# Retail Project

Ishaan Gupta (29735491)

30th May 2021

## Library

```
library(tidyverse)
library(fpp3)
library(stringr)
library(readabs)
library(gridExtra)
```

## Preliminary analysis

In NSW, we can see that monthly turnover has progressively increased over time. Prominent peaks can also be seen and happen at the same time every year. These features suggest there is a trend and seasonality in the data.

The variation in the seasonal pattern appears to be proportional to the trend of the time series, which infers the components of the time series are multiplicative.

## Turnover – Other Recreational Goods (NSW)
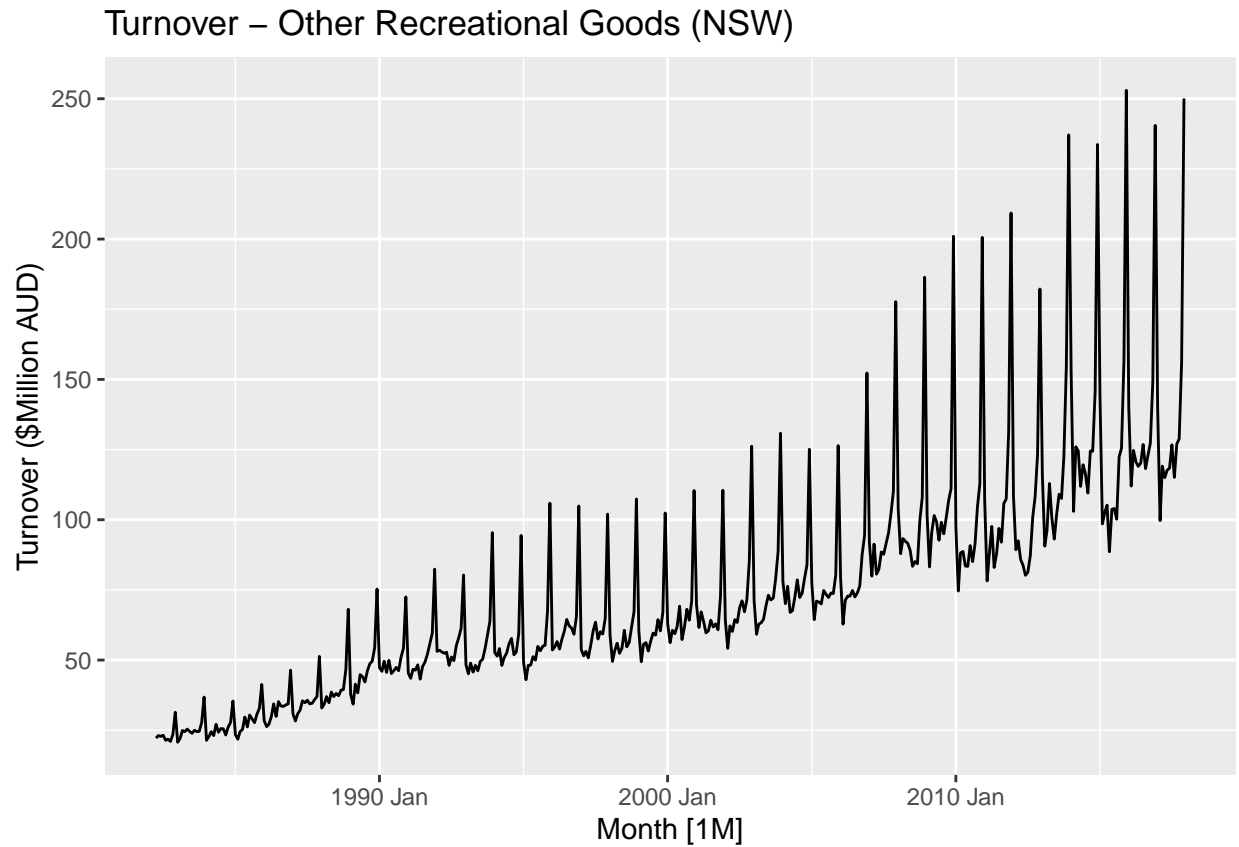


We can obtain a multiplicative decomposition by first log-transforming the data and then back-transforming the components. Then, we can run the STL decomposition. The output of STL decomposition is shown in the below plot. According to the scales, we can see the largest variation in the trend component, followed by the seasonal component and then the remainder component. This implies that trend has the largest impact on turnover.

STL decomposition

`box_cox(Turnover, x)` = trend + season_year + remainder
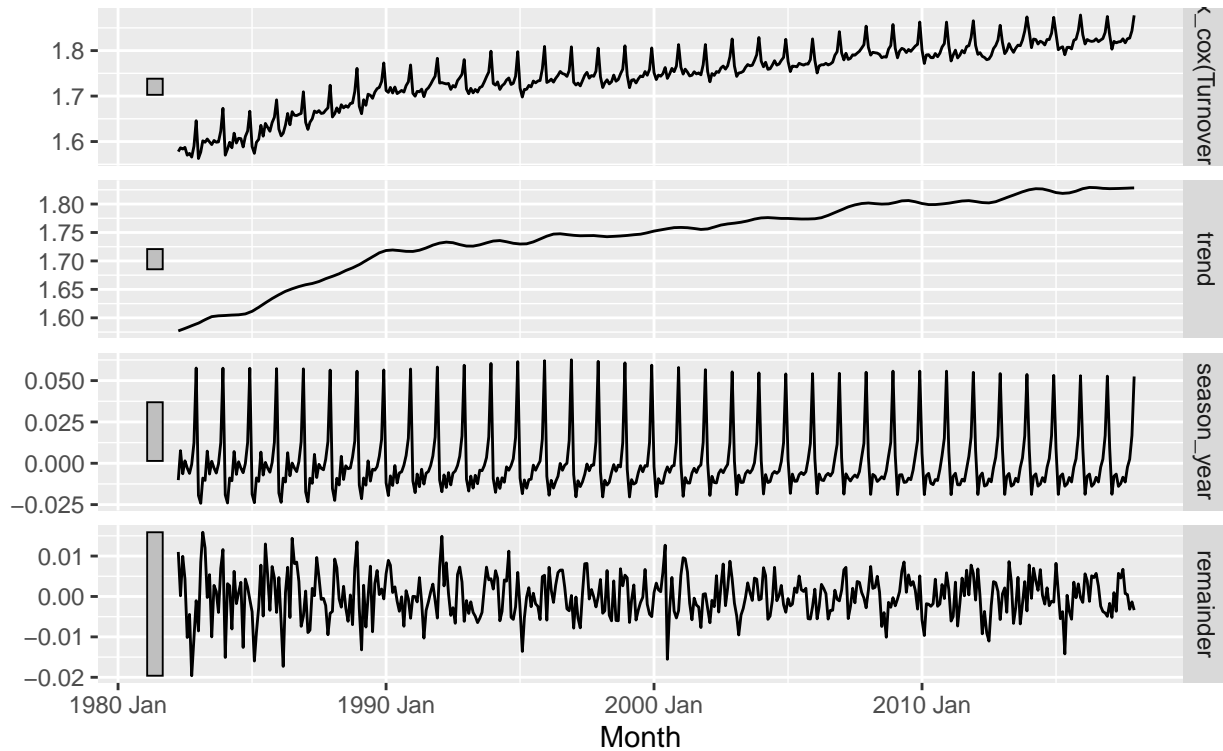
We can also assess the trend and seasonality in the data by looking at the below autocorrelation plot:

- The trended time series shows positive values that slowly decrease as the lags increase, which confirms the data has a trend.

- We can also see strong seasonality as the data follows a scalloped shape. The $r_{12}$, $r_{24}$, $r_{36}$ and $r_{48}$ autocorrelations are higher than for the other lags, which is likely explained by the Christmas and Boxing Day sales.
- There is clearly no white noise in the data. None of the autocorrelations are close to zero nor lay within the bounds of the ACF.

Other Recreational Goods in NSW
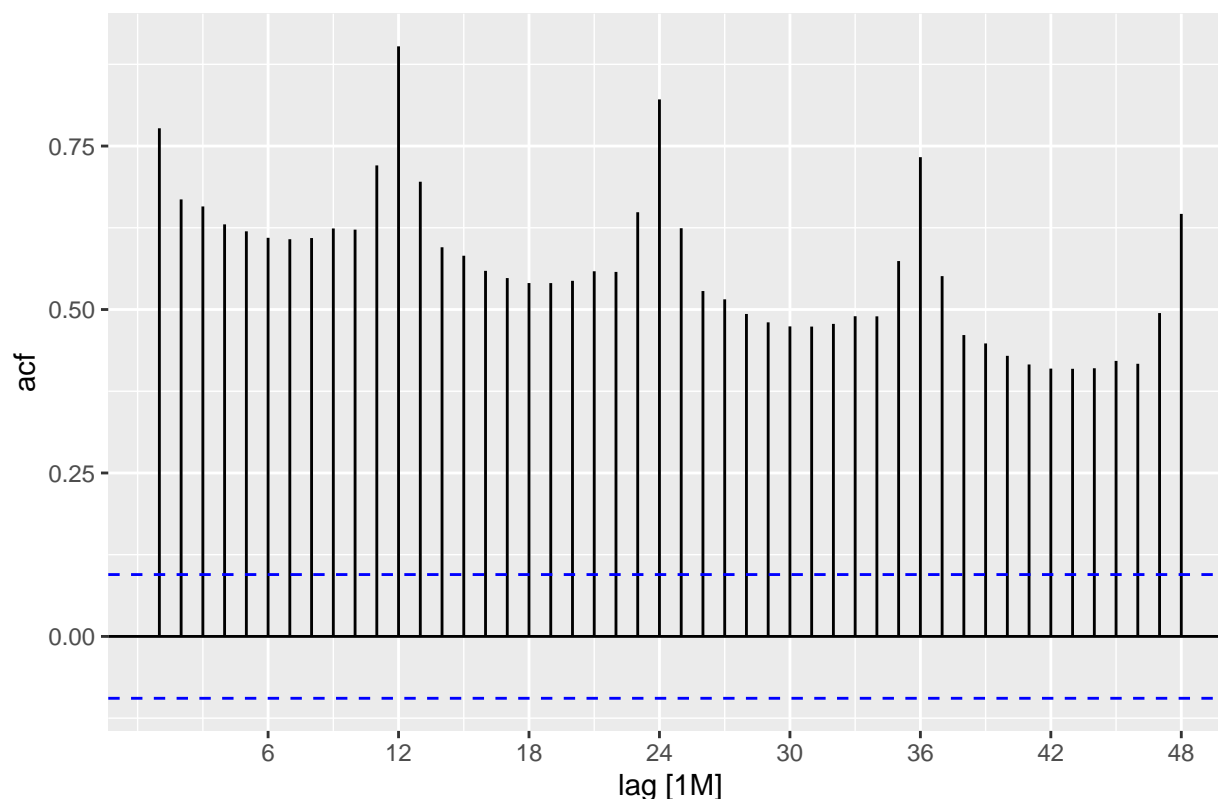
## ARIMA modelling

The above statistical properties suggest the data is non-stationary. We can make the data stationary by applying transformations and differencing the data.

As turnover is highly seasonal, it is worth exploring whether any seasonal differencing is required. The result of the unitroot_nsdiffs function confirms that we should apply one seasonal difference. Furthermore, the seasonal strength of the transformed series is 0.926, which is above the threshold value required to perform seasonal differencing.

The p-value from the unitroot_kpss test is less than 1%. Therefore, there is sufficient evidence at the 5% level to support the alternative hypothesis that the data is non-stationary. Therefore, we should difference the data.

```
## # A tibble: 1 x 3
##   nsdiffs trend_strength seasonal_strength_year
##     <int>          <dbl>                  <dbl>
## 1       1          0.993                  0.926

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1      1.09        0.01
```

From the below plot, it appears the data are completely stationary when we first difference the data. I believe there will be seldom change in the stationarity if we we apply a second difference. This is confirmed by the result of the unit_ndiffs function.

Turnover – Other Recreational Goods (NSW)
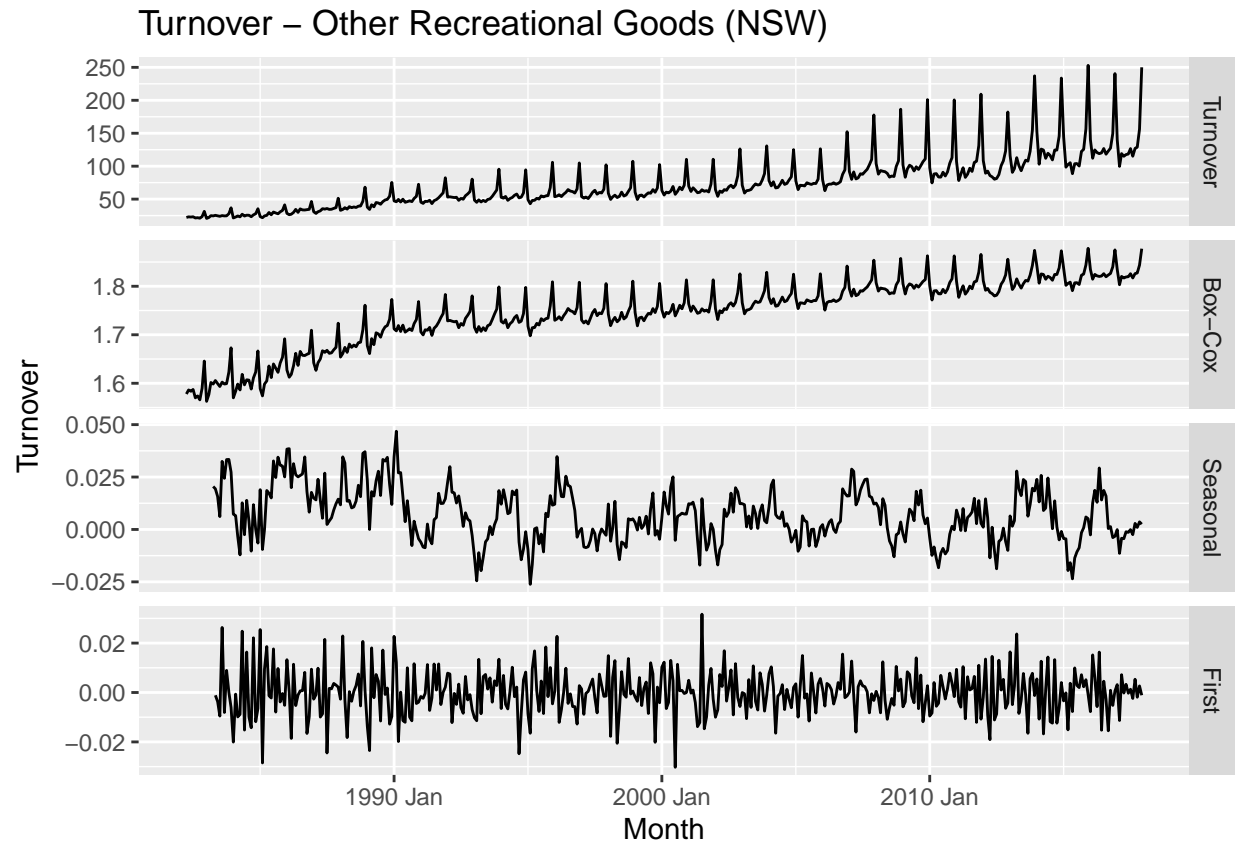
```
## # A tibble: 1 x 3
##   State           Industry                          ndiffs
##   <chr>           <chr>                              <int>
## 1 New South Wales Other recreational goods retailing     1
```

We can now use the ACF and PACF plots to help determine the values of p and q for the seasonal component and the non-seasonal component of the ARIMA model.
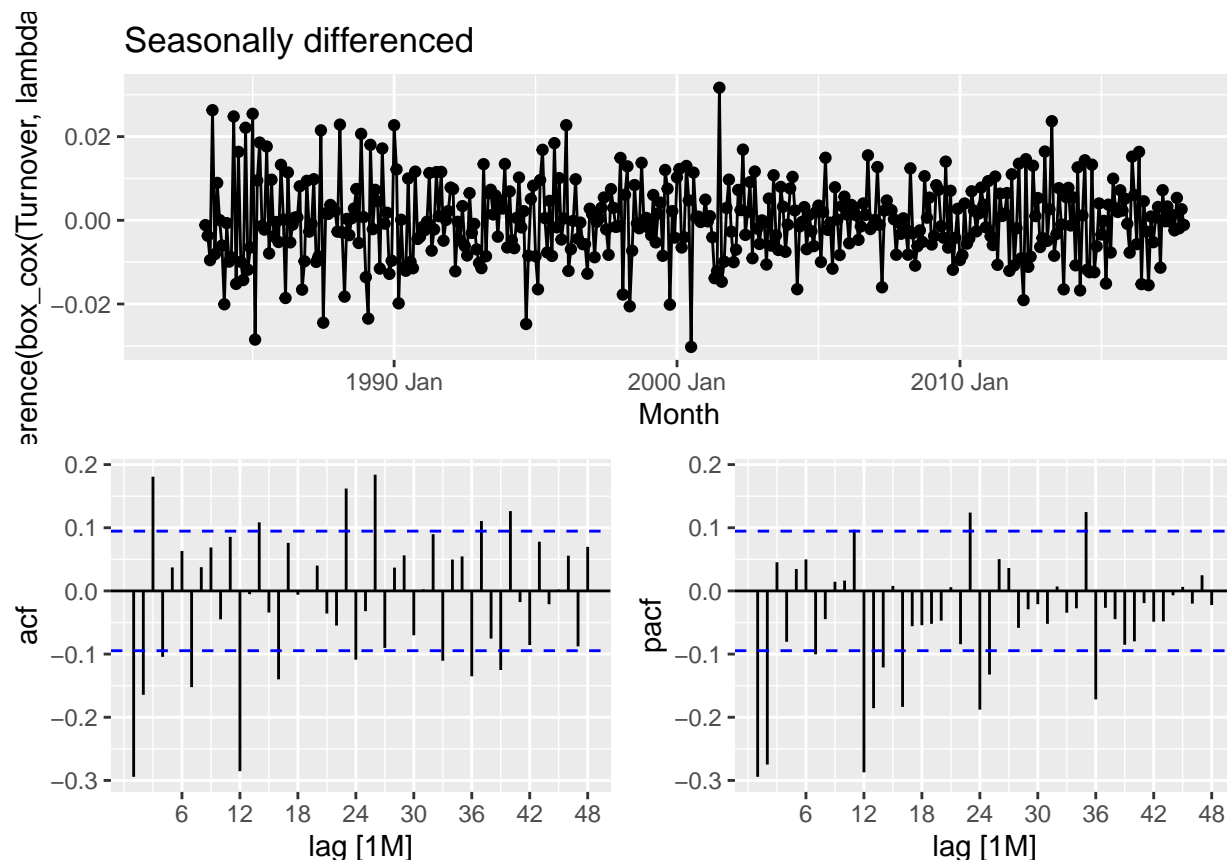
We see exponential decay in the seasonal lags of the PACF and a significant spike at lag 12 in the ACF, which suggests a seasonal MA(1) component. Likewise, the exponential decay in the seasonal lags of the ACF and the significant spike at lag 12 of the PACF suggests a seasonal AR(1) component.

If the data follow the ARIMA(0,d,q) model, we see there are generally less significant spikes beyond lag 3 of the ACF, which suggests a non-seasonal MA(3) component. We also see the ACF plot start with four significant non-seasonal lags up to lag 4. These four significant spikes might suggest a possible non-seasonal MA(4) component.

Alternatively, if the data follow the ARIMA(p,d,0) model, we see there are less significant spikes beyond lag 2 of the PACF, which suggests a non-seasonal AR(2) component.

With seasonal and first-order differencing, the above features combine to form four potential models. They are as follows:

```
## # A mable: 3 x 2
## # Key:     Model [3]
##   Model                          Orders
##   <chr>                          <model>
## 1 arima013111 <ARIMA(0,1,3)(1,1,1)[12]>
## 2 arima014111 <ARIMA(0,1,4)(1,1,1)[12]>
## 3 arima210111 <ARIMA(2,1,0)(1,1,1)[12]>
```

We will now fit the above models and compare the AICc values. The results are as follows:

```
## # A tibble: 3 x 4
##   .model        AIC    AICc    BIC
```

```
##   <chr>           <dbl>  <dbl>  <dbl>
## 1 arima014111 -2877. -2877. -2849.
## 2 arima210111 -2872. -2872. -2852.
## 3 arima013111 -2868. -2868. -2844.
```

We generally use the AICc in model selection. Based on the above results, the arima014111 is the best model with an AICc value of -2876.55.

We will also compare some of the models fitted using a test set consisting of the last 24 months of the data provided. Regardless of which measure is used, we can see the arima014111 outperforms the other models. Therefore, the arima014111 is my chosen model.

```
## # A tibble: 3 x 5
##   .model        RMSE   MAE  MAPE  MASE
##   <chr>        <dbl> <dbl> <dbl> <dbl>
## 1 arima014111   15.1  10.7  7.20  1.78
## 2 arima013111   24.2  15.2  9.74  2.53
## 3 arima210111   24.5  15.4  9.84  2.56
```

We can run a few diagnostic tests to confirm how well the arima014111 model fits the data. We can achieve this by checking the innovation residuals.
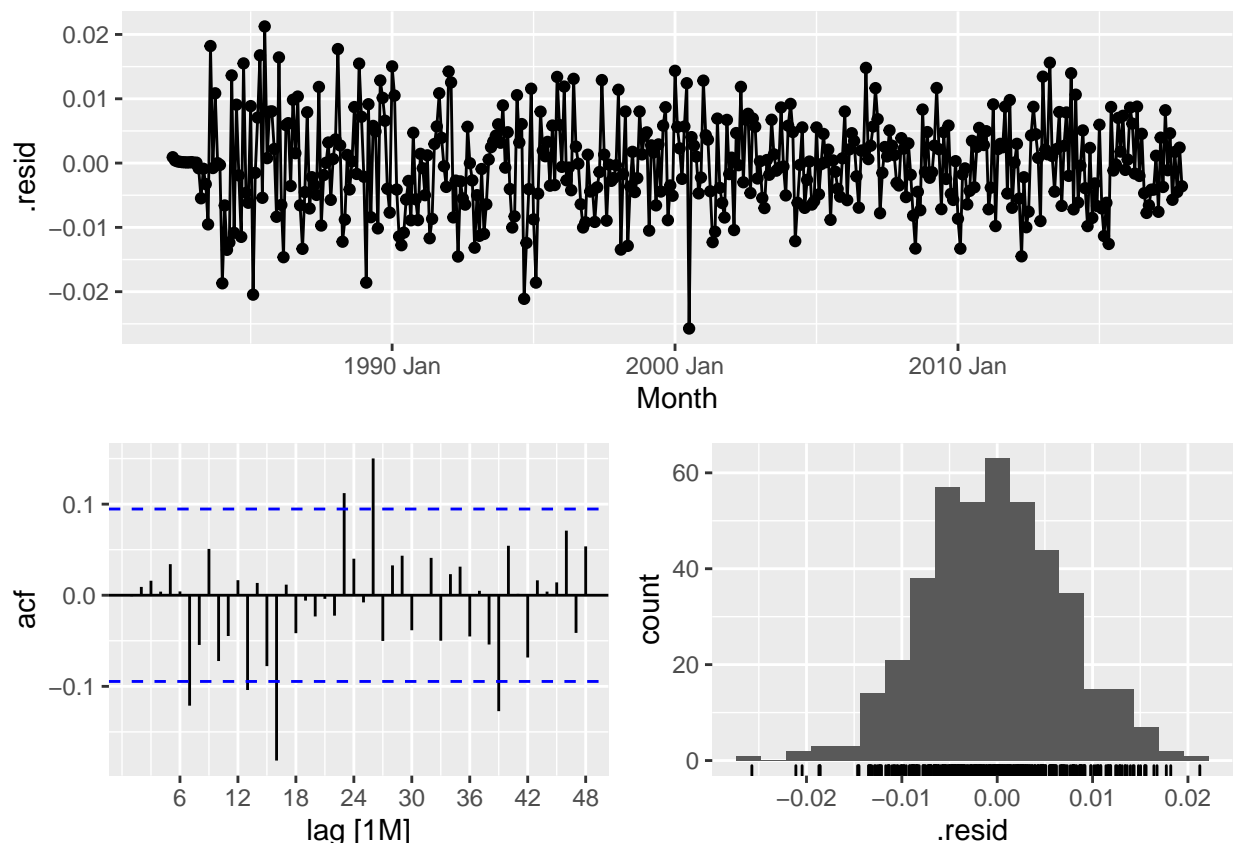


The residuals in the arima014111 model follow a normal distribution. However, there are a few significant spikes in the ACF that are a cause for concern, so we should run the Ljung-Box test to confirm the accuracy of the arima014111 model.

```
## # A tibble: 3 x 3
```

```
##    .model      lb_stat  lb_pvalue
##    <chr>         <dbl>      <dbl>
## 1 arima013111    75.3 0.00000915
## 2 arima014111    60.2 0.000875
## 3 arima210111    64.6 0.000245
```

The arima014111 model fails the Ljung-Box test. The p-value for the arima014111 model is less than 1%. Therefore, there is sufficient evidence at the 5% level to support the alternative hypothesis that the residuals are distinguishable from a white noise series. This means the arima014111 model can be used for forecasting, but the prediction intervals may not be accurate due to the correlated residuals.

The 24 month forecast from the arima014111 model is shown below:

```
## # A tsibble: 24 x 5 [1M]
## # Key:       .model [1]
##    .model         Month .mean                 '80%'                 '95%'
##    <chr>          <mth> <dbl>                 <hilo>                <hilo>
##  1 arima014111 2016 Jan  140. [124.82991, 156.9261]80 [117.98011, 167.4995]95
##  2 arima014111 2016 Feb  103. [ 91.40610, 114.8774]80 [ 86.39570, 122.6075]95
##  3 arima014111 2016 Mar  111. [ 97.36296, 125.3242]80 [ 91.52596, 134.7484]95
##  4 arima014111 2016 Apr  114. [ 98.42755, 130.5963]80 [ 91.89962, 141.7623]95
##  5 arima014111 2016 May  102. [ 88.03304, 116.3556]80 [ 82.26493, 126.1499]95
##  6 arima014111 2016 Jun  109. [ 93.48949, 126.2722]80 [ 86.94646, 137.8499]95
##  7 arima014111 2016 Jul  112. [ 94.93072, 130.1201]80 [ 88.00297, 142.7273]95
##  8 arima014111 2016 Aug  110. [ 92.83999, 128.2714]80 [ 85.91611, 141.0641]95
##  9 arima014111 2016 Sep  125. [104.15629, 148.7393]80 [ 95.70768, 165.3644]95
## 10 arima014111 2016 Oct  132. [108.85598, 158.7384]80 [ 99.58519, 177.7252]95
## # ... with 14 more rows
```

## ARIMA(0,1,4)(1,1,1) Forecast – Other Recreational Goods (NSW)
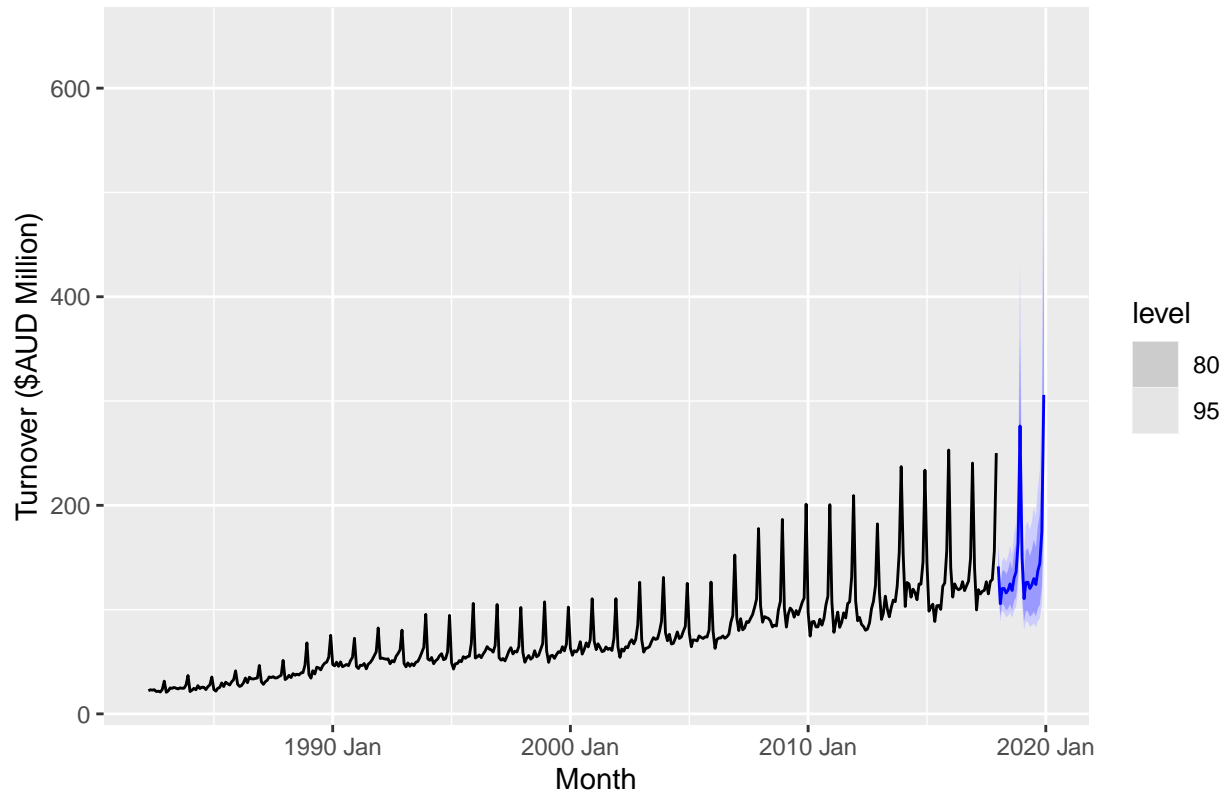


## ETS modelling

I believe over-forecasting will not be an issue as we are forecasting over a relatively short time horizon of 2 years. Using an additive trend in the ETS model should therefore be sufficient, though we will also consider the damped additive trend in our analysis.

We know the variation in the seasonal pattern is proportional to the trend of the time series. This suggests the season component should be multiplicative, but by applying the box-cox transformation, we can stabalise the seasonal variation in the data and employ an additive season component in the ETS model.

Based on the above statistical features, I have created a short-list of ETS models that I believe will be most appropriate in forecasting turnover. These cover two sets of groups for the multiplicative error and additive error. They are as follows:

```
## # A mable: 4 x 2
## # Key:    Model [4]
##   Model      Orders
##   <chr>      <model>
## 1 MAA    <ETS(M,A,A)>
## 2 MAdA   <ETS(M,Ad,A)>
## 3 AAdA   <ETS(A,Ad,A)>
## 4 AAA    <ETS(A,A,A)>
```

We will now fit the above models and compare the AICc values.

```
## # A tibble: 4 x 4
##   .model    AIC   AICc    BIC
##   <chr>   <dbl>  <dbl>  <dbl>
## 1 AAA    -1530. -1528. -1461.
## 2 AAdA   -1516. -1515. -1443.
## 3 MAA    -1512. -1510. -1443.
## 4 MAdA   -1487. -1486. -1414.
```

We generally use the AICc in model selection. Based on the above results, the AAA is the best model with an AICc value of -1528.42.

We will also compare some of the models fitted using a test set consisting of the last 24 months of the data provided. Regardless of which measure is used, we see the AAdA model outperforms the other models.

```
## # A tibble: 4 x 5
##   .model  RMSE   MAE  MAPE  MASE
##   <chr>  <dbl> <dbl> <dbl> <dbl>
## 1 AAdA    11.4  9.21  6.47  1.54
## 2 MAdA    20.0 13.0   8.39  2.16
## 3 MAA     21.0 13.9   9.02  2.31
## 4 AAA     24.6 15.3   9.75  2.54
```

With these conflicting results, we can run a few diagnostic tests to confirm how well the AAA model and AadA model fit the data. We can achieve this by checking the innovation residuals.
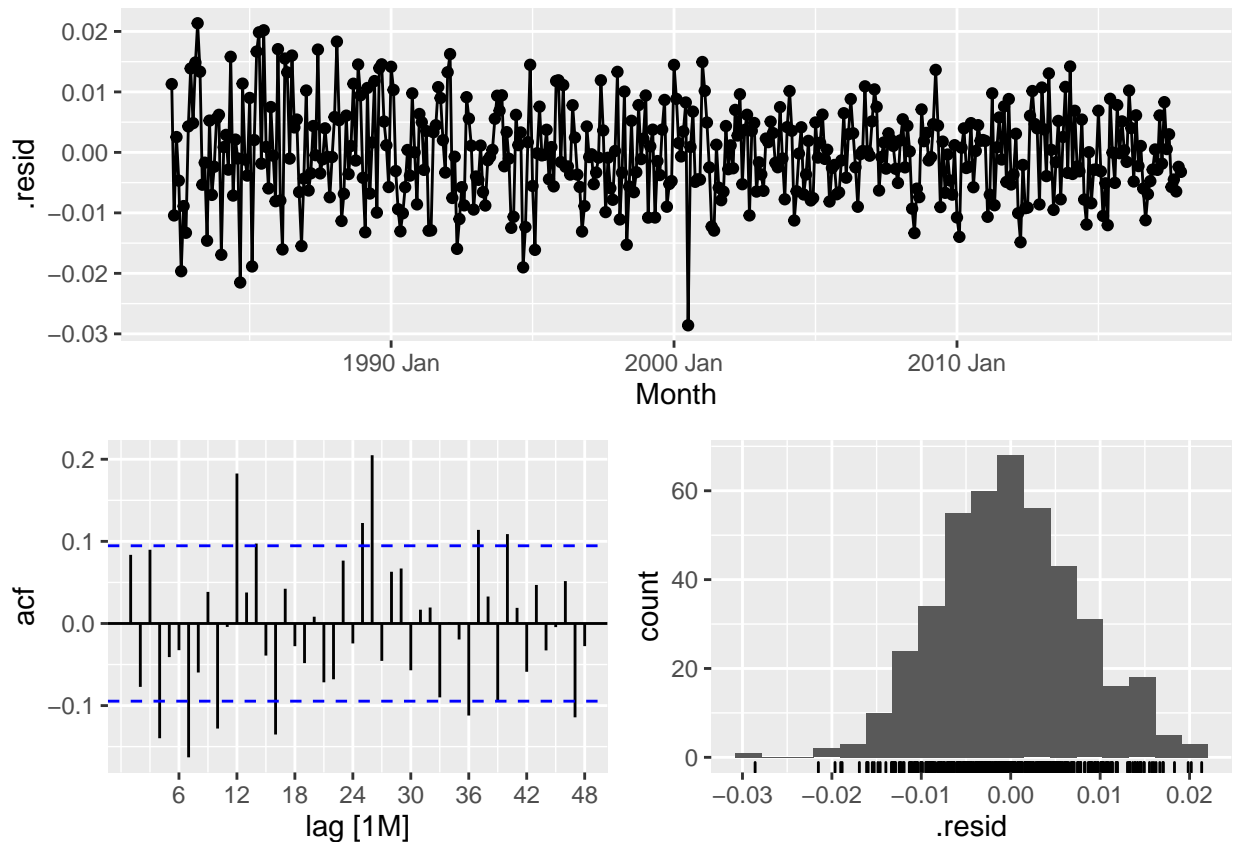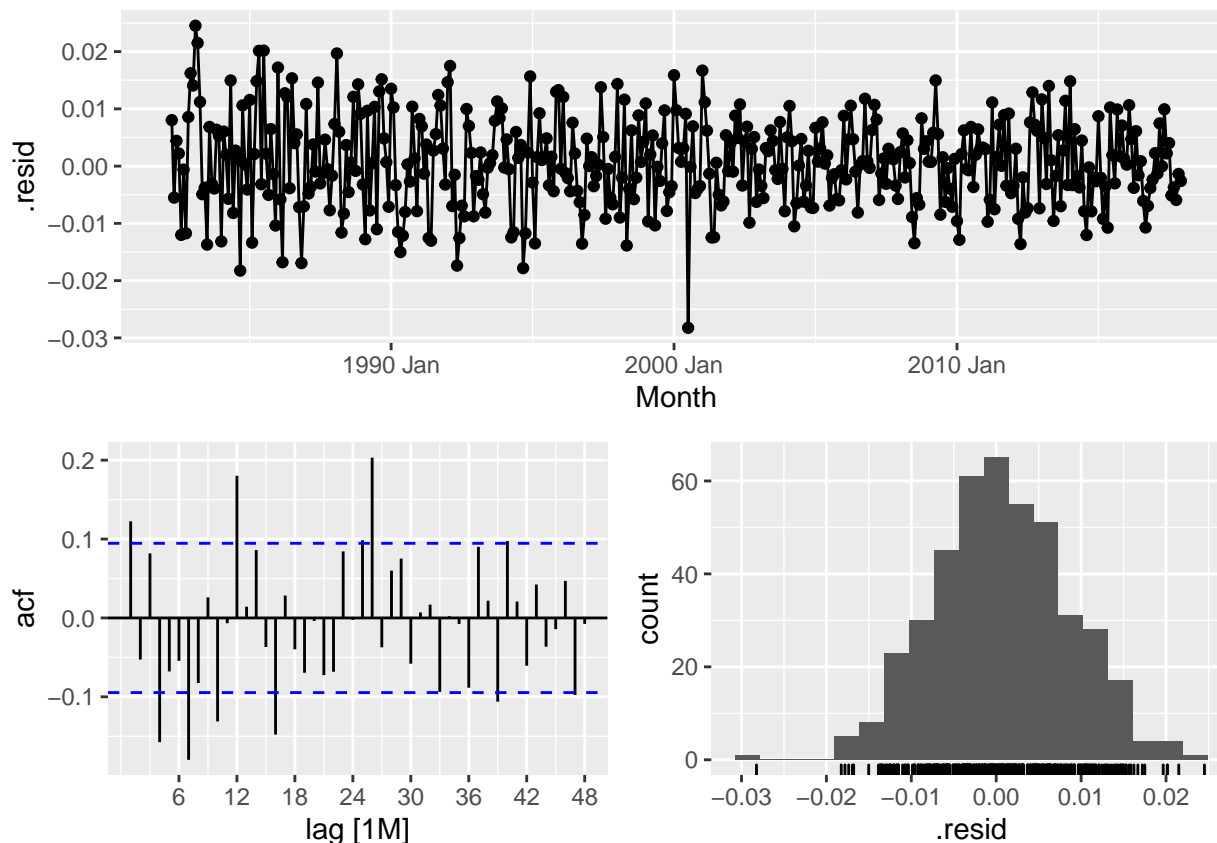
Both models follow somewhat of a normal distribution and have residuals that follow a constant variance. There are a few significant spikes in the ACF that are also a cause for concern, so we should run the Ljung-Box test to confirm the accuracy of the models.

```
## # A tibble: 4 x 3
##    .model lb_stat lb_pvalue
##    <chr>    <dbl>     <dbl>
## 1 AAA       121.  8.43e-14
## 2 AAdA      127.  7.33e-15
## 3 MAA       138.  1.11e-16
## 4 MAdA      153.  0.
```

The AAA model and AAdA both fail the Ljung-Box test. The p-values for the models are less than 1%. Therefore, there is sufficient evidence at the 5% level to support the alternative hypothesis that the models are distinguishable from a white noise series. This means the models can be used for forecasting, but the prediction intervals may not be accurate due to the correlated residuals.

Considering all of the above, I would prefer the AAdA model. The reason I prefer the AAdA model is because it yields much lower values in all the accuracy measures compared to the AAA model. Though the AAA model yields a lower AICc score, the conflicting results suggested by the accuracy measures confirms my belief that the AAdA model is the better model.

The 24 month forecast from the AAdA model is shown below:

```
## # A tsibble: 24 x 5 [1M]
## # Key:       .model [1]
##    .model    Month .mean                '80%'                '95%'
```

11

```
##    <chr>     <mth> <dbl>                 <hilo>                <hilo>
##  1 AAdA    2016 Jan  142. [125.02408, 159.6842]80 [117.73598, 171.2788]95
##  2 AAdA    2016 Feb  100. [ 89.35418, 112.0608]80 [ 84.49741, 119.5238]95
##  3 AAdA    2016 Mar  112. [ 97.73069, 126.6468]80 [ 91.73216, 136.4566]95
##  4 AAdA    2016 Apr  116. [100.06584, 132.9155]80 [ 93.40654, 144.3300]95
##  5 AAdA    2016 May  104. [ 89.76252, 119.6618]80 [ 83.72173, 130.0876]95
##  6 AAdA    2016 Jun  111. [ 94.17073, 129.0081]80 [ 87.30880, 141.4826]95
##  7 AAdA    2016 Jul  113. [ 95.29917, 133.4031]80 [ 87.94315, 147.3376]95
##  8 AAdA    2016 Aug  113. [ 93.99692, 133.7822]80 [ 86.43396, 148.5696]95
##  9 AAdA    2016 Sep  128. [104.43015, 154.8482]80 [ 95.20398, 174.3562]95
## 10 AAdA    2016 Oct  134. [107.78204, 164.2160]80 [ 97.71185, 186.6435]95
## # ... with 14 more rows
```



ETS(A,Ad,A) Forecast – Other Recreational Goods (NSW)
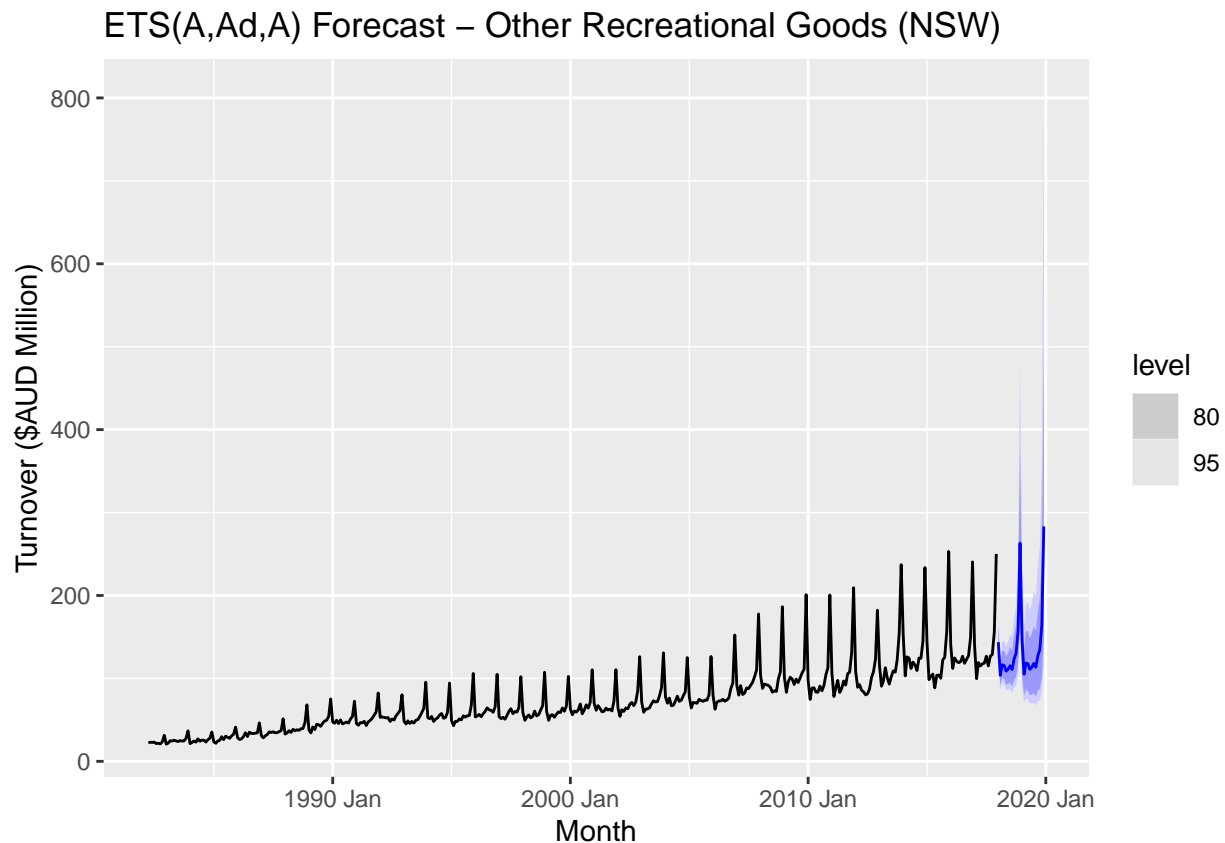
## ARIMA vs ETS

Based on the diagnostic tests, I would say the ETS model gives the better forecast. First, we can see in that
the ETS model outperforms the ARIMA model across all accuracy measures. Second, though both models
fail the Ljung-Box test, the innovation residuals of the ETS model follow a more symmetric distribution,
which makes calculating prediction intervals easier.

```
## # A tibble: 2 x 5
##   .model       RMSE   MAE  MAPE  MASE
##   <chr>       <dbl> <dbl> <dbl> <dbl>
## 1 AAdA         11.4  9.21  6.47  1.54
## 2 arima014111  15.1 10.7   7.20  1.78
```

## ETS(A,Ad,A) – Histogram of Innovation Residuals
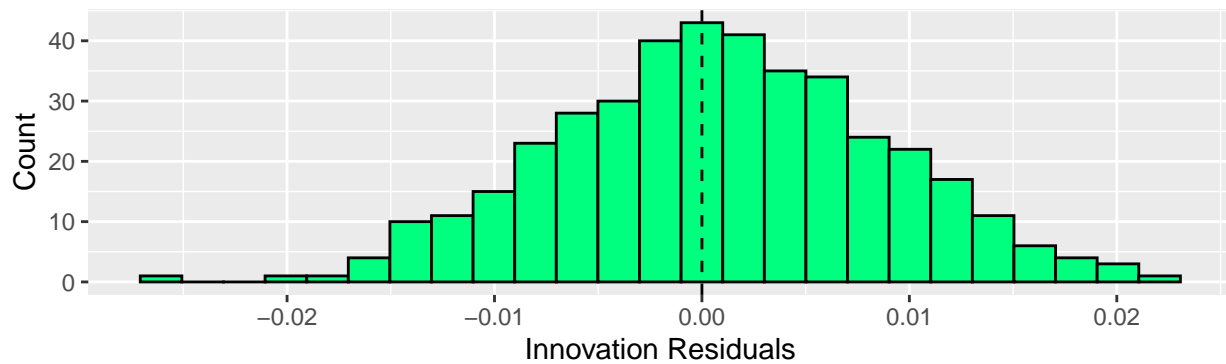
## ARIMA(0,1,4)(1,1,1) – Histogram of Innovation Residuals
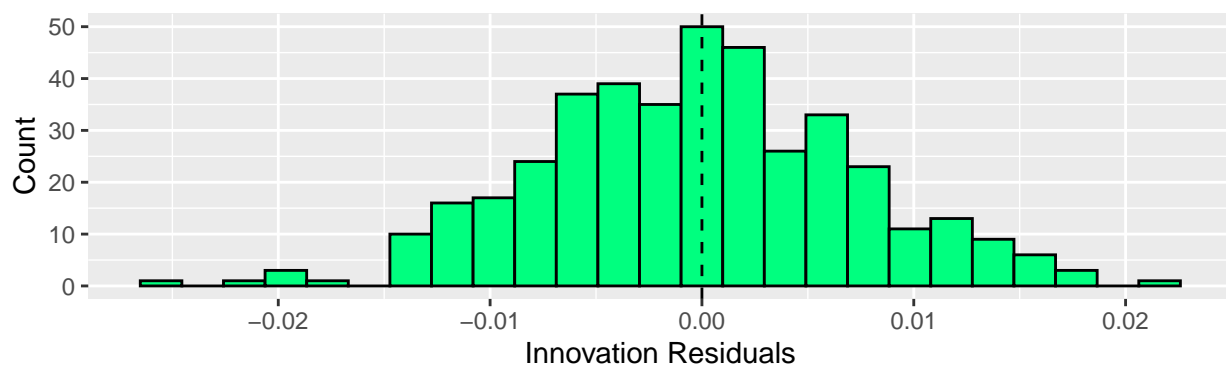
## Out-of-sample point forecasts and 80% prediction intervals

We can see there is a slight stabilization in the trend of the series when we forecast the data over the 24 month period. This could be explained by the general decrease in turnover in 2018.

```
## # A tsibble: 48 x 4 [1M]
## # Key:       .model [2]
##    .model       Month .mean                 '80%'
##    <chr>        <mth> <dbl>                 <hilo>
##  1 arima014111 2019 Jan 133.  [118.81251, 147.9039]80
##  2 arima014111 2019 Feb  99.1 [ 88.61063, 110.3942]80
##  3 arima014111 2019 Mar 112.  [ 98.47430, 126.3634]80
##  4 arima014111 2019 Apr 107.  [ 93.13209, 121.9858]80
##  5 arima014111 2019 May 103.  [ 88.94637, 117.2794]80
##  6 arima014111 2019 Jun 102.  [ 88.17676, 117.5920]80
##  7 arima014111 2019 Jul 113.  [ 96.16696, 131.7015]80
##  8 arima014111 2019 Aug 102.  [ 86.72017, 118.1345]80
##  9 arima014111 2019 Sep 110.  [ 92.78535, 129.3820]80
## 10 arima014111 2019 Oct 122.  [101.13871, 145.0485]80
## # ... with 38 more rows
```

# ARIMA(0,1,4)(1,1,1) Forecast – Other Recreational Goods (NSW)
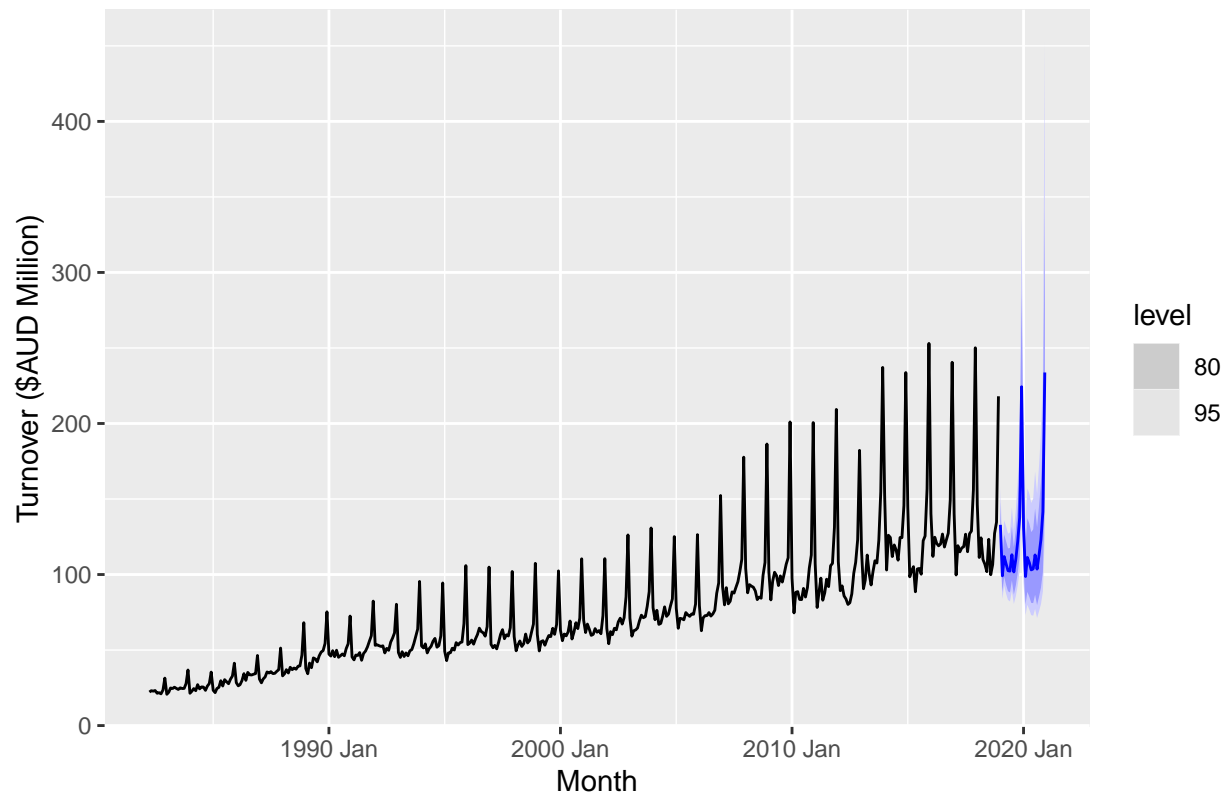
## ETS(A,Ad,A) Forecast – Other Recreational Goods (NSW)
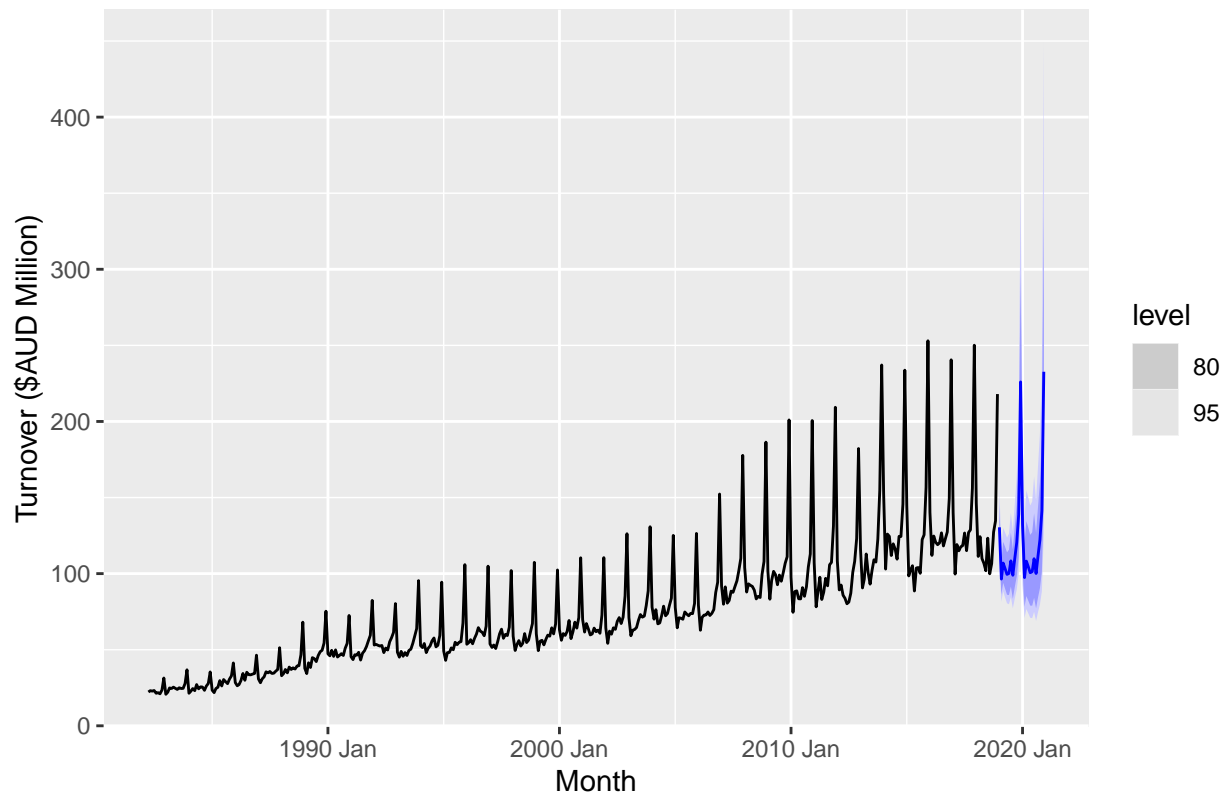


## Obtain up-to-date data from the ABS website and compare forecasts

The accuracy measures suggest the forecast on the up-to-date ABS data did not perform as well as the forecast on the test-set. I speculate this is the case because of the unprecedented impacts of the COVID-19 pandemic. Though many industries in the retail sector suffered, the other recreational goods industry generally performed better in 2020 compared to prior years. Various lock-downs were imposed in NSW, which may have inspired a greater need to buy recreational goods as people needed to find ways to stay mentally and physically fit while at home.

The ARIMA model yields a lower RMSE value compared to the ETS model. Though the ETS model outperforms the ARIMA model in terms of the other accuracy measures, the large forecast errors on the 2020 data are undesirable, which is why I prefer RMSE. Therefore, my belief is that the ARIMA model does a better job at forecasting the most up-to-date ABS data.

```
## # A tibble: 2 x 5
##   .model       RMSE   MAE  MAPE  MASE
##   <chr>       <dbl> <dbl> <dbl> <dbl>
## 1 arima014111  38.4  25.5  15.2  4.01
## 2 AAdA         39.1  25.3  14.7  3.98
```

## Benefits and limitations of models

Overall, I think my ARIMA model did a decent job at forecasting turnover. I think the ARIMA model does a better job at capturing the increase in the trend of the data in the long-term, which is why the model was

better at forecasting the up-to-date ABS data. Though, it was easy to determine whether any differencing or seasonal differencing was required with the unit root tests, I did find it challenging to choose appropriate values for p,q,P and Q as there were many significant autocorrelations and partial autocorrelations that laid outside the bounds of the ACF and PACF.

Overall, I was also pleased with the performance of the ETS model. I think the ETS model does a better job at forecasting turnover in the short-term as the damped additive component helps to not over-forecast turnover. Though the ETS model is relatively easy to implement, I felt confused at times as I wasn't sure how the components worked together to forecast the data. In that sense, I felt I had gained a better of the data when I forecasted turnover with the ARIMA model.

## References

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3