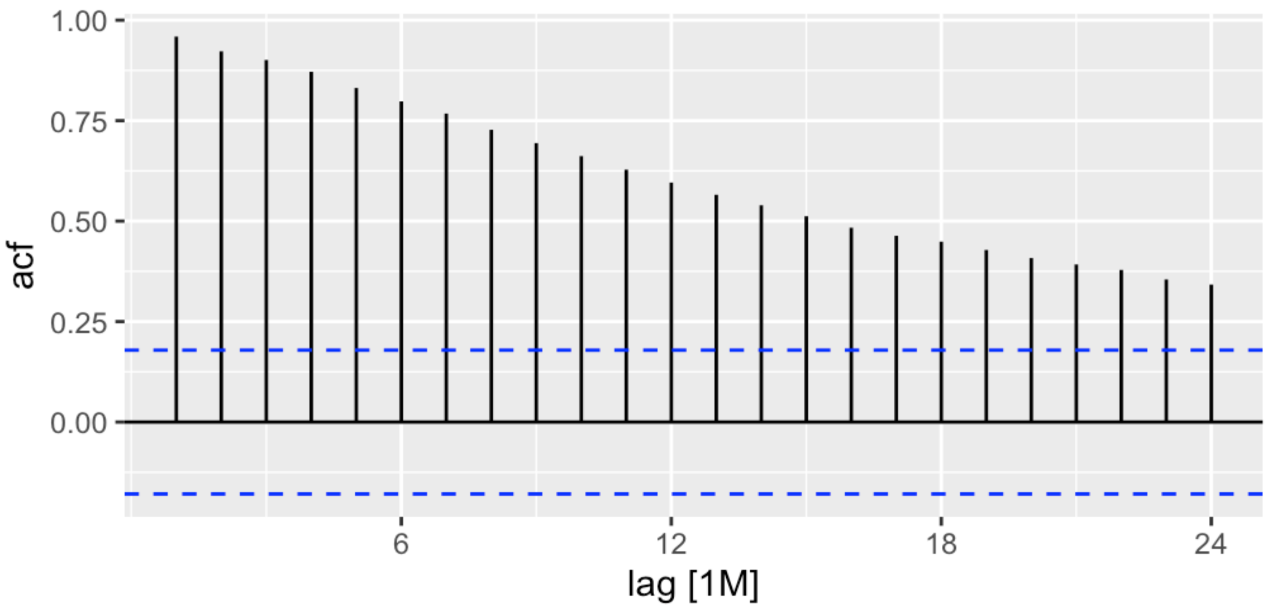


Figure 3: Correlogram of Time Series

3.4 Seasonality

Figure 2 seems to show a clear seasonality of the data. However, this should be interpreted with caution, since classical decomposition inherently assumes a seasonal component. Looking again at the time series (figure 1), it is difficult to see a clear seasonality. One could argue that there is a certain regularity of falling prices towards the end of the year (especially at the end of 2018 and 2019), but this is not always the case. A clearer picture is provided by the seasonal plot in figure 4, which shows that price developments over the different years do not follow a consistent pattern. Consequently, seasonality of the data can be ruled out. This decision is supported by the result of the "unitroot_nsdiffs()" function from the "feasts" package in R. The function indicates that no seasonal differentiation is necessary, which consequently means that there is no seasonality.

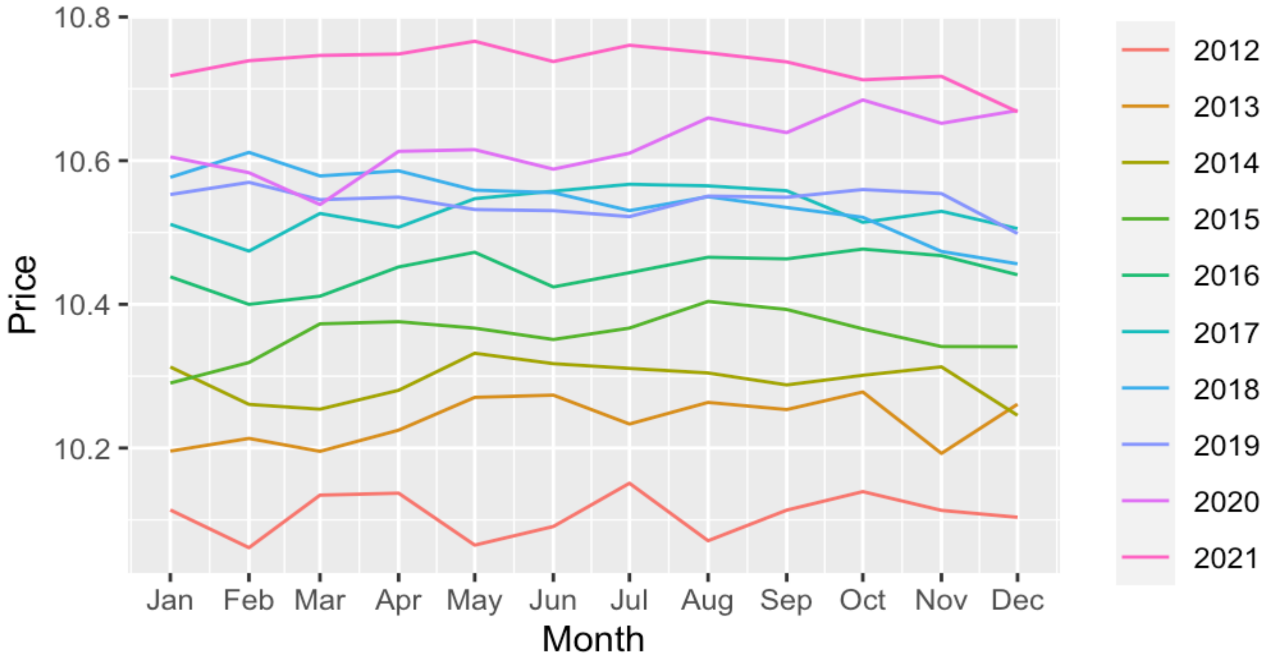


Figure 4: Seasonal Plot of Time Series

3.5 Stationarity

Non-stationarity can cause forecasting models to perform worse, since most models assume independent data points. Therefore, in the following we investigate whether housing prices are stationary or non-stationary and in the latter case transformations are applied to make the data stationary. A first strong indicator of non-stationarity was the significance of all lags in the ACF plot in section 3.3 (figure 2). For certainty, two tests for (non) stationarity are further applied in the following: Kwiatkowski-Phillips-Schmidt-Shin Test (KPSS) and the Augmented Dickey-Fuller Test (ADF). The null hypothesis of the KPSS indicates stationarity, while the null hypothesis of the ADF tests for the presence of a unit root (and thus non stationarity). Therefore, a rejection of the H0 of the KPSS as well as an acceptance of the H0 of the ADF means that non-stationary data is present.

Because of the trend assumed in section 3.3, the KPSS is first of the type "tau" and the ADF of the type "trend". The number of lags of the ADF are defined by the Akaike Information Criterion (AIC) to obtain a more conservative test. The KPSS test gives a test-statistic value of 0.2632, which is higher than all critical values. The H0 of the KPSS can therefore be rejected. The null hypothesis of a unit root with drift and trend of the ADF can be accepted, because the first value of the test-statistic is higher than the significance level at 1% and both other values of the test-statistic are lower than the significance level at 1%. Both tests thus demonstrate non-stationarity of the data. These results are additionally supported by the fact that the "unitroot_ndiffs()" function from the "feasts" package in R recommends a differentiation of the data. Although this function is also based on the KPSS test, it is suitable as a quick cross-check in case errors might have occurred during the tests themselves.

To make the data stationary, I take the first difference and test ADF and KPSS again. Now that the value of the first test-statistic is significantly smaller and the second and third values are significantly larger than the respective critical values, the H0 of the ADF ca. n be rejected. As learned in the lecture, I will now also perform the ADF tests of the type "drift" and "none". Their null hypothesis can also be rejected. The KPSS shows for the differentiated data a lower value than the significance level at 10%. The KPSS of the type "mu" also comes to this result, which is why in both cases the null hypothesis of stationarity cannot be rejected and thus accepted.

As a result, the time series is stationary after one differentiation and no further differentiations are necessary. All test reports for the stationarity ADF and KPSS tests can be found in appendix 2-8.

3.6 Structural Breaks

As a last step of the exploratory data analysis, the time series is examined for possible structural breaks. A visual analysis based on figure 1 does not reveal any clearly recognizable breaks (Stock & Watson, 1996). Therefore, the Quandt likelihood ratio (QLR) test (Quandt, 1960) is used to identify endogenous breaks in the time series. Figure 5 shows the results of the QLR test. Accordingly, the housing price time series does not contain any (endogenous) structural breaks. Thus, the time series does not need to be truncated.

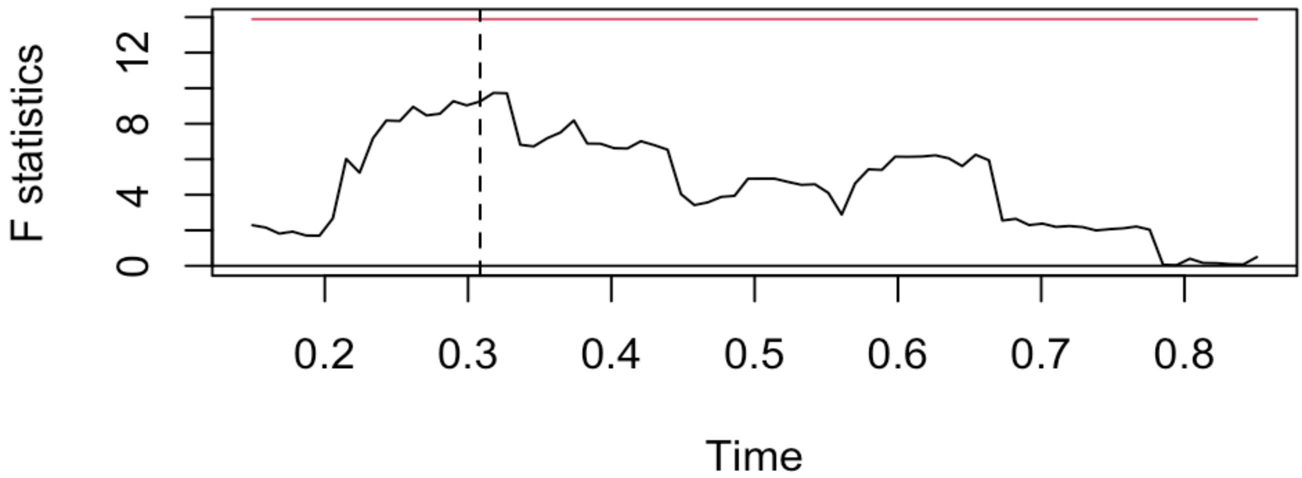


Figure 5: QLR Test

3.7 Sample Split

To prepare the data for the models, the time series is split into a train and a test set. The train set covers 8 years (from January 2012 to December 2019) and the test set 2 years (from February 2020 to December 2021). The ratio of train to test set is therefore 80:20.

4. Methodology

This section is devoted to the definition of suitable ETS and ARIMA models for the prediction of housing prices in Copenhagen. The architecture of the models is determined using statistical methods. In addition, an automatically generated model (i.e., another automatically generated ETS and ARIMA model) is used for comparison. After creating the models, the respective residuals are then examined for potential finetuning. Based on the three information criteria (namely Akaike information criterion (AIC), AIC with a correction for small sample sizes (AICc), and the Bayesian information criterion (BIC)), the best model of the respective class is then determined.
The forecasts of the two best models are then evaluated and the overall best model is used to forecast the development of housing prices in Copenhagen over the next two years.

4.1 ETS Model Creation

Exponential Smoothing (ETS) is one of the most widely used forecasting techniques today (Holt, 2004). The ETS model combines Error, Trend and Seasonality in a smoothing calculation. Each of these three can become an additive or multiplicative component of the function or be omitted entirely.

How the different components are weighted in this case can be derived from the preceding data analysis. A logarithmic transformation was performed in the preprocessing, which leads to an additive error (A). Furthermore, both decomposition (figure 2) and correlogram (figure 3) indicated an additive trend (A). Since seasonality was excluded as a component of the time series, this parameter is also declared null (N) in the ETS model. In summary, the self-selected model is an ETS (A,A,N). The automatic ETS model also includes an additive error component. Unlike my model, however, it assumes a damped additive trend and an additive seasonality and thus has the following form: ETS (A,Ad,A).

	ETS (A,A,N)	ETS (A,Ad,A)
AIC	-195.8868	-225.0918
AICc	-195.2274	-216.3226
BIC	-183.0132	-178.7270

Table 1: Comparison of both ETS Models using AIC, AICc, BIC

The comparison of both models in table 1 suggests that the automatic model has lower values for the AIC and the AICc, but a slightly higher BIC value than the model. This is plausible, since BIC penalizes higher complexity in models. Since the higher complexity of the automatic model does not weigh so heavily in my case, I decide to choose the automatic ETS as the best ETS model and will now examine its properties a bit more closely.

α	β	γ	ℓ	b	s
0.4140	1.05e-4	2.73e-4	10.0565	0.0120	-0.0517 to 0.0232

Table 2: Estimated Smoothing Parameters and Initial States of chosen ETS (A,Ad,A)

As seen in table 2, α is about 0.41. This indicates that newer observations are not weighted much stronger than older ones. The low values for β and γ show that both the seasonal component and the slope have a strong inertia.

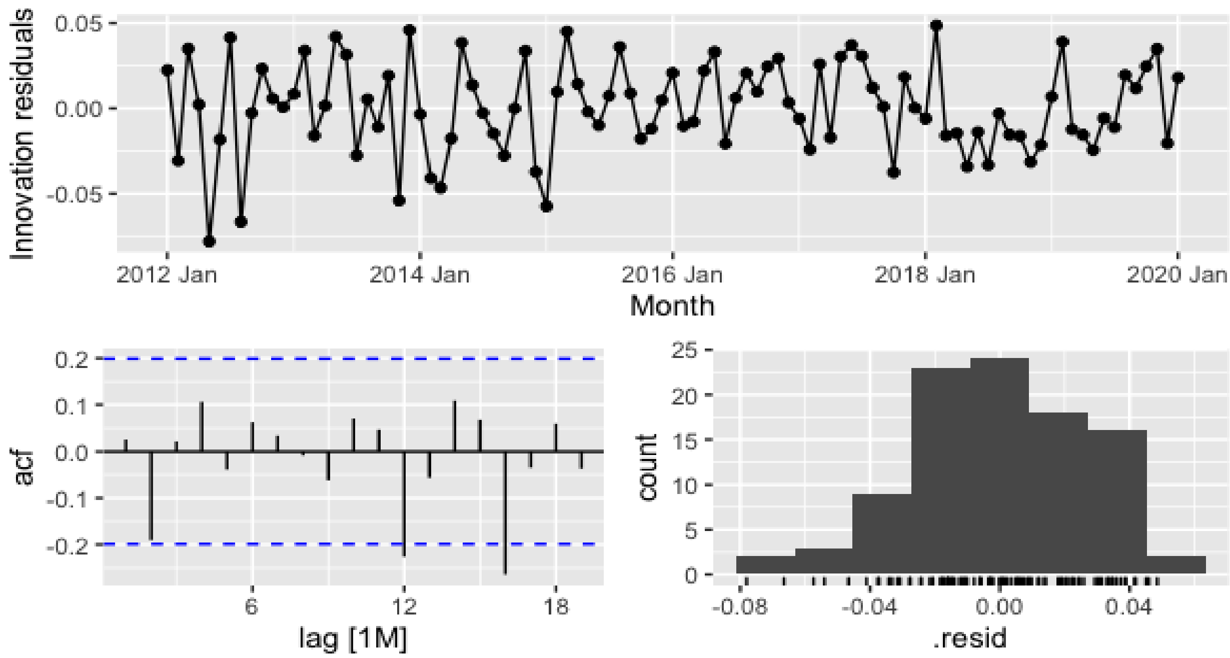


Figure 6: Residuals of ETS (A,Ad,A) Model

Figure 6 does not reveal a good picture with respect to the residuals. Although only two spikes in the ACF plot are significant, the variance of the residuals over time does not seem to be constant and the distribution of the residuals does not correspond to a normal distribution. Furthermore, the null hypothesis of the Ljung Box test is rejected with a p-value of 0.0335, implying a correlation of the residuals. Thus, the ETS (A,Ad,A) unfortunately does not capture all variation in the data and has potential for improvement. However, these improvements will not be covered in this paper.

4.2 ARIMA Model Creation

The Autoregressive integrated moving average (ARIMA) is another class of statistical models for the prediction of time series data. ARIMA models also combine three components: an autoregression (AR), an integration (I) and a moving average (MA). A typical ARIMA model has the following structure: (p,d,q)(P,D,Q), where lower case letters represent the non-seasonal part and upper case letters the seasonal part (Box et al., 2015).

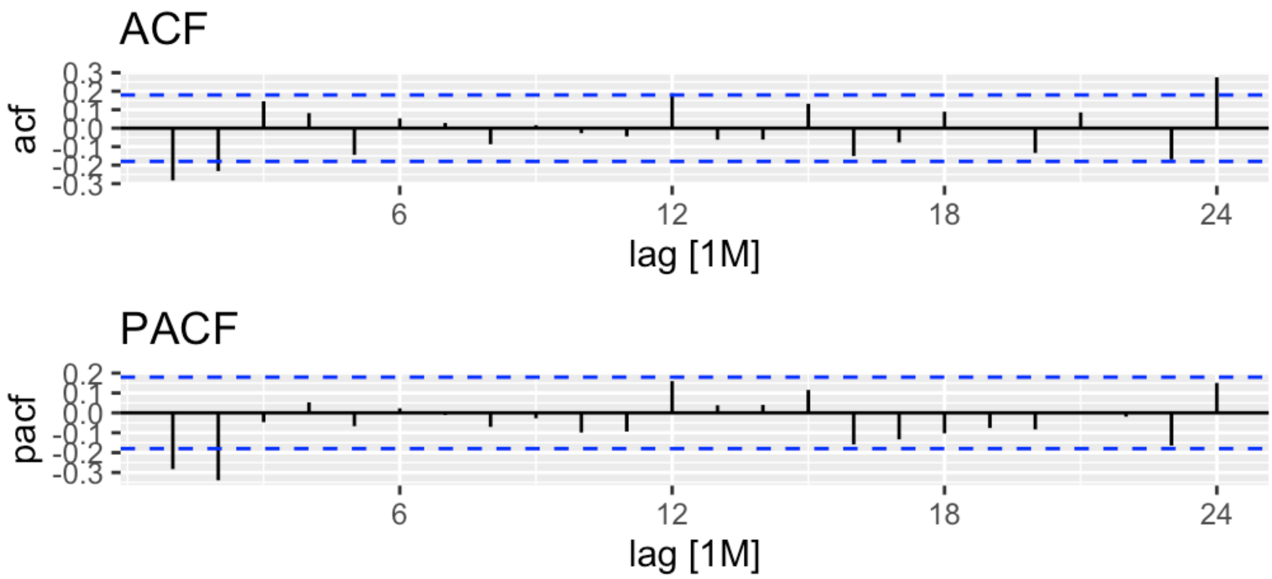


Figure 7: ACF and PACF of Stationary Time Series

For the creation of an own ARIMA model, ACF and PACF of the stationary time series (figure 7) are examined more closely. Neither the spikes in the ACF nor in the PACF seem to follow a certain pattern. While the PACF plot shows only two significant spikes at lag 1 and 2, the ACF plot has a total of four significant spikes (at lag 1,2, 12 and 24). If the ACF were exponentially decaying or sinusoidal, q could be equated to 0 according to Hyndman & Athanasopoulos (2018). However, as previously described, neither ACF nor PACF seem to follow any particular pattern. The two significant spikes in lag 1 and 2 of the PACF therefore lead to the choice of a value of 2 for q. Since the time series was differentiated once to obtain stationarity, d can be set equal to 1. Because seasonality was excluded in the course of the data analysis, the self-generated model is an ARIMA(0,1,2) model. The examination of the residuals (see appendix 9) shows that the ACF plot has a significant spike at lag 12. To cover an unexpected seasonality in the data, a second model ARIMA(0,1,2)(0,0,1) is investigated as an alternative model. Although the ACF of the residuals (appendix 10) of the second model no longer shows a significant spike at lag 12, it does show a new spike at lag = 16. Therefore, no further finetuning is attempted and instead a third, automatic ARIMA model is added. Auto.arima creates an ARIMA(0,2,1)(0,0,2) model, i.e., with a seasonal MA(2) instead of a seasonal MA(1) as in the second model but unchanged beyond that.

	ARIMA (0,1,2)	ARIMA (0,1,2)(0,0,1)	ARIMA (0,1,2) (0,0,2)
AIC	-364.25	-365.86	-368.72
AICc	-363.81	-365.86	-367.78
BIC	-353.99	-353.71	-353.34

Table 3: Comparison of all ARIMA Models using AIC, AICc, BIC

Table 3 shows the comparison of the three models using AIC, AICc, and BIC. The automatic model performs better in terms of AIC and AICc. However, the lowest value in terms of BIC can be seen in the first self-generated ARIMA model. I follow the recommendation of Hyndman & Athanasopoulos (2021) and attach a higher importance to the AICc. Thus, the best ARIMA model is the ARIMA (0,1,2)(0,0,2) model.

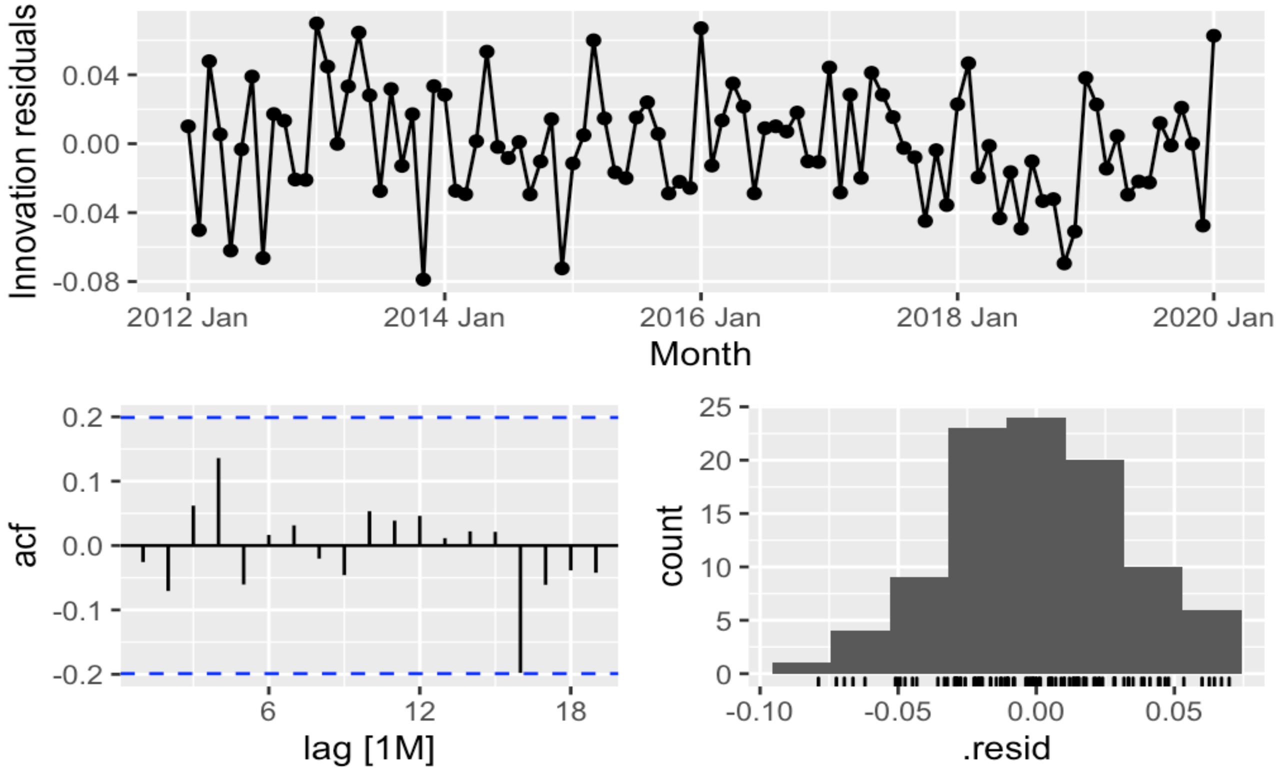


Figure 8: Residuals of ARIMA (0,1,2)(0,0,2) Model

As figure 8 shows, the residuals seem to follow almost a normal distribution with a slight skew to the left. The ACF plot also shows no significant spikes, which is a good sign. Nevertheless, the residuals do not seem to have a uniform variance. Furthermore, the Ljung Box test is rejected with a p-value of 0.0245, so the residuals are correlated. All in all, the model seems to capture some information of the data well, but also has room for improvement.

5. Results

In this section the two chosen ETS and ARIMA forecasting models are compared. First, the logarithmically transformed data is transformed back and then forecasts of both models are evaluated visually and numerically. Error analysis is performed by comparing four known statistical ratios: Root-mean-square-error (RMSE), mean-absolute-error (MAE), mean-absolute-percentage error (MAPE), and mean absolute scaled error (MASE).

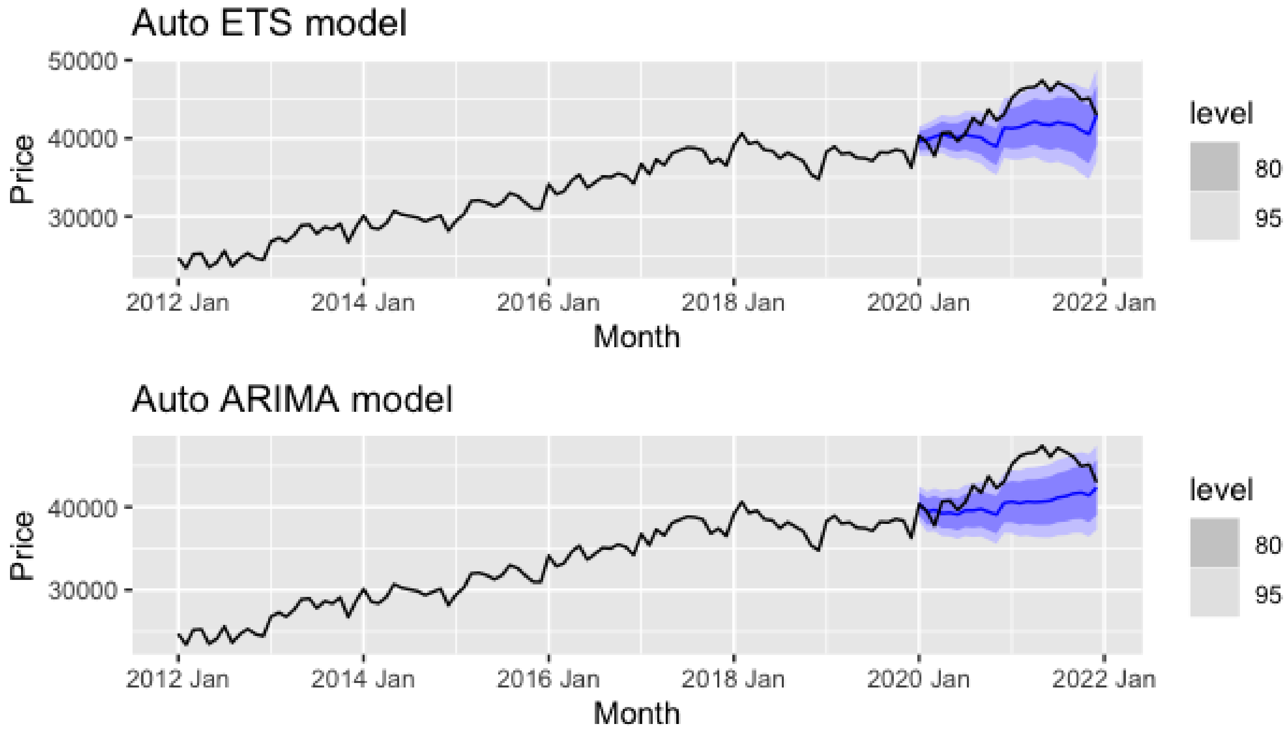


Figure 9: Forecasts of ETS and ARIMA Model among Entire Time Series

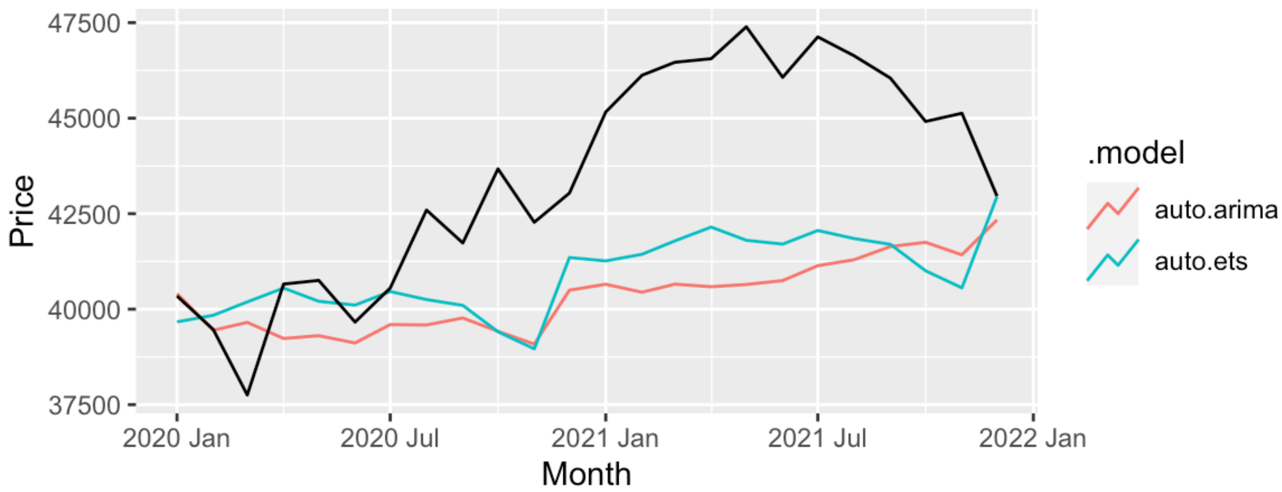


Figure 10: Forecast of Both Models on Test Set

Figure 9 and figure 10 show the forecasts of both models individually with the original total data and a direct comparison of both models with the test set, respectively. A first glance reveals that both models undershoot the real data. The ETS model seems to be able to reproduce the fluctuation of the last year (even if slightly shifted) a bit better than the ARIMA model. Whether this is just a coincidental development, or a seasonality is present in the time series indeed, could be evaluated in future work. The ARIMA model does not show such a strong seasonality. Both models have relatively large confidence intervals. All in all, the ETS model seems to perform a bit better than the ARIMA model, although an even better model could perhaps be created from an average of both models. This would also be a topic for future work.

	RMSE	MAE	MAPE	MASE
ARIMA	3,875	3,274	7.28	1.44
ETS	3,414	2,843	6.33	1.25

Table 4: Accuracy Measures of Best ETS and ARIMA Models

Table 4 shows the forecast errors for the test set in comparison. The forecast errors for the training set were deliberately omitted, as I do not think they add any value (a good fit to the training data does not necessarily indicate a high forecast quality). The ETS model outperforms the ARIMA model with respect to all accuracy measures. Thus, the ETS model is chosen as the final model and for the forecast of the next years.
Since I have used real prices for the analysis so far, but for the overall forecast nominal prices are of greater interest, the forecast is corrected by an inflation of 6%. This two-year forecast is shown in figure 11 and shows a further increase in prices per square meter in Copenhagen.

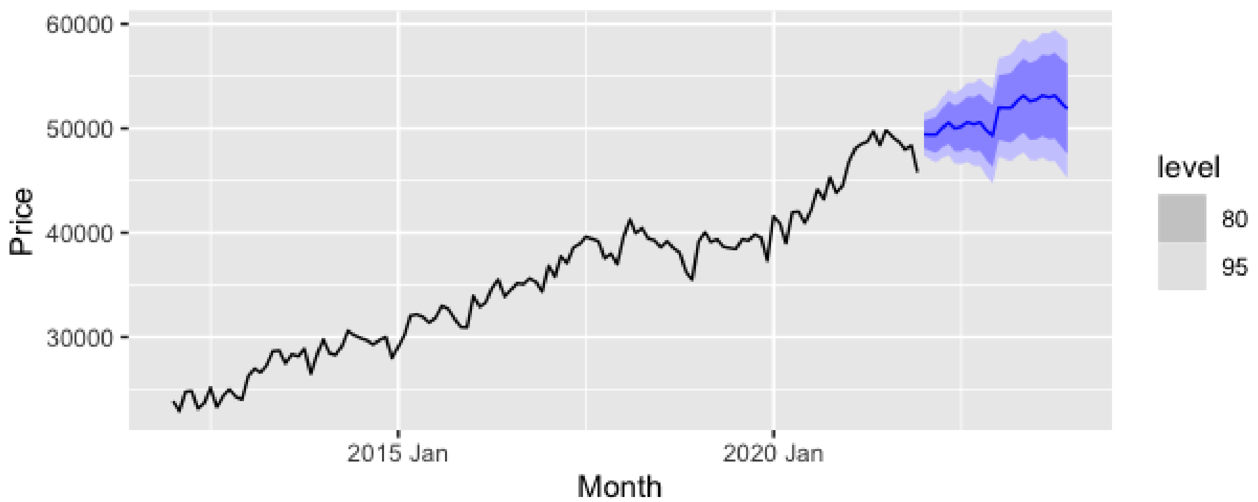


Figure 11: 2-Year Forecast of Final Model ETS (A,Ad,A)

6. Conclusion

The aim of this paper was to investigate whether it is possible with current forecasting models to make reliable forecasts for housing prices in Copenhagen. The study shows that it is difficult to build a reliable model due to many irregular fluctuations in prices. The best of the models was the ETS (A,Ad,A) model, although even this is relatively far from a good forecast in terms of performance.
Certainly, there is some ground for future work. As already mentioned in section 5, the two best models of the time series data have assumed a seasonality. I could not detect this seasonality from my results. Future work could investigate in more detail whether this seasonality is indeed present. In any case, much more data would be needed for future investigations, which could allow a more gauged, long-term prediction. In this context, the use of quarterly data would also be advisable to mitigate the partially strong monthly fluctuations. Future efforts could also be made to optimize the forecast by combining several models. I suspect that in general a temporal analysis of price developments alone is not sufficient to identify all influences on housing prices that need to be taken into account for a good forecast. The housing bubble and crisis around 2008/2009 shows this very vividly. Therefore, it would be interesting to integrate other developments or features into such a model in the future.

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th edition). Wiley.

Forbes. (2022, May 26). *Inflation And The European Central Bank: It Is A Mess*. Forbes. https://www.forbes.com/sites/philippsandner/2022/05/26/inflation-and-the-european-central-bank-it-is-a-mess/

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Hyndman, R., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. https://research.monash.edu/en/publications/forecasting-principles-and-practice-3

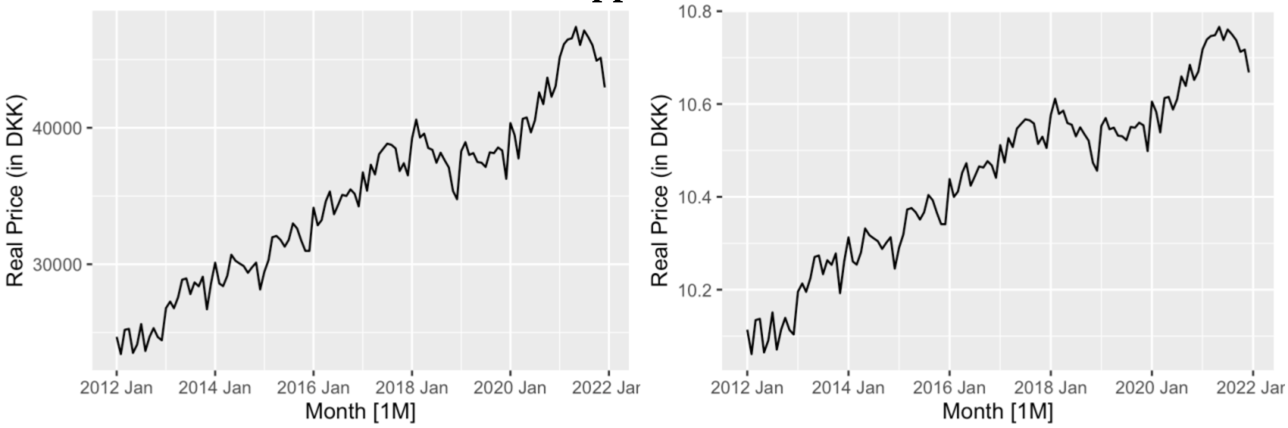
Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. https://research.monash.edu/en/publications/forecasting-principles-and-practice-2

Quandt, R. E. (1960). Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association*, *55*(290), 324–330. https://doi.org/10.1080/01621459.1960.10482067

Statistics Denmark. (2022). *Consumer price index*. https://www.dst.dk/en/Statistik/emner/oekonomi/prisindeks/forbrugerprisindeks

Stock, J., & Watson, M. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business & Economic Statistics*, *14*(1), 11–30.

Appendix



Appendix 1: Comparison of Time Series Before (left) and After (right) Logarithmic Transformation

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.085657 -0.021733 -0.000015  0.022431  0.081250

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.4935382  0.7157515   3.484 0.000702 ***
z.lag.1     -0.2450870  0.0706643  -3.468 0.000740 ***
tt           0.0011860  0.0003754   3.159 0.002028 **
z.diff.lag  -0.1668061  0.0920581  -1.812 0.072624 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03454 on 114 degrees of freedom
Multiple R-squared:  0.175,     Adjusted R-squared:  0.1533
F-statistic:  8.06 on 3 and 114 DF,  p-value: 6.45e-05


Value of test-statistic is: -3.4683 5.6433 6.34

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

Appendix 2: ADF Test (type = ”trend”) of Log-Transformed Time Series

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: tau with 4 lags.

Value of test-statistic is: 0.2632

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

Appendix 3: KPSS Test (type = “tau”) of Differentiated Time Series

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.090800 -0.021759 -0.000294  0.020076  0.079452

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0123769  0.0064798   1.910 0.058658 .
z.lag.1     -1.7188592  0.1411164 -12.180  < 2e-16 ***
tt          -0.0000659  0.0000930  -0.709 0.480006
z.diff.lag   0.3450480  0.0872916   3.953 0.000135 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03392 on 113 degrees of freedom
Multiple R-squared:  0.6823,    Adjusted R-squared:  0.6739
F-statistic:  80.9 on 3 and 113 DF,  p-value: < 2.2e-16


Value of test-statistic is: -12.1804 49.4778 74.1843

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

Appendix 4: ADF Test (type = ”trend”) of Differentiated Time Series

```
################################################
# Augmented Dickey-Fuller Test Unit Root Test #
################################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.088259 -0.021017 -0.001469  0.022175  0.080388

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.008395   0.003219   2.608 0.010334 *
z.lag.1     -1.713659   0.140617 -12.187  < 2e-16 ***
z.diff.lag   0.343002   0.087053   3.940 0.000141 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03385 on 114 degrees of freedom
Multiple R-squared:  0.6809,    Adjusted R-squared:  0.6753
F-statistic: 121.6 on 2 and 114 DF,  p-value: < 2.2e-16


Value of test-statistic is: -12.1867 74.29

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```
Appendix 5: ADF Test (type = ”drift”) of Differentiated Time Series

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-0.08054 -0.01334  0.00800  0.02872  0.08981

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -1.62774    0.14011 -11.618  < 2e-16 ***
z.diff.lag  0.30146    0.08771   3.437  0.00082 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03469 on 115 degrees of freedom
Multiple R-squared:  0.662,     Adjusted R-squared:  0.6561
F-statistic: 112.6 on 2 and 115 DF,  p-value: < 2.2e-16


Value of test-statistic is: -11.6178

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

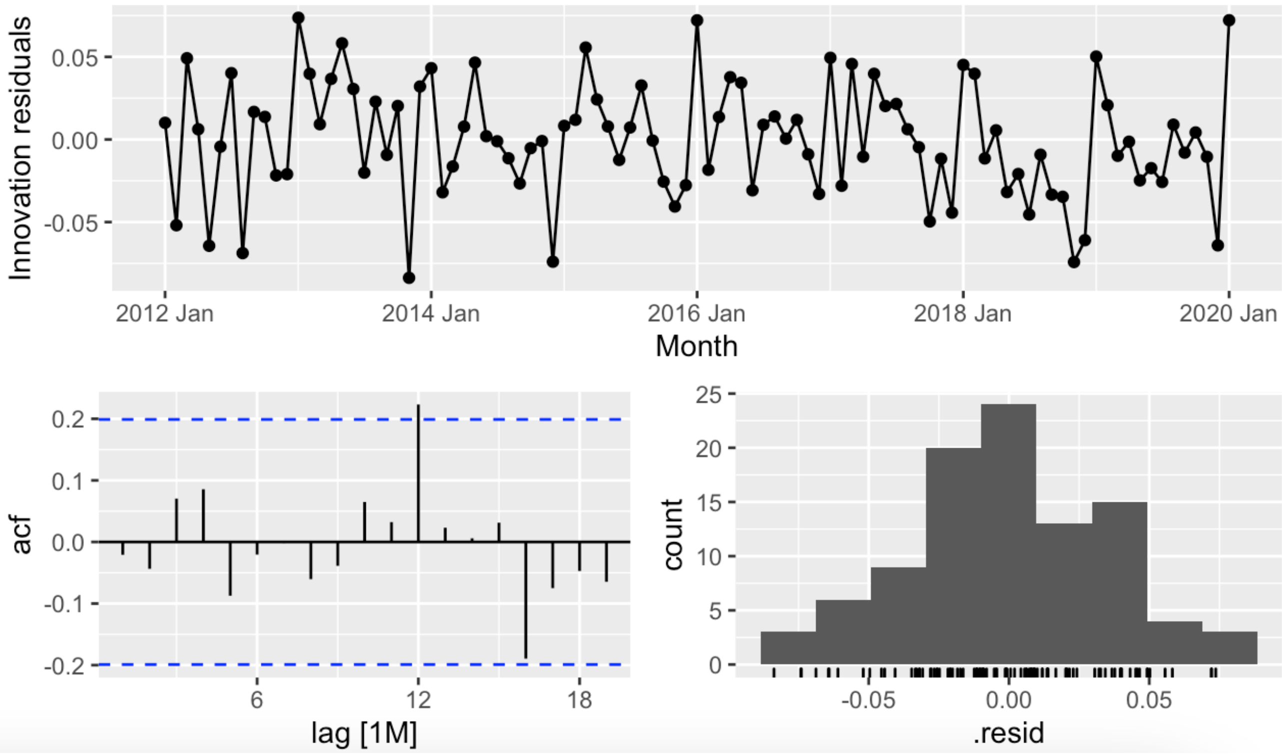
Appendix 6: ADF Test (type = ”none”) of Differentiated Time Series

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: tau with 4 lags.

Value of test-statistic is: 0.0332

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

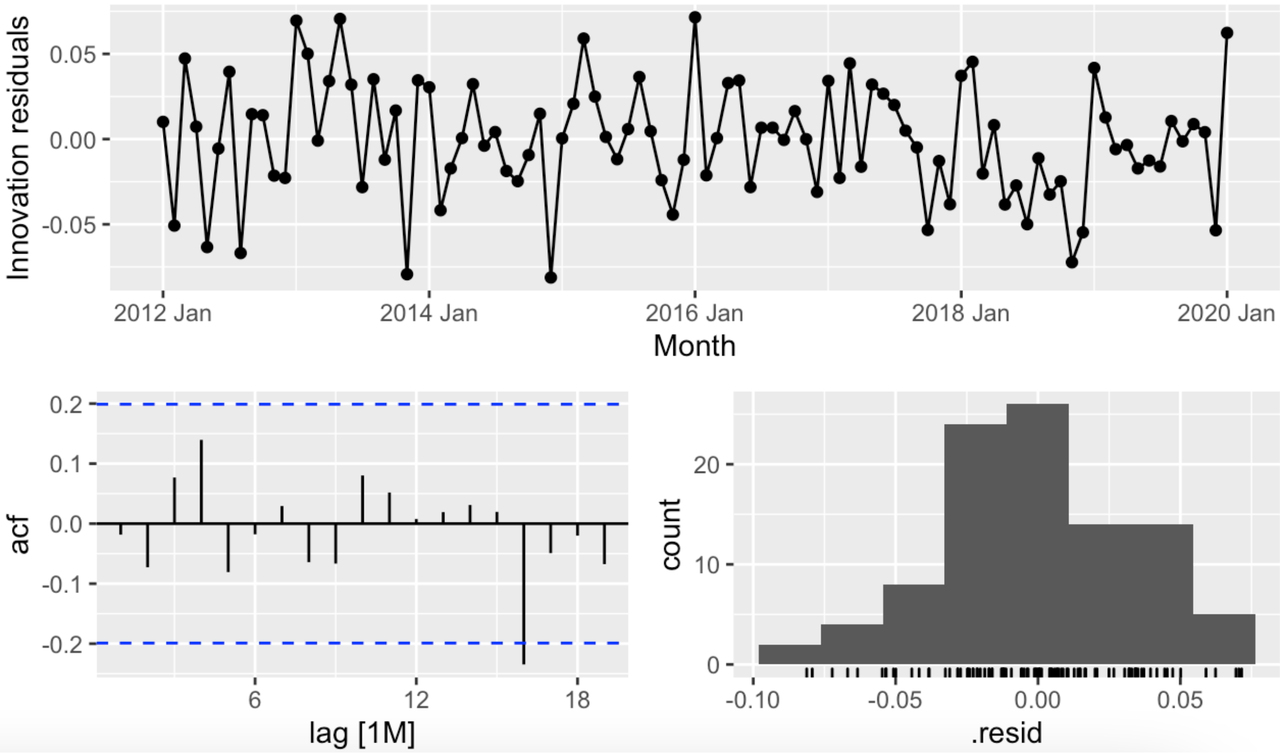
Appendix 7: KPSS Test (type = “tau”) of Differentiated Time Series

```
######################
# KPSS Unit Root Test #
######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.0747

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Appendix 8: KPSS Test (type = “mu”) of Differentiated Time Series



Appendix 9: Residuals from self-generated ARIMA (0,2,1) Model



Appendix 10: Residuals from self-tuned ARIMA (0,2,1)(0,0,1) Model