Project Task 2

Aleksej Hoffaerber, Egor Zmaznev

3/4/2020

Contents

List of Figures

Time-series Analysis and Summaries

Analysis of Trend, Seasons, and Spurious Regression Issues

As subject of study serves the Australian consumption and income data. Having a look at the time series, we see a high amount of correlation between both series that needs to be analysed using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test, the Augmented Dickey-Fuller (ADF) Test, and corrected using differencing and other transformations. It can be observed that the seasonality and trend components are similar for both series, as seen in *Figure 1*. The distance between consumption and income increases with the later years, meaning that people in Australia earn more than they spend (see *Figure 1*). Also, spurious regression issues can be discarded as a potential issue, as disposable available income is impacts the consumption of the population because consumption depends on earnings and debt available.

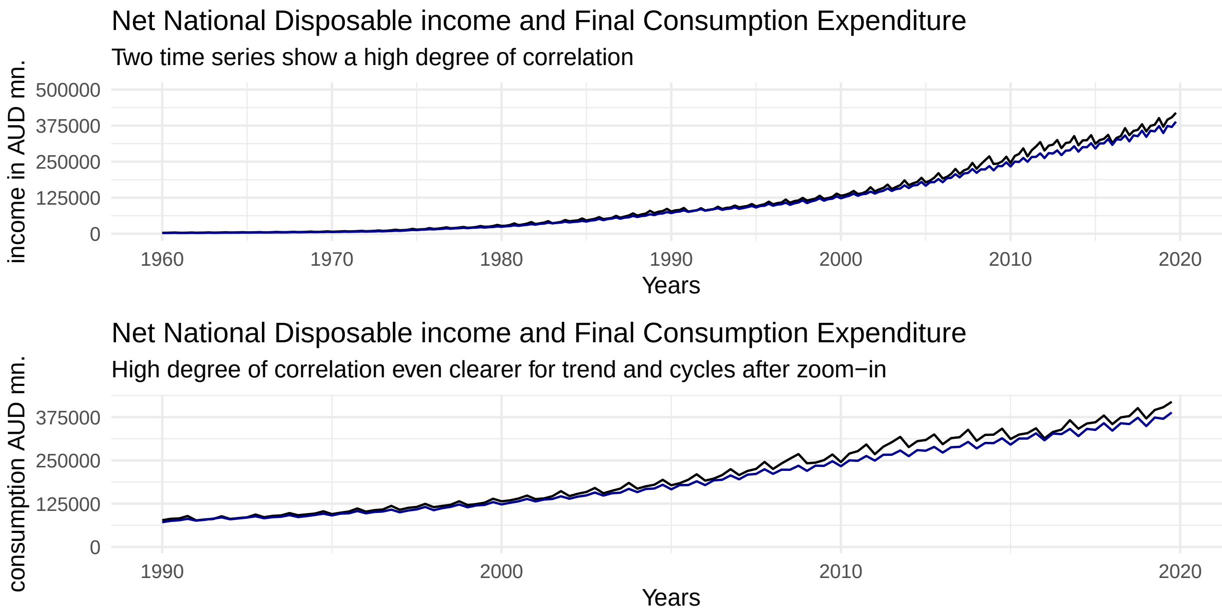

Figure 1: Correlation and Trend of Australian Consumption and Income Data

Stationarity Analysis

The KPSS tests null hypothessis is that the data is stationary (Hyndman & Athanasopoulos, 2018). Our test statistic is with `10.8` much bigger than the 1% critical level, meaning that the data is stationary, as p>0.01. The ADF test tests also for the presence of a unit root process in the time series, with the null hypothesis indicating the stationarity of the underlying time series. We use the `drift` characteristic because of the observed characteristics of the time series and in order to test for a variety of test scenarios (Enders, 2014). ADF indicates that we fail to reject our null hypothesis for `tau2` as p>0.05, meaning that gamma is unequal zero and therefore our data is non-stationary. But, we keep our the null hypothesis for `phi1` meaning that `alpha0` and `gamma` are equal to each other and to zero (Enders, 2014), meaning that we have a stationary series and a drift component. Same applies for an ADF test with trend. Despite rejecting the null hypothesis for `tau3`, saying that gamma is unequal to zero and therefore our data stataionary, `phi2` and `phi3` indicate to a significance level of 5% that we have a `drift` and `trend` component (see *Appendix*). Based on these characteristics adn the similarities in the time series, both time series are not stationary:

```
##
```

```
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 4 lags.
##
## Value of test-statistic is: 4.4141
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
##
## ###############################################
## # 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
## -13554.7   -792.3   -273.5    734.8  20410.6
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 408.195325 276.668476   1.475  0.14147
## z.lag.1       0.008923   0.002464   3.621  0.00036 ***
## z.diff.lag1  -0.216599   0.044570  -4.860 2.17e-06 ***
## z.diff.lag2  -0.195532   0.044699  -4.374 1.84e-05 ***
## z.diff.lag3  -0.206351   0.044619  -4.625 6.25e-06 ***
## z.diff.lag4   0.789050   0.044586  17.697  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2937 on 231 degrees of freedom
## Multiple R-squared:  0.9186, Adjusted R-squared:  0.9169
## F-statistic: 521.6 on 5 and 231 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: 3.6214 10.0668
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

Autocorrelation and Residual Analysis

Checking the residuals indicates that lots of autocorrelation remains in the residuals. In other words: valuable information that is currently not used to predict the data. This autocorrelation pattern is also typical for this type of economic data and time series. The significance of autocorrelation issue is very high, with the p-value being zero, as indicated by the Ljung-Box test in the *Appendix*. Additionally, strong trend cycles and seasonality, visible in the ACF and residual plots indicate that both series are non-stationary. Based on the distribution of the residuals we also see, that stabilisation of variance and mean are particularly important in order to design a suitable forecasting model based on the AR(I)MA process.

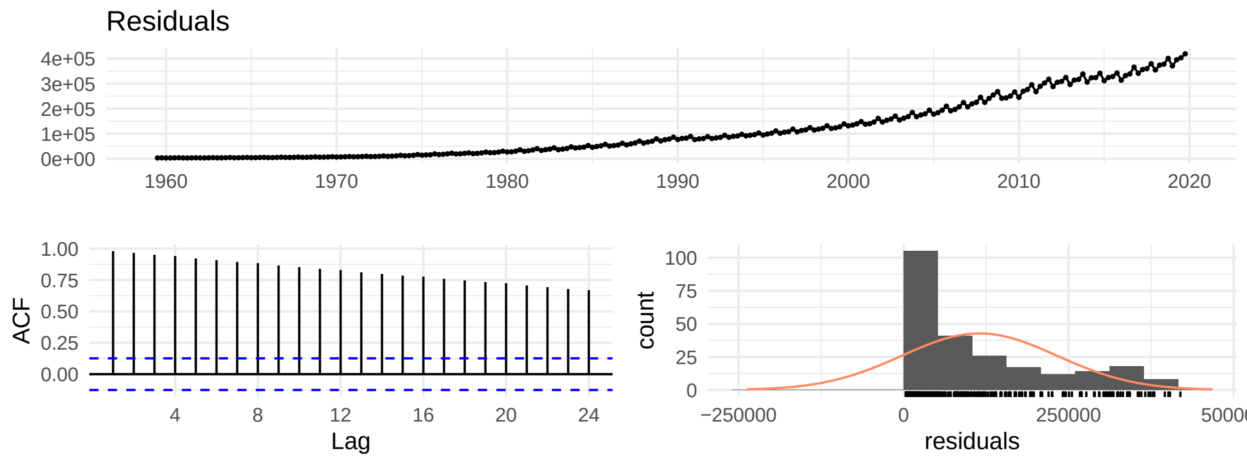

Figure 2: Residual Analysis of Disposable Income

Disposable Income Analysis

To deal with the economic characteristic we apply a Box-Cox transformation to the data. But, this is not enough, as after the transformation not only the autocorrelation and heteroscedasticity issues remain, but also because the distibution of the residuals does not have a fitting Gauss distribution, as can be seen in *Figure 2*. To fix this, we apply differencing methods to stabilize the mean and maintain the interpretability of the model and its results. As we saw with the previous KPSS test result that rejected the null hypothesis for all p-levels from 10 to 1 percent, we need to apply a differencing method.

To calculate a fitting number of differences we must scrutinise the ACF plots: Based on the previous analyses, we know that seasons and trends play a role. A way to account for both issues and to keep the interpretability of the results is to apply first-order and seasonal differencing. In this way, we account for quarterly and seasonal (so yearly) difference in the series and can interpret the results as *quarterly changes* in the respective variables.

As can be seen in *Appendix*, in *Figures 21 and 22* and after applying Box-Cox transformation, seasonal, and first-order differencing, the ACF plots look less impacted by issues, such as autocorrelation, heteroscedasticity, and in total: non-stationarity. The main and variance are now stabilised. But, there are still many spikes in the ACF and residuals, but those do not follow a specific pattern. Still, the Ljung-Box test indicates that a significant amount of autocorrelation remains in the residuals, which we can not get rid off based on our pre-processing toolset. Specific assumptions and calculations in the ARIMA setting will account for the last issues. Applying the KPSS test, we see that the data is now stationary. The difference in both test results is based on autocorrelation or serial correlation, which is a much stronger indication than stationarity.

##

```
##  Box-Ljung test
##
## data:  diff(diff(log(tse[, 1]), 4), 1)
## X-squared = 46.203, df = 10, p-value = 1.318e-06
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 4 lags.
##
## Value of test-statistic is: 0.0193
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

To be completely sure this transformation is correct we apply KPSS functions in order to determine lag values for the differencing. Unsuprisingly, KPSS indicates to use first-order and seasonal differencing.

Consumption Expenditure Analysis

Exactly the same test and procedure will be applied to the Final consumption Expenditure and seperate data will not be displayed. Because both time series look very similar (see *Figure 1*), it can be inferred that a very similar transformation must be applied. This holds true, as the plots *Appendix*, *Figure 23* show. We incorporate both Box-Cox, seasonal and first-order differencing transformed series into the data frame and ts object.

Long-Term Relationship Analysis

As already indicated above, just regressing both variables on each other might lead to spurious regressions, identifiable by a high Adj. R-sqaured and high residual autocorrelation (Hyndman & Athanasopoulos, 2018). This occurence impacts the reliance of our forecast in the long-term horizon. Based on the previously conducted KPSS test, we already know that non-stationarity and possible cointegration of our series are an issue. These unit root tests already indicated that in order to guarantuee that the characteristic equation lies within the unit circle (Hyndman & Athanasopoulos, 2018), we must take the differences in order to assure stationarity.

As we see, from the analyses, incorporating our transformed series for consumption into the regression does resolve our stationarity issues in the regression setting. This indicates, that both variables have a significant, cointegrated long-term relationship, which can be used to design our forecasts. This also holds true if we apply the alternative values for `tau2`, being -3.43, -2.86, and -2.57 respectively.

```
##
## Call:
## tslm(formula = ts.train[, "di.adj"] ~ ts.train[, "ce.adj"])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13982.0   -696.7     -3.1    591.7   7994.6
##
```

```
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           18.18071  152.90762   0.119    0.905
## ts.train[, "ce.adj"]  -0.01602    0.14930  -0.107    0.915
##
## Residual standard error: 2142 on 195 degrees of freedom
## Multiple R-squared:  5.902e-05,  Adjusted R-squared:  -0.005069
## F-statistic: 0.01151 on 1 and 195 DF,  p-value: 0.9147
```

```
##
## ###############################################
## # 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
## -13131.4   -703.9    -10.0    649.4   7594.7
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.60735  153.50005   0.037    0.971
## z.lag.1      -0.86793    0.09819  -8.839 6.15e-16 ***
## z.diff.lag   -0.01465    0.07346  -0.199    0.842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2143 on 192 degrees of freedom
## Multiple R-squared:  0.4337, Adjusted R-squared:  0.4278
## F-statistic: 73.53 on 2 and 192 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -8.8394 39.0922
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

Identify ARIMA Model for Consumption Expenditure

Manual ARIMA Selection Approach

Firstly, based on the previously determined time series characteristics, we know that we must employ a first-order lag differences for our autoregression (AR) part and a first-order seasonal component for our quarterly data, meaining that `d` and `D` are equal to one in order to reflect the observations and KPSS test from before.

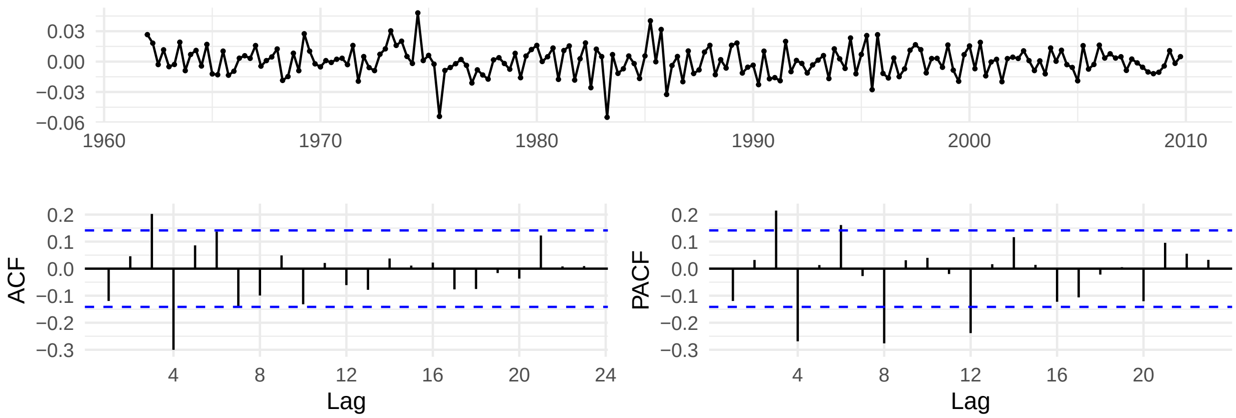

Figure 3: Residual Analysis for Log-Transformed and Differenced Consumption Data

Based on the resulting assumptions, we conlude that an ARIMA(0,1,3)(0,1,2) might be suitable, as indicated in the *Appendix* in *Figure 24 and 25.* Because of the significant spikes seen in the PACF for lag 3 and 6 we determine determine `q` = 3 and `Q` = 2. The subsequent check shows some autocorrelation left, visible in the ACF plot at spike 6. Because we already adjusted the MA(q) part, this must be a detail to be adjusted in the AR(p) part. We therefore set `p` = 1 and 2 and compare their AICc, autocorrelation, stationarity, and white noise residuals.

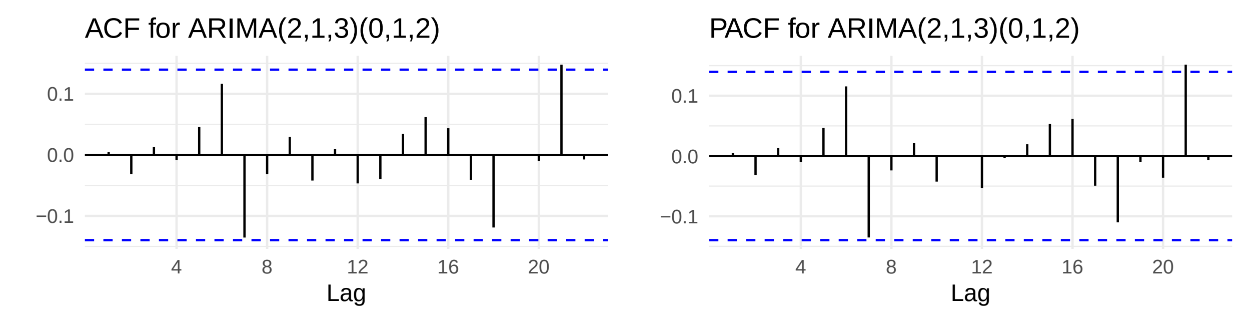

Figure 4: ACF and PACF Analysis for ARIMA (2,1,3)(0,1,2) Model

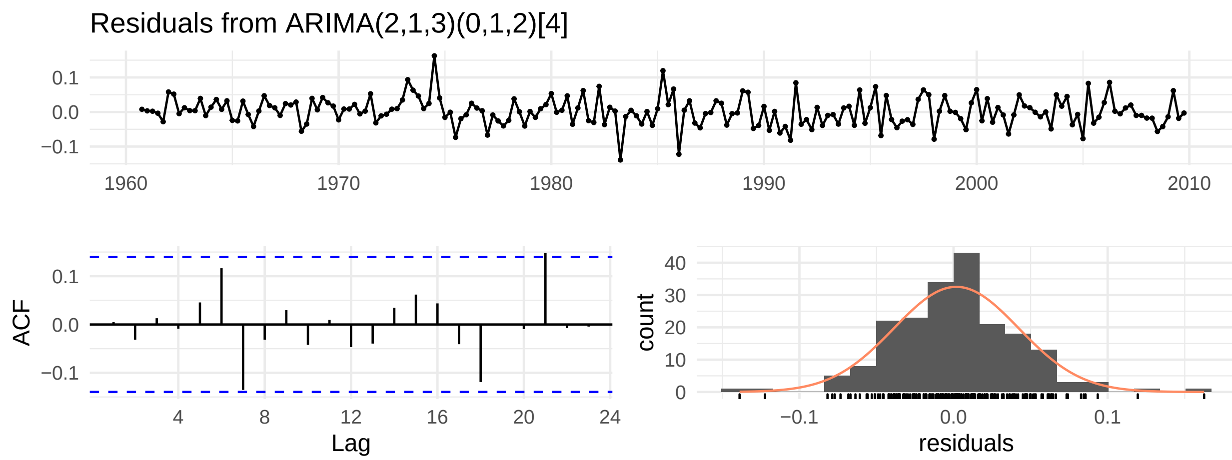

Figure 5: Residual Analysis for ARIMA (2,1,3)(0,1,2) Model

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,3)(0,1,2)[4]
## Q* = 8.0037, df = 3, p-value = 0.04594
##
## Model df: 7.   Total lags used: 10
```

Comparing our models for ARIMA(1,1,3)(0,1,2) (see *Appendix*, *Figure 26 and 27*) and ARIMA(2,1,3)(0,1,2), we see that the latter has better ACF and PACF spike conditions, illustrated in *Figure 5*. Also, the Ljung-Box test is supporting this assumptions as the autocorrelation in the ARIMA(1,1,3)(0,1,2) and ARIMA(0,1,3)(0,1,2) is still highly significant as opposed by the latter model. The resulting series is now a white-noise series. The distribution of the residuals also fits the assumed distribution pattern. Looking at the coefficients, we observe that `ar1`, `ma1`, and `sma2` are not statistically significant, and `ma3` being almost not statistically significant. Because `ar1` should not be altered because of high significance of `ar2`, we try new model variants:

```
## Series: ts.train[, 2]
## ARIMA(2,1,2)(0,1,1)[4]
## Box Cox transformation: lambda= 0.1175096
##
## Coefficients:
##          ar1      ar2      ma1     ma2     sma1
##       1.1879  -0.3847  -1.3026  0.6588  -0.7242
## s.e.  0.1902   0.2050   0.1617  0.1621   0.0676
##
## sigma^2 estimated as 0.001755:  log likelihood=338.32
## AIC=-664.65   AICc=-664.19   BIC=-645.1
```

Because in our `fit4` model (see *Appendix*, *Figure 28*) we still have some issues with autocorrelation we set `P` = 1 because of the trend that can still be observed in the residuals. We design and end up with an ARIMA(2,1,2)(1,1,1) model that also surpasses all previous models in terms of coefficient significance, and AICc, AIC, and BIC. Also the Ljung-Box p-value is maximized, the residual distibution feed the ARIMA process requirements, and the unit root theorem is also satisfied, as seen in *Figures 6 to 8*:

```
## Series: ts.train[, 2]
## ARIMA(2,1,2)(1,1,1)[4]
## Box Cox transformation: lambda= 0.1175096
##
## Coefficients:
##           ar1      ar2     ma1     ma2    sar1     sma1
##       -0.3639  -0.6482  0.2266  0.8426  0.4537  -0.7889
## s.e.   0.1512   0.1093  0.1212  0.0660  0.1431   0.1078
##
## sigma^2 estimated as 0.001724:  log likelihood=340.4
## AIC=-666.8   AICc=-666.19   BIC=-644
```

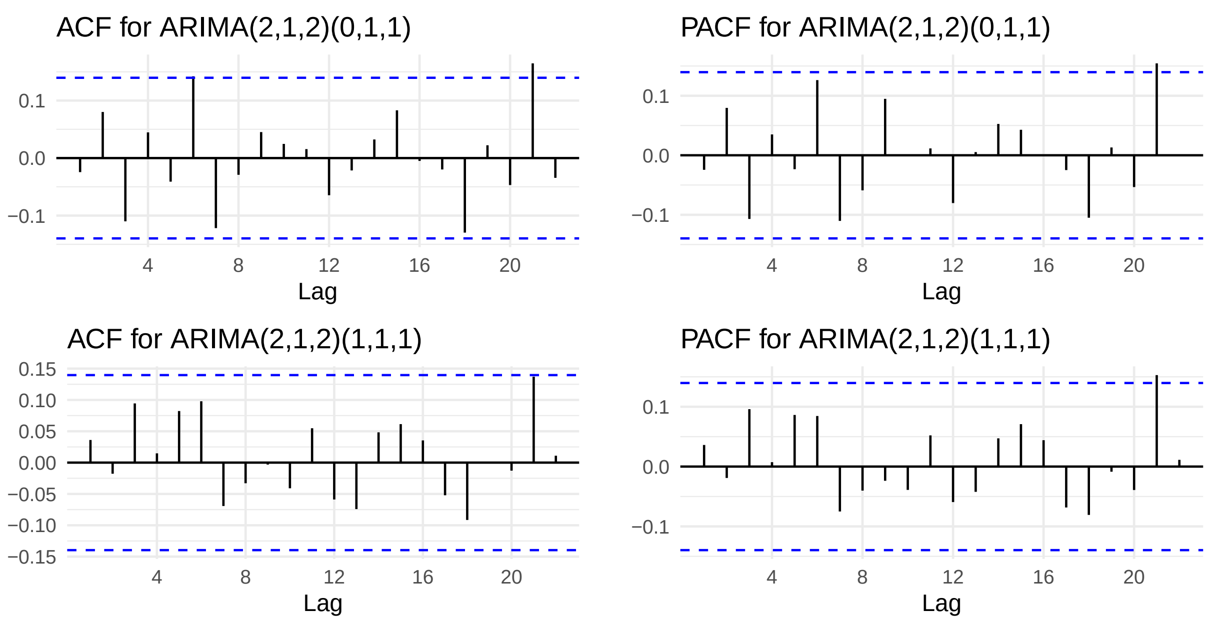

Figure 6: ACF and PACF Comparison of ARIMA(2,1,2)(0,1,1) and ARIMA(2,1,2)(1,1,1)

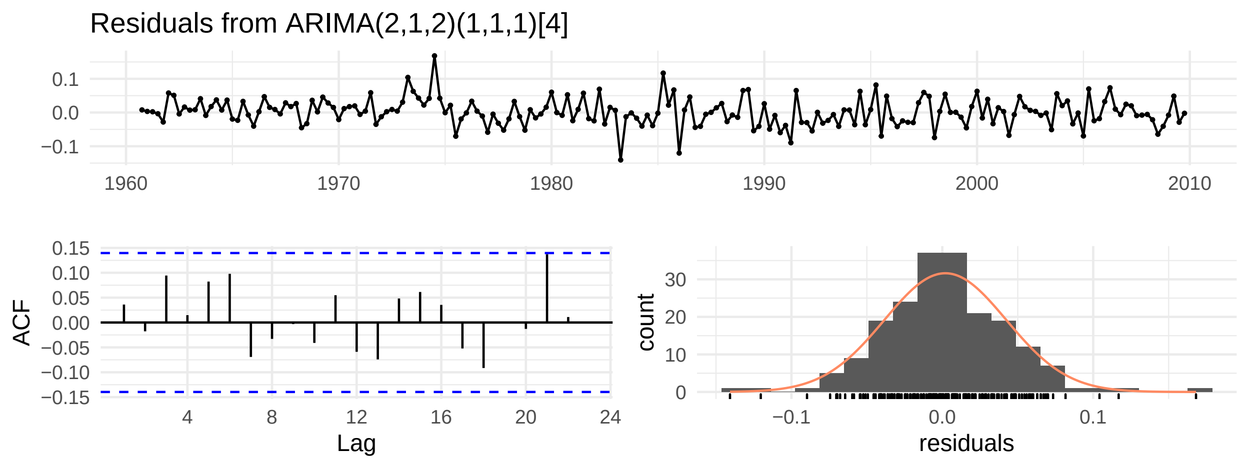

Figure 7: Residual and ACF Analysis for ARIMA(2,1,1)(1,1,1)

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,2)(1,1,1)[4]
## Q* = 6.7349, df = 3, p-value = 0.08084
##
## Model df: 6.   Total lags used: 9
```

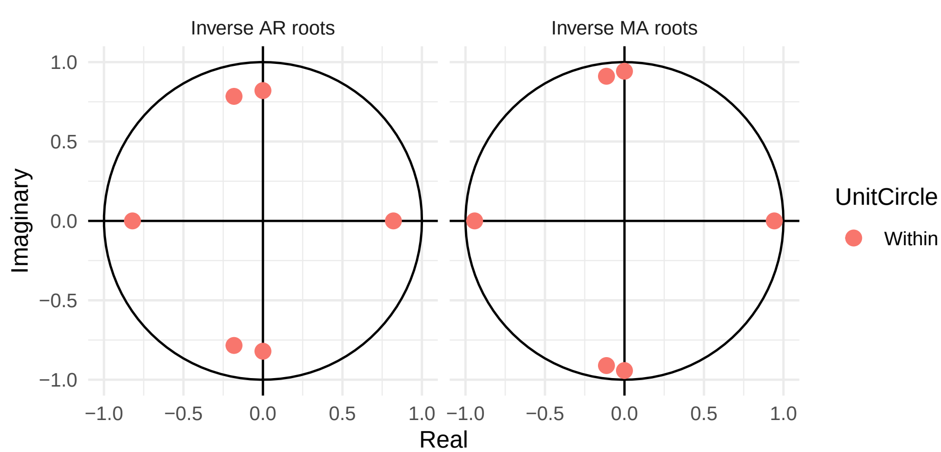

Figure 8: Unit Root Theorem Analysis for ARIMA(2,1,1)(1,1,1)

For detailed figures and graphs showing the other scenarios, please see the *Appendix*. A comprehensive evaluation of ARIMA models can also be taken from this table:

Table 1: Evaluation of ARIMA models without Regression Component

model	LB.p.value	aicc	bic
ARIMA(2,1,2)(1,1,1)	0.081	-666.195	-644.001
ARIMA(2,1,3)(0,1,2)	0.046	-663.645	-638.372
ARIMA(1,1,3)(0,1,2)	0.024	-664.072	-641.878
ARIMA(2,1,2)(0,1,1)	0.007	-664.195	-645.104
ARIMA(1,1,1)(0,1,2)	0.002	-660.958	-641.867

Automated ARIMA Selection Approach

Because KPSS can only be used to determine `d` and `D`, we need to employ Information Criteria, such as AICc, to pick the correct `p`,`q`,`P`,`Q` values. This is already incorporated in the automated ARIMA model selection that calculates different ARIMA models and picks the best models based on those Informaiton Criteria. In fact, the same ARIMA(2,1,2)(1,1,1) is picked, based on the unit root space optimization to guarantee stationarity.

Comparison of ACF and PACF with ARIMA and Raw Consumption Expenditure

Comparing the ACF and PACF plots of the raw Final consumption Expenditure and ARIMA data, we can observe the following: 1) The autocorrelation in the residuals is resolved. This was important to resolve, as ARIMA assumes that historical patterns will not change during the forecast. 2) The issue of a high PACF spike at lag 1, indicating correlation between the error terms of consumption between different lags was resolved. This is important to resolve, because ARIMA assumes uncorrelated future errors.

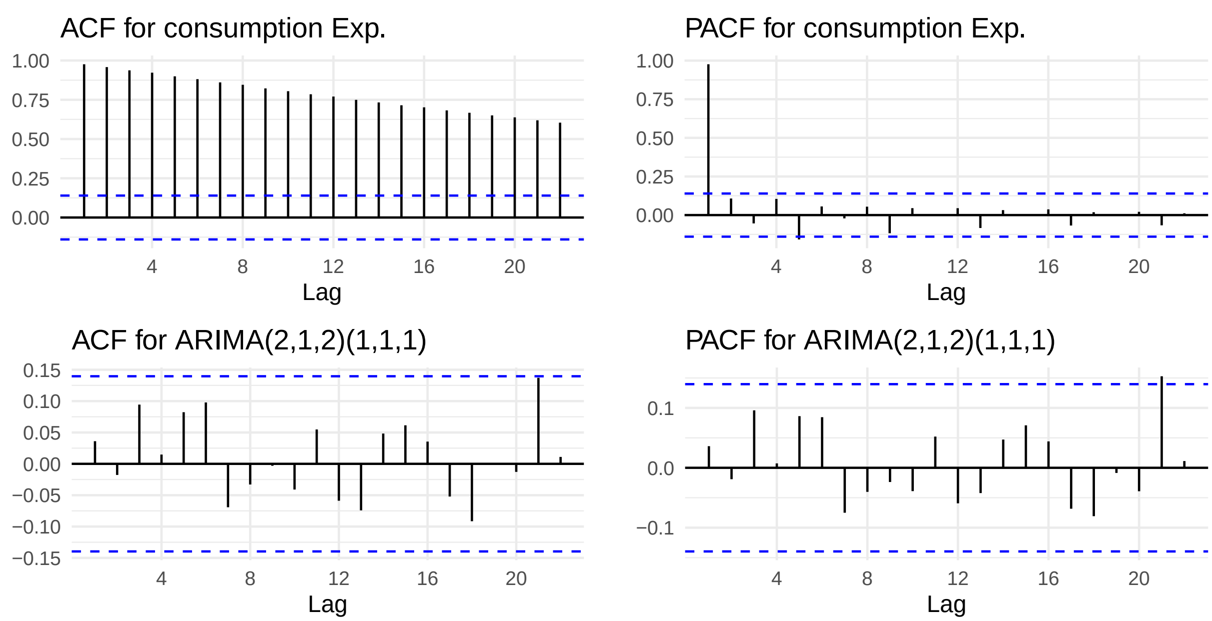

Figure 9: ACF and PACF Comparison between Original Data and ARIMA TRANSFORMED Results

ARIMA Forecast on Test Data Set

As can bee seen in *Figure 10*, the manual ARIMA model fits the data quite well, despite some minor overestimation. The prediction interval increases in size throughout time because of the included differences method.

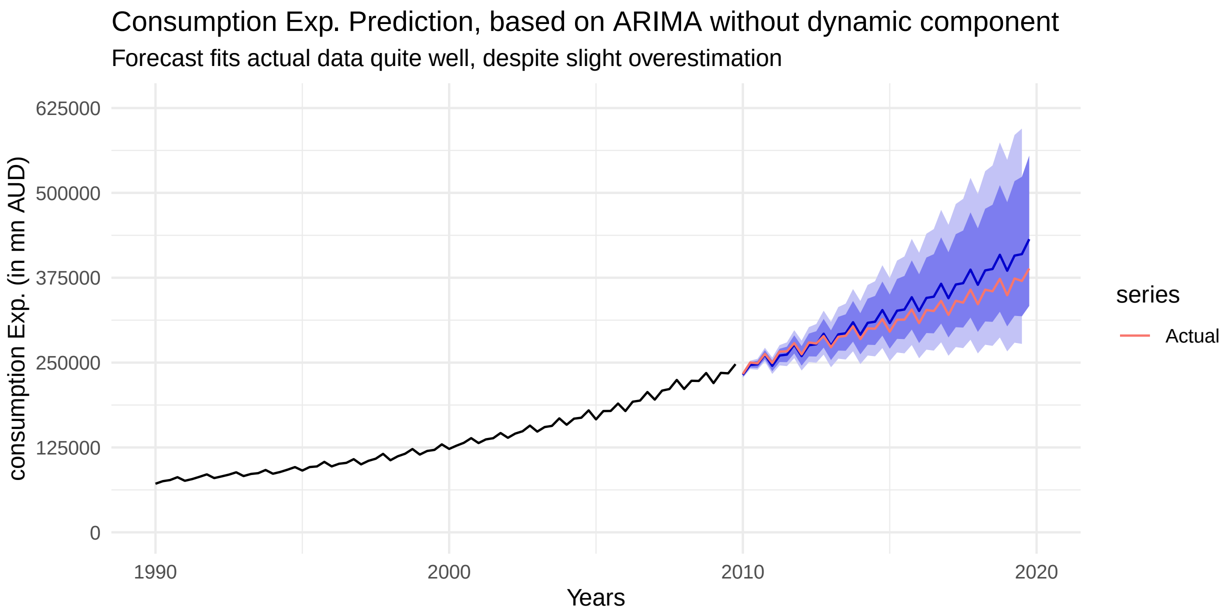

Figure 10: Predicted Australian Consumption Expenditure Based on ARIMA(2,1,2)(1,1,1) Model

Dynamic Regression with Explanatory Variable

Manual Dynamic Regression Model Selection

The inclusion of a new explanatory variable in the ARIMA model requires us to check the errors terms of the regression model (eta) and our ARIMA model (epsilon). In our case, our two variables for consumption and income are cointegrated. That's why we can rely on non-stationary time series (Hyndman & Athanasopoulos, 2018). In our first model, that is already adjusted with $\texttt{d} = \texttt{D} = 1$, as we observed with the KPSS test before in order to guarantee non-stationarity of the data, we still observe significant ACF spikes for lag 1,3, and 4 (see *Appendix, Figure 29*), suggesting a `Q`-value of 1. PACF spikes for lag 1,3, and 4 also indicate that $\texttt{P} = 1$. Coefficients and AICc, and BIC values will be showed at the end.

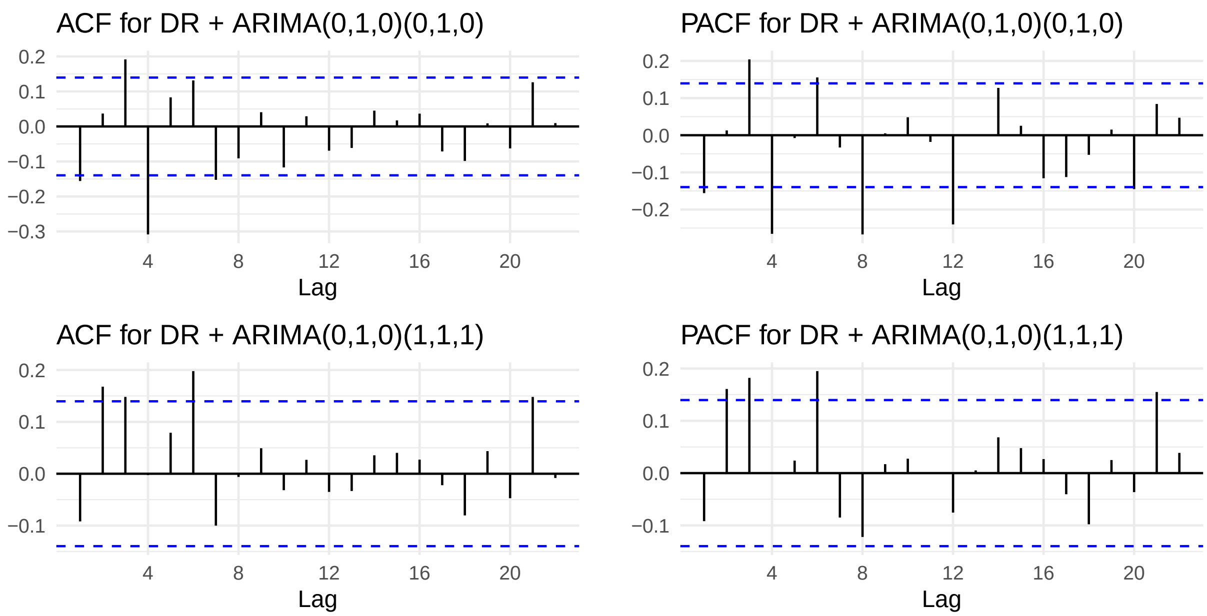

Figure 11: ACF and PACF Analysis Comparison for ARIMA(0,1,0)(0,1,0) and ARIMA(0,1,0)(1,1,1)

This setting shows us significant ACF and PACF spikes for lags 2 and 3 as well as 6 (see *Figure 11*) and potential for improvement for the distribution of residuals (see *Appendix, Figure 30*). We set $\mathtt{p} = \mathtt{q} = 2$, as in the previous model to balance the ACF and PACF values against each other. This results in optimal models considering ACF/PACF and white-noise behaviour, residual distribution, heteroscedasticity, stationarity, and coefficient significance (see *Figure 12* and *Figure 13*). When looking at the coefficients, we observe that `ma1` and `xreg`, are not significant, but we include the latter because of the task. We do not delete `ma1`, as this would impact the significant `ma2` coefficient and because of the needed transformation towards autocorrelation decrease.

```
## Series: ts.train[, 2]
## Regression with ARIMA(2,1,2)(1,1,1)[4] errors
## Box Cox transformation: lambda= 0.1175096
##
## Coefficients:
##           ar1      ar2     ma1     ma2    sar1     sma1  xreg
##       -0.3616  -0.6468  0.2254  0.8424  0.4516  -0.7876     0
## s.e.   0.1287   0.0902  0.1156  0.0646  0.0976   0.0926   NaN
##
## sigma^2 estimated as 0.001732:  log likelihood=340.5
## AIC=-665   AICc=-664.21   BIC=-638.94
```

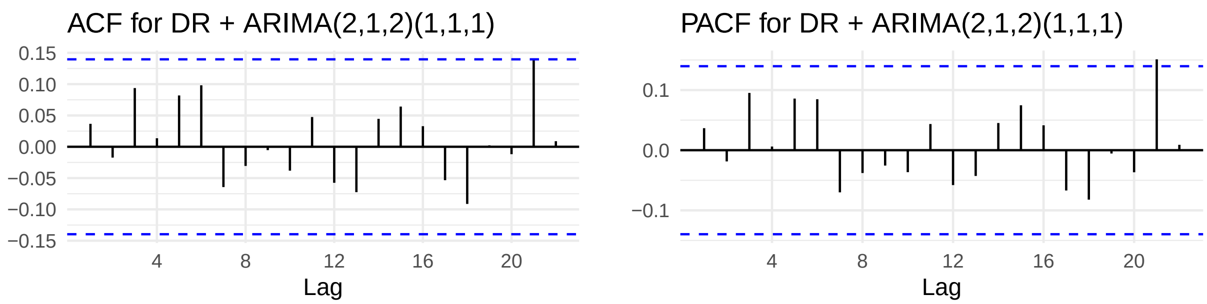

Figure 12: ACF and PACF Analysis for ARIMA(2,1,2)(1,1,1)

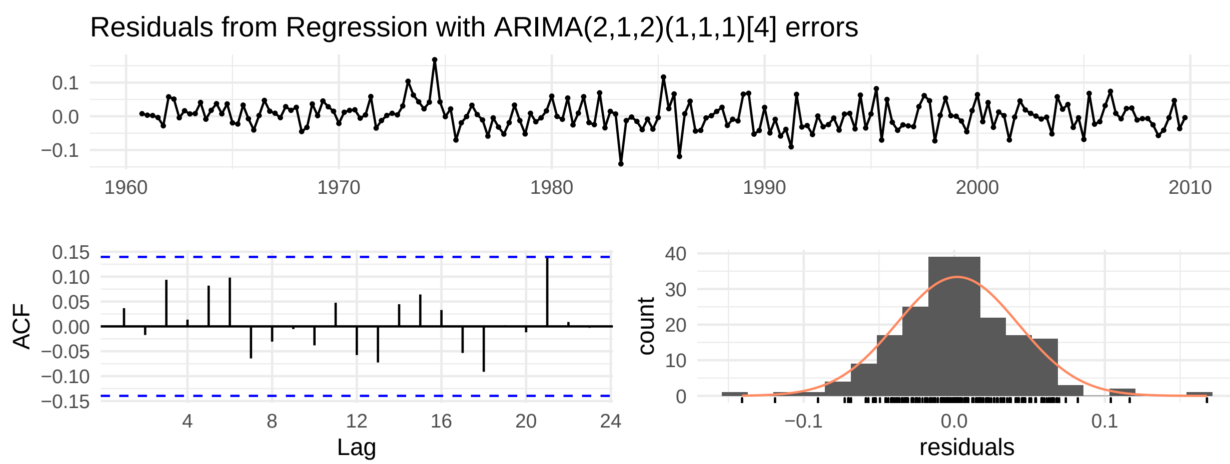

Figure 13: Residual Analysis for ARIMA(2,1,2)(1,1,1)

```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(2,1,2)(1,1,1)[4] errors
## Q* = 6.8499, df = 3, p-value = 0.07684
##
## Model df: 7.   Total lags used: 10
```

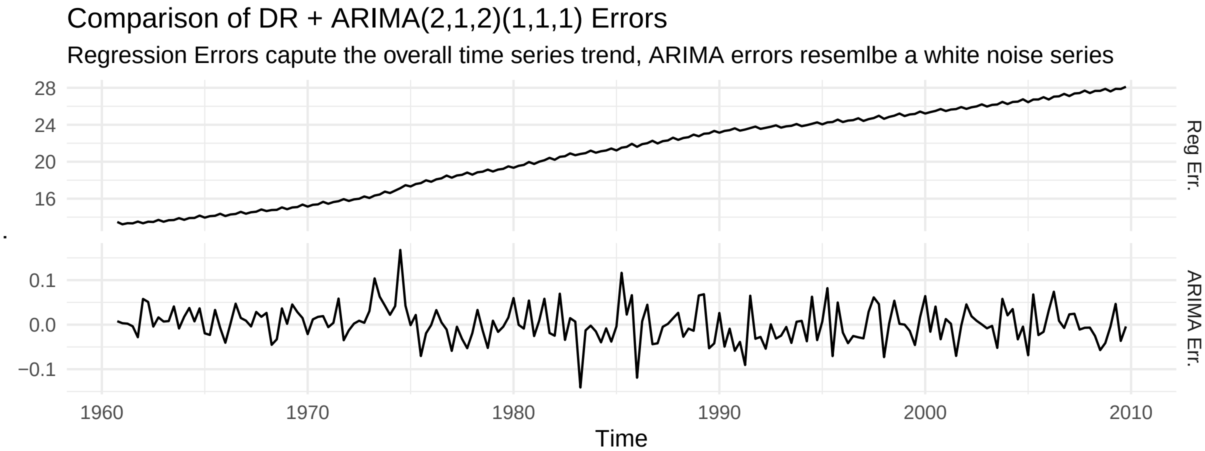

Figure 14: Error Comparison

This reestimation yields in a suitable model, considering the white noise type of ARIMA residuals, ACF and PACF specifics, as well as a fitting residual distribution that is only slightly skewed because of the observation outliers during the financial crisis in 2007/08. Additionally, we could include a constant in order to mimick the trend that is displayed in our regression residuals (see *Figure 14*). For this, and because a drift cannot be included if the order of difference > 2, we must set d = 0 and also q = 0, because this drift should explain the information conveyed in the regression residuals. But because this change yields in more autocorrelation, we refrain from doing so.

Automated ARIMA Model Selection

On the other side, the automated approach yields in a different model variation, that was already discussed above but discarded because of its negative impact on ACF and PACF plots and white-noise properties. It yields a lower AICc and does not yield in autocorrelation reduction, as seen in *Figure 15*. In sum, the automated dynamic regression model is a worse forecast model than our ARIMA(2,1,2)(1,1,1) model, as can be seen in *Figure 16*.

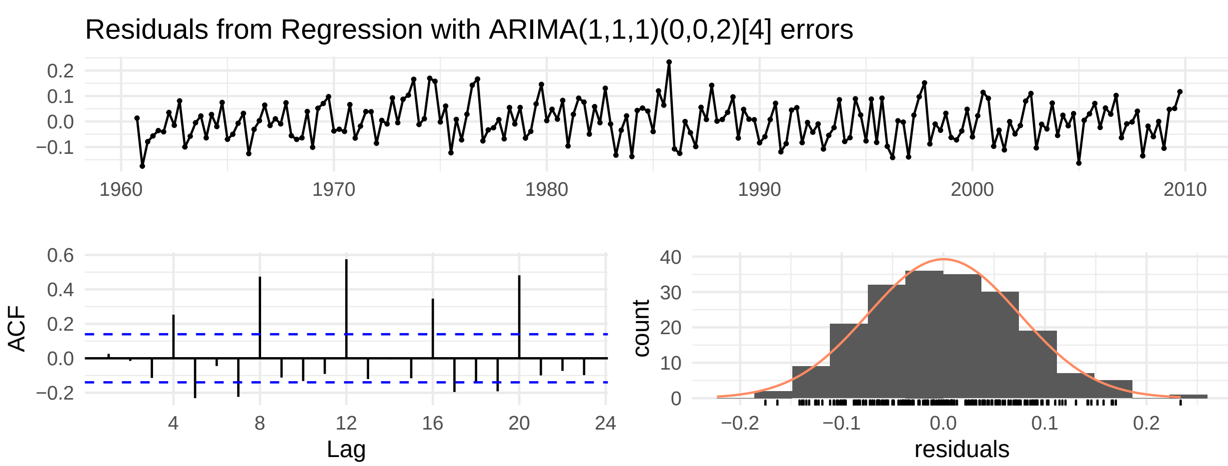

Figure 15: Residual Analysis for Automated Selected ARIMA model

```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(1,1,1)(0,0,2)[4] errors
## Q* = 86.72, df = 3, p-value < 2.2e-16
##
## Model df: 6.   Total lags used: 9
```

Table 2: Evaluation of ARIMA models with regression

model	LB.p.value	aicc	bic
ARIMA(2,1,2)(1,1,1)	0.077	-664.214	-638.941
ARIMA(0,1,0)(1,1,1)	0.000	-654.671	-641.855
ARIMA(0,1,0)(0,1,0)	0.000	-618.404	-611.953
ARIMA(1,1,1)(0,0,2)	0.000	-441.553	-419.202

Figure 16: Forecast Performance Comparison between Manual and Automated Approach

Forecast Combination: Comparison, Plotting and Measurment

In the following part, we will combine different forecasting models and assess its performance. Specifically, we will test 3 different cases in forecasting combination:

1. ARIMA model combined with the dynamic regression(DR). In this scenario, we will combine our best performing ARIMA model (in our case, it is `ARIMA(2,1,2)(1,1,1)[4]`) and same ARIMA model with the specified `xreg` parameter.
2. AutoARIMA with regression model and manually picked ARIMA with regression. Previously, the automated model has been compared with the manually adjusted one, and we have seen the advantage of the manual approach. Yet, it is interesting to see the performance of the combined model and to compare it to other variations of the forecast combination.
3. Manual ARIMA with regression and TBATS/ETS In this case, we want to combine the best performing ARIMA model with other forecasting approaches. In this case, for the comparison, we decided to compare the combination with ETS and TBATS models.

Every forecast combination has two variations: averaged forecast combination and using optimal weights. Optimal weights have been calculated by the following formulas: $w = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$ and $1 - w = \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$

Table 3 below provides the comparison of both sole and combined models and measures of ME, RMSE and MAE for the test set. In the performance assessment and comparison, we are primarily looking at the RMSE, therefore, the ordered comparison of the model performance can be seen in the *Figure 20* below.

Table 3: Evaluation of forecasting of sole and combined models

	ME	RMSE	MAE
ARIMA(2,1,2)(1,1,1)	-13759.711	19936.764	15383.120
ARIMA(1,1,1)(0,0,2) with reg	-43648.055	53350.354	43648.055
ARIMA(2,1,2)(1,1,1) with reg	-12602.044	18709.743	14461.070
ETS	10403.266	11609.098	10403.266
TBATS	-26308.101	45426.611	29673.195
Averaged ARIMA+DR	-13180.878	19321.676	14922.095
Optimal weights: ARIMA+DR	-13174.158	19314.554	14916.743
Averaged: DRs(auto+manual)	-28125.049	35716.730	28273.289
Optimal: DRs(auto+manual)	-24235.763	31356.277	24496.598
Averaged: DR + ETS	-1099.389	5187.795	4445.283
Optimal weights: DR + ETS	3370.362	4198.252	3444.653
Averaged: DR + TBATS	-19455.072	31831.652	21922.485
Optimal weights: DR + TBATS	-17181.374	27391.689	19445.203

As for the first combination case, where the ARIMA model is combined with the same model with regression, we can observe that the RMSE for the sole models is comparably close. So, in fact, there is no direct need to combine the forecasts, as the performance of the combined models and sole models will differ just slightly.

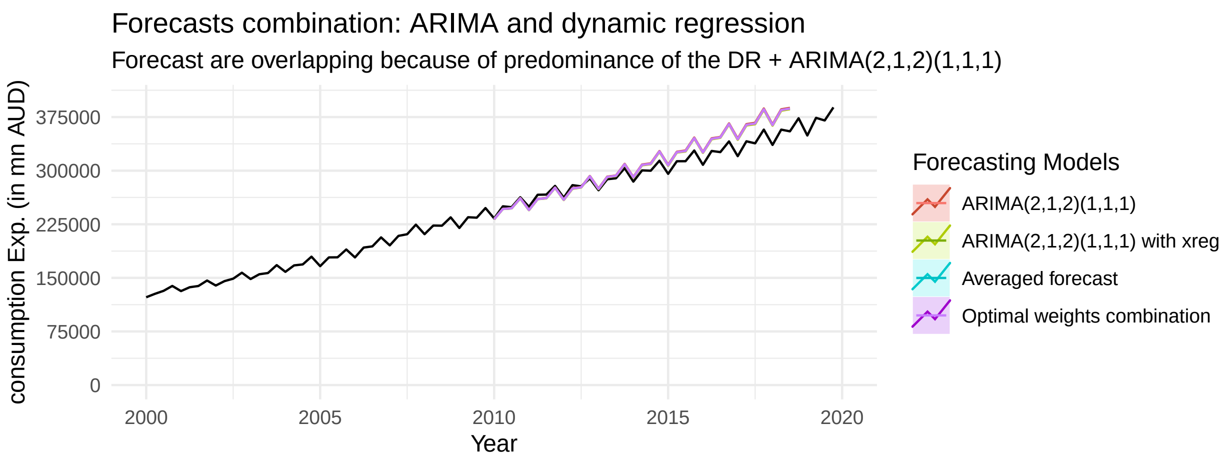

Figure 17: Forecast Combination for Dynamic Regressions using ARIMA

From the *Figure 17*, we can see that all of the forecasting models perform almost equally as good, with just a slight advantage of the ARIMA(2,1,2)(1,1,1) with regression. This can be explained with the fact that both models are providing forecasts with estimations "above" the actual forecast and with a slight difference comparing to the models we will observe in the following two cases. Also, the 4 forecast are incredibly similar because they are all based on ARIMA(2,1,2)(1,1,1) model.

For the second combination, we decided to combine `auto.arima` with regression and manually adjusted model with regression, used in the previous case. As we previously observed, the difference between both cases is observable with the measured RMSE of 19351.334 and 52235.143, for the manual and auto models correspondingly.

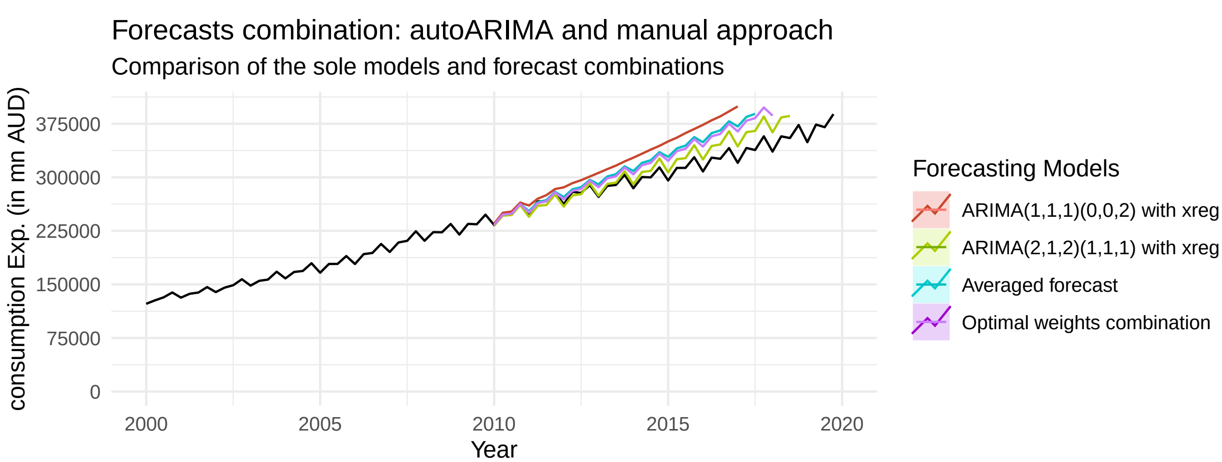

Figure 18: Forecast Combination for Dynamic Regressions and Automated Regression

From the *Figure 18*, presented above, we can observe the similar pattern as in the previous case: combination model with optimal weights performs slightly better than the averaged model, yet the sole manually adjusted model is still seen as the best performer due to both forecasts laying above the actual data of the test set.

In the previous cases, we have looked into the combination of the different variations of the ARIMA models that we have estimated in the for the previous tasks. All of them share similar characteristics being above the test set in terms of forecast and primarily varying in the "closeness" to the actual data and capturing of the seasonality (especially visible in the last case). However, it is interesting to see the difference between the ARIMA models and other estimation approaches. For this case, we have made forecasting using TBATS and ETS models both separately combining with the best performing ARIMA model with regression.

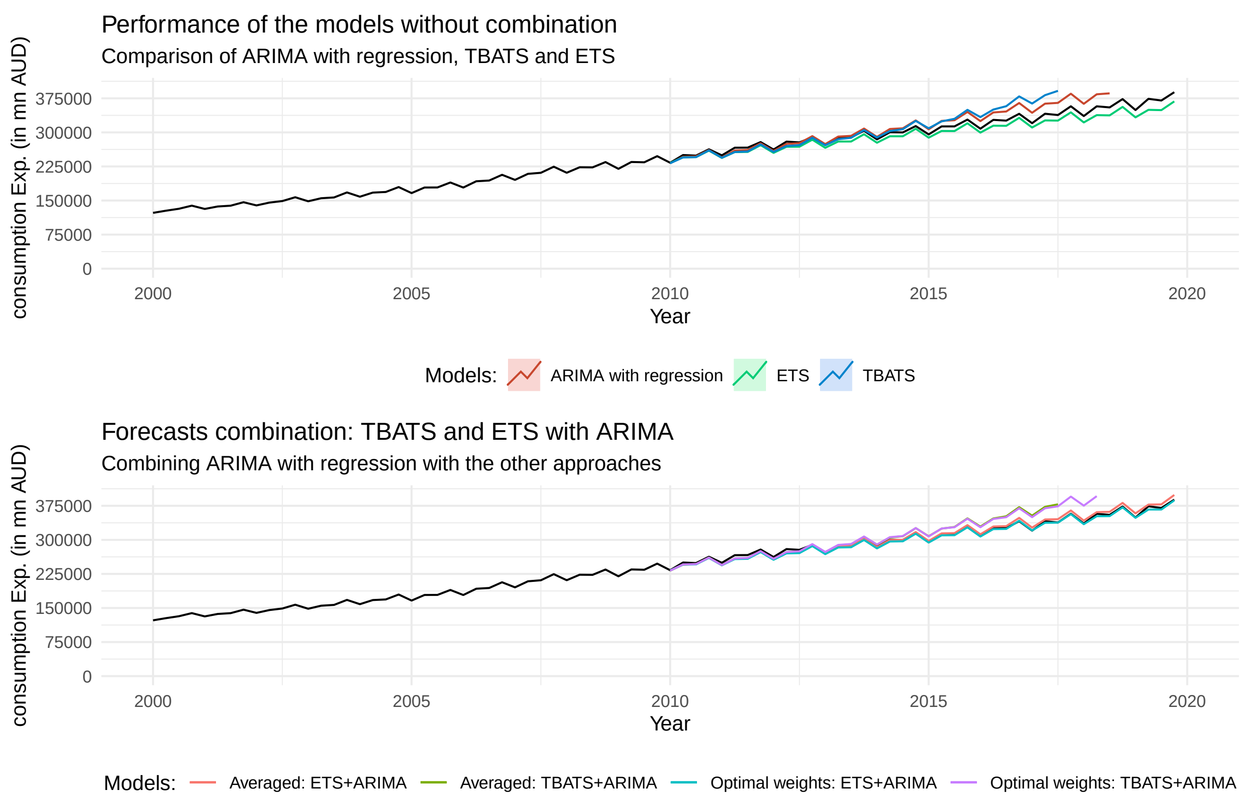

Figure 19: Forecast Comparison for ETS, TBATS Approaches

From *Figure 19* above, we can see that TBATS model and ARIMA with regression are very similar in its estimations, which is logical as far as TBATS is similar to the dynamic harmonic regression in many aspects, apart from seasonality. Different estimation characteristics are observed for the ETS (M,A,M): the model itself performs very well for based on the training set, and we can see that it captures data trend and seasonality quite well. Apart from that, ETS estimations are mapped slightly below the actual data, which should be very beneficial in the combination of two models.

The results of the combined models can be found in the lower part of the *Figure 19*. High performance of the forecast combination can be observed: a combination of the ARIMA with ETS has the best results among all models. Applying the use of the optimal weights, allows us to enhance the model even further, producing the overall RMSE of 4098.904. TBATS model forecast combination with ARIMA is weaker, due to the aforementioned estimation characteristics of TBATS. In the *Figure 20* below, we can see that overall TBATS model, as with many automated modelling frameworks, performs rather poorly with just a slighter better estimation, than autoARIMA.

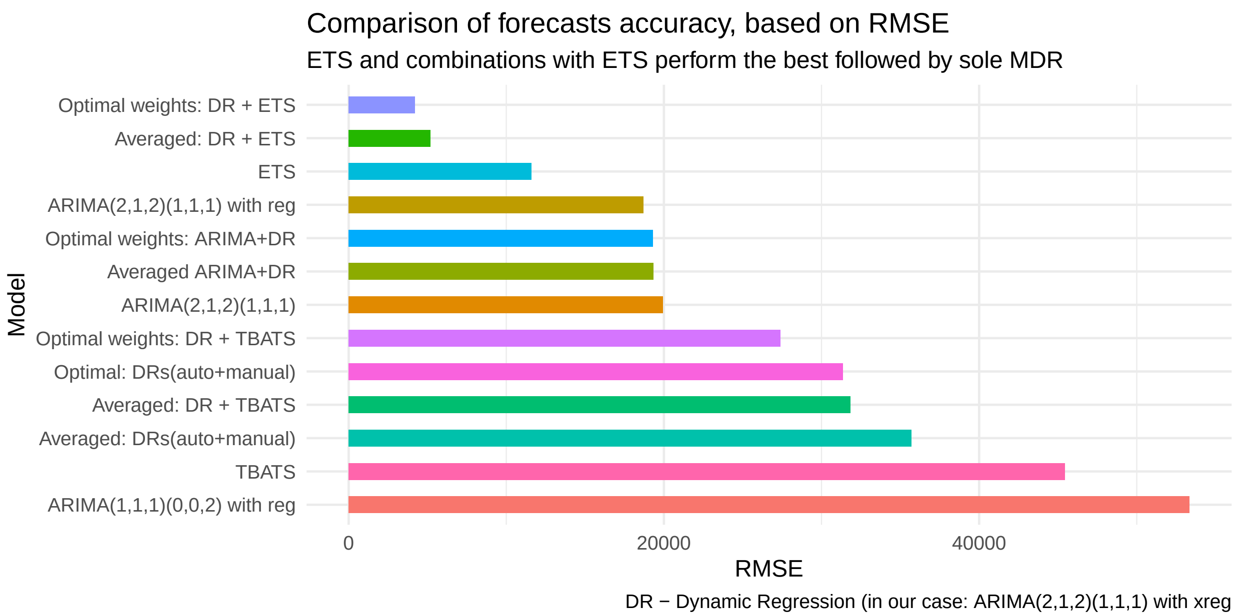

Figure 20: Forecast Accuracy Comparison

From *Figure 20*, we can see the overall comparison of all the models (combined and sole) performance based on the RMSE. Noticeably, the combination of two different best performing sole models (ARIMA with regression and ETS), provides us with the best results. In particularly this case, the poor performance of the automated modelling frameworks is visible with TBATS and autoARIMA having the biggest RMSE. In all the cases of forecasts combination, the advantage of the application of the optimal weights over averaging can be mentioned.

Appendix

```
##
## ###############################################
## # 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
## -27208.4  -3436.7    469.6   3483.9  25030.2
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.142e+03  1.497e+03  -0.763    0.446
## z.lag.1     -1.328e-03  1.332e-02  -0.100    0.921
## tt           3.217e+01  2.212e+01   1.455    0.147
## z.diff.lag  -5.304e-01  5.595e-02  -9.478   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8560 on 236 degrees of freedom
## Multiple R-squared:  0.2943, Adjusted R-squared:  0.2853
## F-statistic:  32.8 on 3 and 236 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -0.0997 11.7072 7.0162
##
## 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
```

```
##
##  Box-Ljung test
##
## data:  tse[, 1]
## X-squared = 2099.3, df = 10, p-value < 2.2e-16
```

```
##
##  Box-Ljung test
##
## data:  tse[, 2]
## X-squared = 2106.5, df = 10, p-value < 2.2e-16
```

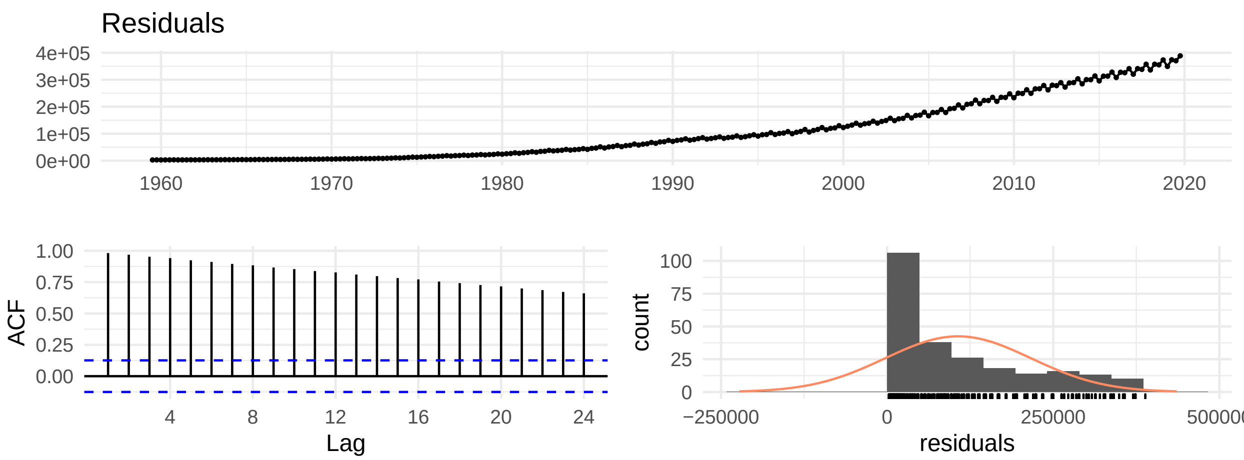

Figure 21: Residual Analysis for Income

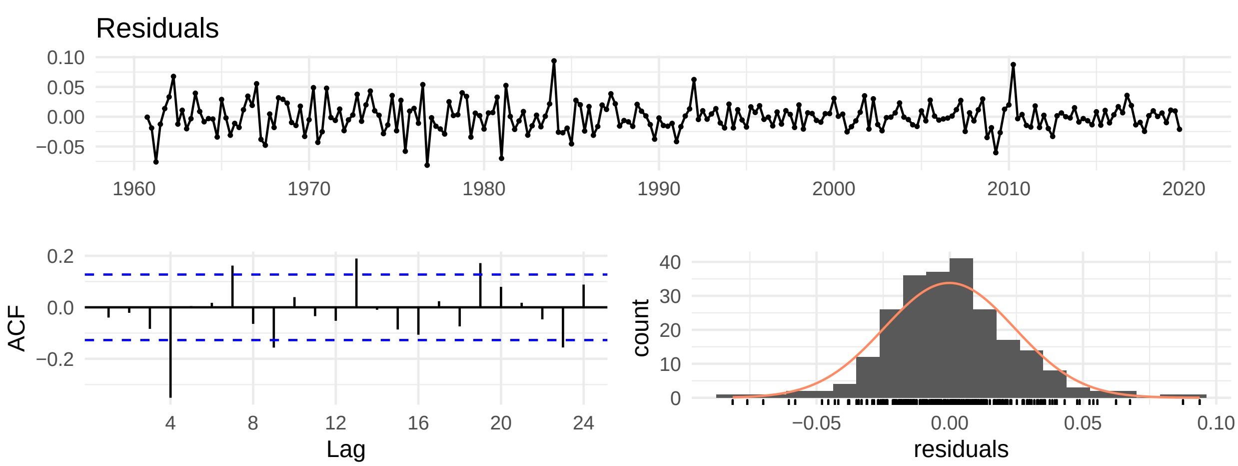

Figure 22: Residual Analysis for First-Order Seasonal Differencing with Box-Cox for Income

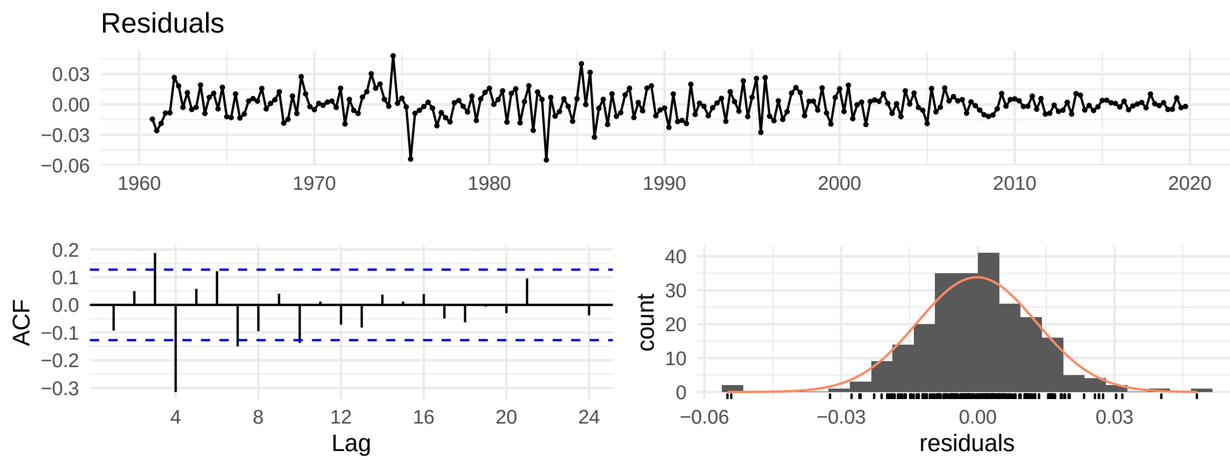

Figure 23: Residual Analysis for First-Order Seasonal Differencing for Consumption

```
## Series: ts.train[, 2]
## ARIMA(0,1,3)(0,1,2)[4]
## Box Cox transformation: lambda= 0.1175096
##
## Coefficients:
##           ma1     ma2     ma3     sma1     sma2
##       -0.0972  0.1924  0.0735  -0.4897  -0.1955
## s.e.   0.0733  0.0793  0.0691   0.0708   0.0695
##
## sigma^2 estimated as 0.001785:  log likelihood=336.71
## AIC=-661.41   AICc=-660.96   BIC=-641.87
```

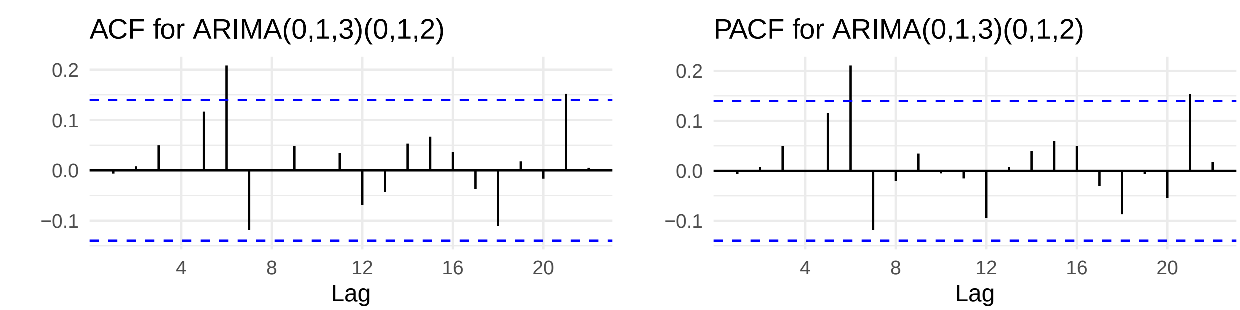

Figure 24: Residual Analysis for First ARIMA(0,1,3)(0,1,2)

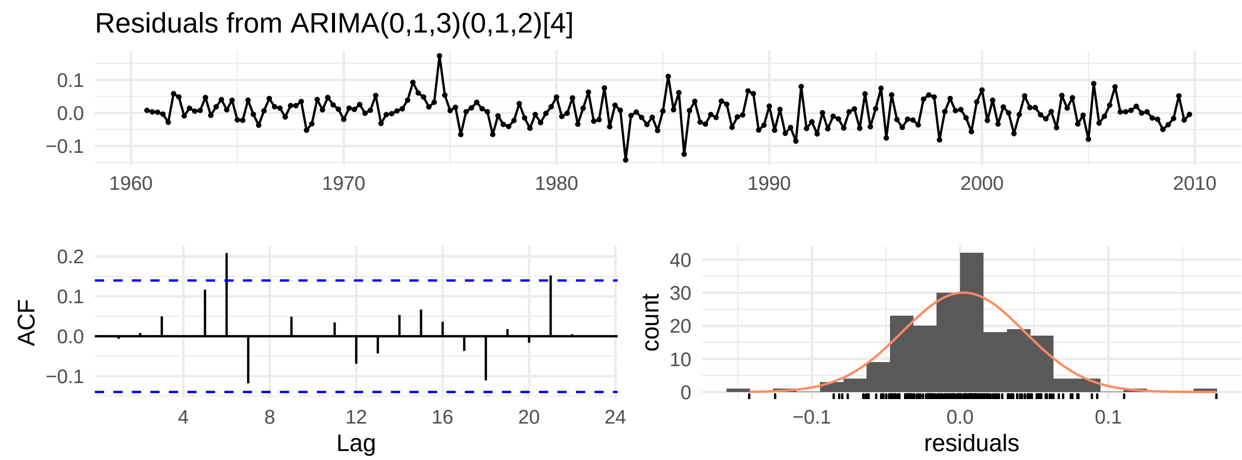

Figure 25: Residual Analysis for First ARIMA(1,1,3)(0,1,2) Model

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,3)(0,1,2)[4]
## Q* = 15.086, df = 3, p-value = 0.001745
##
## Model df: 5.   Total lags used: 8

## [1] 0.00174501

## Series: ts.train[, 2]
## ARIMA(1,1,3)(0,1,2)[4]
## Box Cox transformation: lambda= 0.1175096
##
## Coefficients:
##          ar1      ma1     ma2      ma3     sma1     sma2
##       0.8425  -0.9554  0.2893  -0.0180  -0.6105  -0.1573
## s.e.  0.1119   0.1287  0.0964   0.1015   0.0862   0.0830
##
## sigma^2 estimated as 0.001743:  log likelihood=339.34
## AIC=-664.68   AICc=-664.07   BIC=-641.88
```

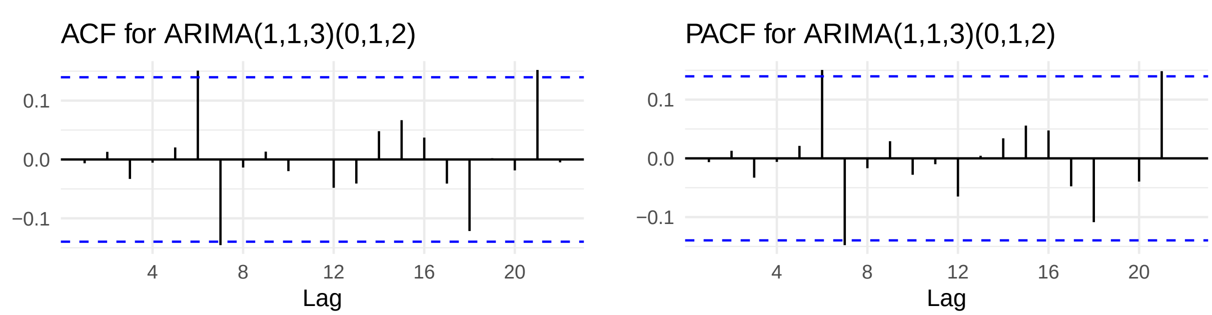

Figure 26: PACF & ACF Analysis for Second ARIMA(1,1,3)(0,1,2) Model

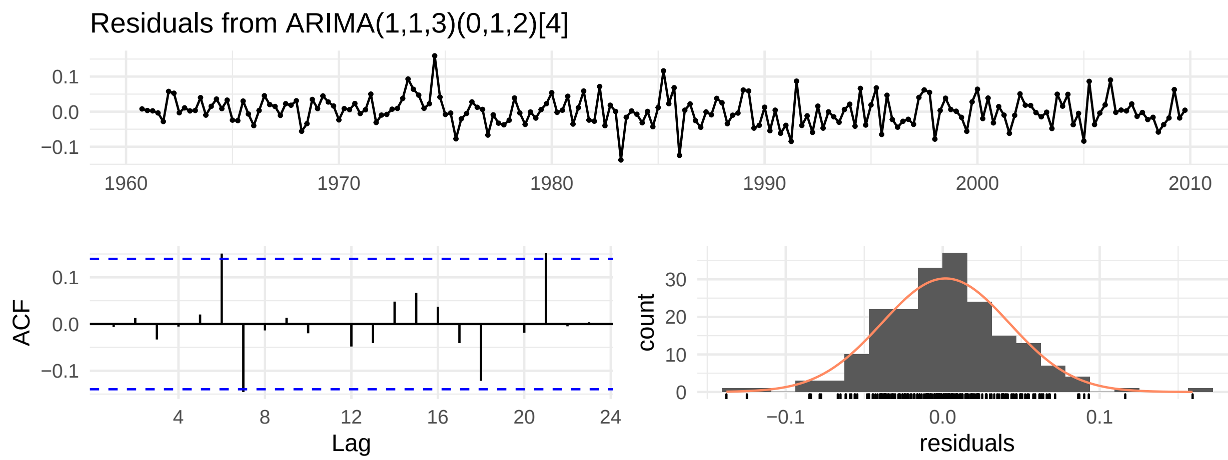

Figure 27: Residual Analysis for Second ARIMA(1,1,3)(0,1,2) Model

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,3)(0,1,2)[4]
## Q* = 9.4732, df = 3, p-value = 0.02362
##
## Model df: 6.   Total lags used: 9
```

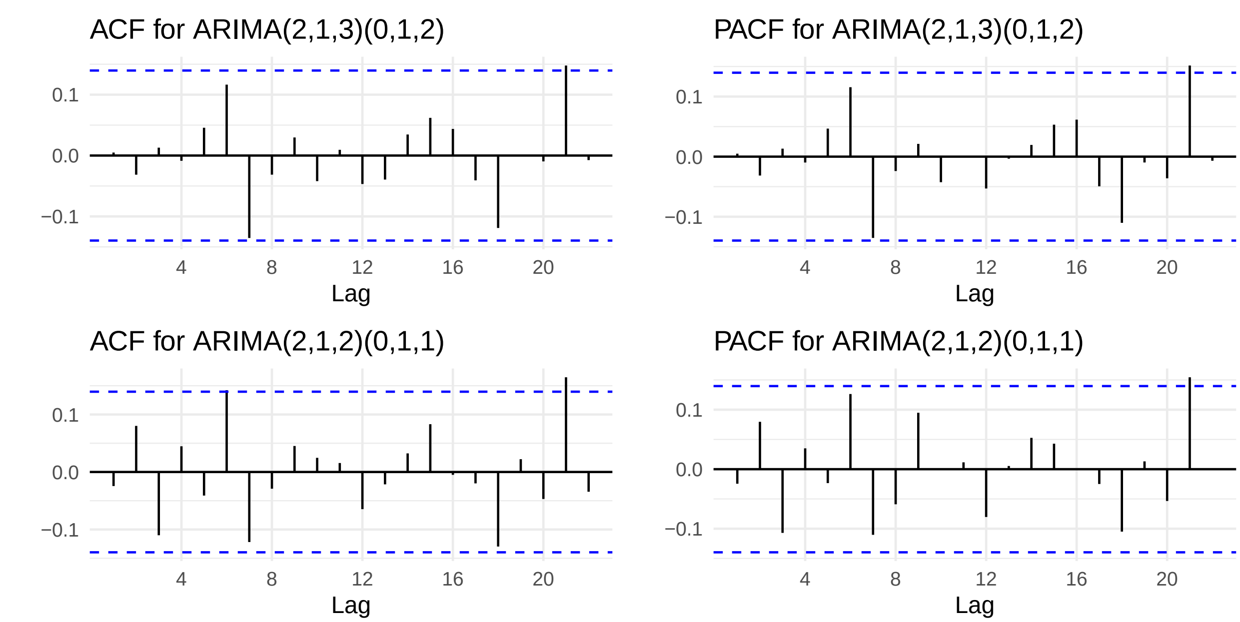

Figure 28: PACF & ACF Analysis for Third ARIMA(2,1,3)(0,1,2) and Fourth ARIMA(2,1,2)(0,1,1) Model

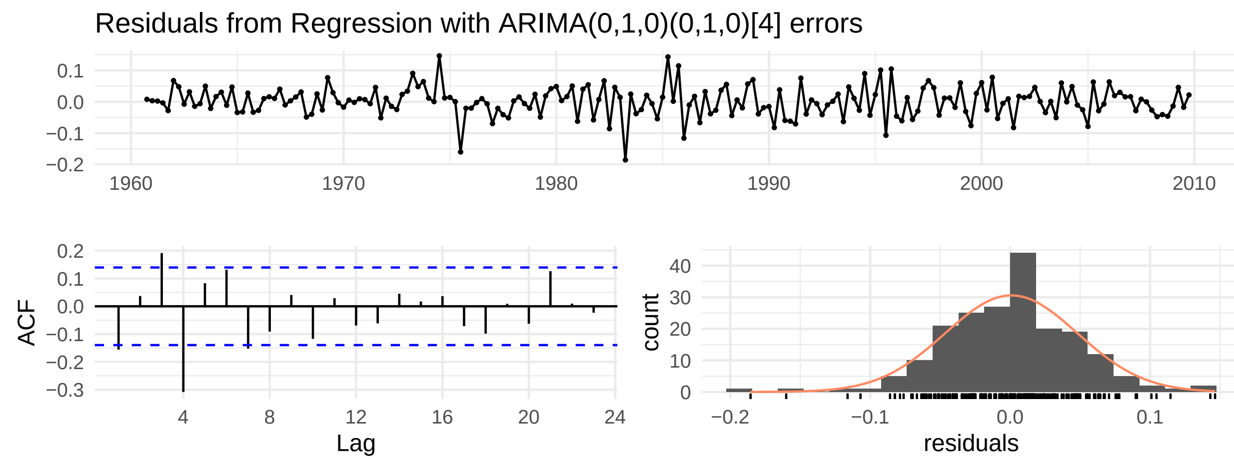

Figure 29: Residual Analysis for First DR + ARIMA(0,1,0)(0,1,0) Model

```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(0,1,0)(0,1,0)[4] errors
## Q* = 43.333, df = 7, p-value = 2.877e-07
##
## Model df: 1.   Total lags used: 8
```

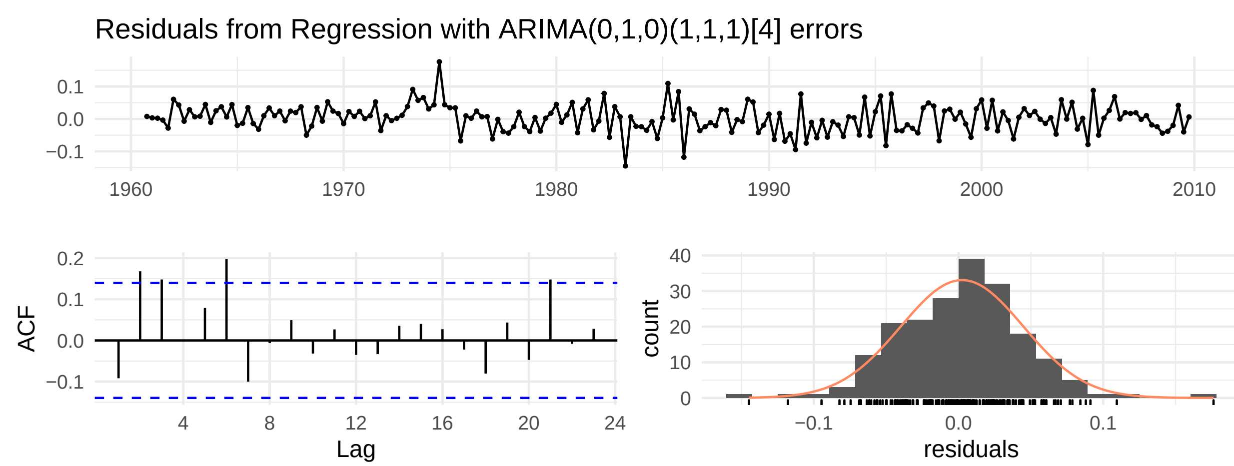

Figure 30: Residual Analysis for Second DR + ARIMA(0,1,0)(1,1,1) Model

```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(0,1,0)(1,1,1)[4] errors
## Q* = 23.175, df = 5, p-value = 0.0003126
##
## Model df: 3.   Total lags used: 8
```

References

- Enders, Walter (2014) Applied Econometric Time Series, 4th Edition, ISBN: 978-1-118-80856-6, Accessed on 01.04.2020
- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 29.03.2020