Drinks ARIMA

Jackson Cates

10/21/2020

Libraries

```
library(readr)
library(tsibble)
library(dplyr)
library(tidyr)
library(ggplot2)
library(fable)
library(feasts)
library(urca)
library(gridExtra)
library(lubridate)
```

Reading in data

```
drinks = read_csv('../../Data/Drinks.csv')
```

```
## Parsed with column specification:
## cols(
##   Date = col_date(format = ""),
##   PEP = col_double(),
##   COKE = col_double(),
##   SBUX = col_double(),
##   MNST = col_double()
## )
```

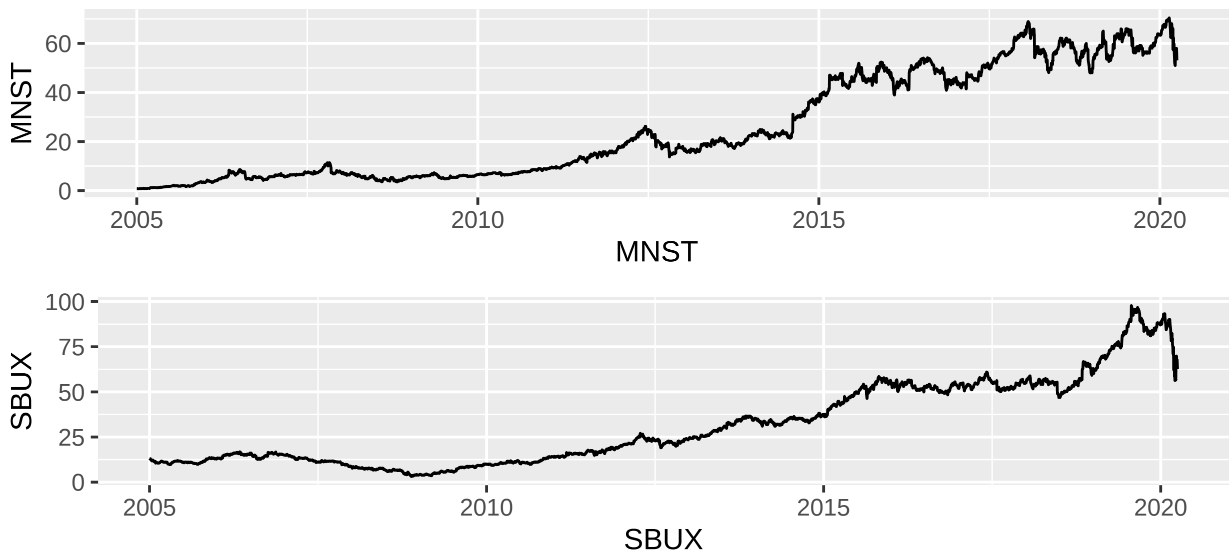
```
drinks = drinks[,-3]
drinks = drinks[,-2]

drinks = drinks %>% filter(Date > "2004-12-31")

# Turns all of it into a tsibble
ts = drinks %>%
  mutate(index = as_date(Date)) %>%
  select(-Date) %>%
  as_tsibble(index = index)
ts
```

```
## # A tsibble: 5,570 x 3 [1D]
##     SBUX  MNST index
##    <dbl> <dbl> <date>
##  1  13.2 0.761 2005-01-01
##  2  13.1 0.763 2005-01-02
```

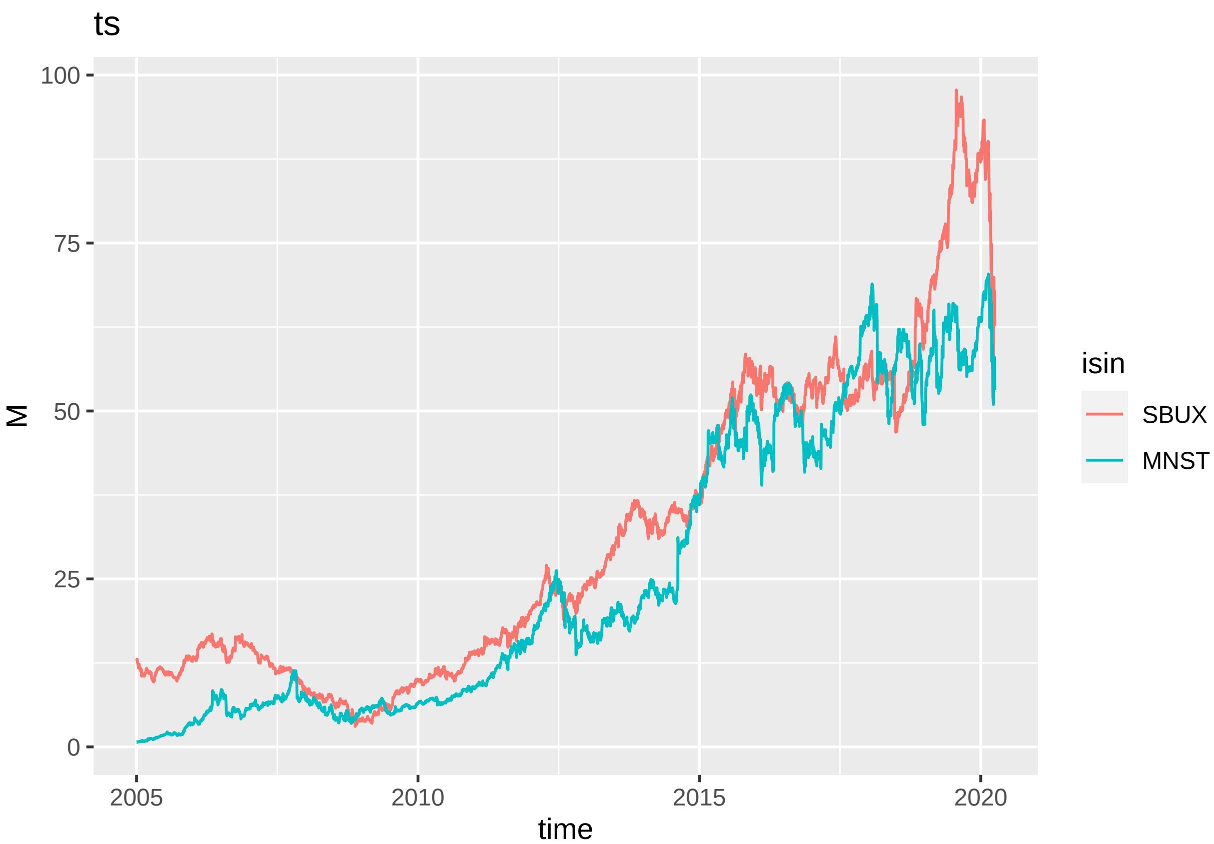
```
## 3   13.0 0.765 2005-01-03
## 4   13.0 0.734 2005-01-04
## 5   13.1 0.698 2005-01-05
## 6   12.7 0.721 2005-01-06
## 7   12.7 0.705 2005-01-07
## 8   12.5 0.709 2005-01-08
## 9   12.4 0.712 2005-01-09
## 10  12.3 0.716 2005-01-10
## # ... with 5,560 more rows
```

Time-plot

```
plot1 = ts %>% autoplot(MNST) + xlab("MNST")
plot2 = ts %>% autoplot(SBUX) + xlab("SBUX")
grid.arrange(plot1, plot2, nrow=3)
```




```
df1 = data.frame(time = ts$index, M = ts$SBUX, isin = "SBUX")
df2 = data.frame(time = ts$index, M = ts$MNST, isin = "MNST")
df = rbind(df1, df2)
ggplot(df, aes(x = time, y = M, color = isin)) + geom_line() + ggtitle("ts")
```

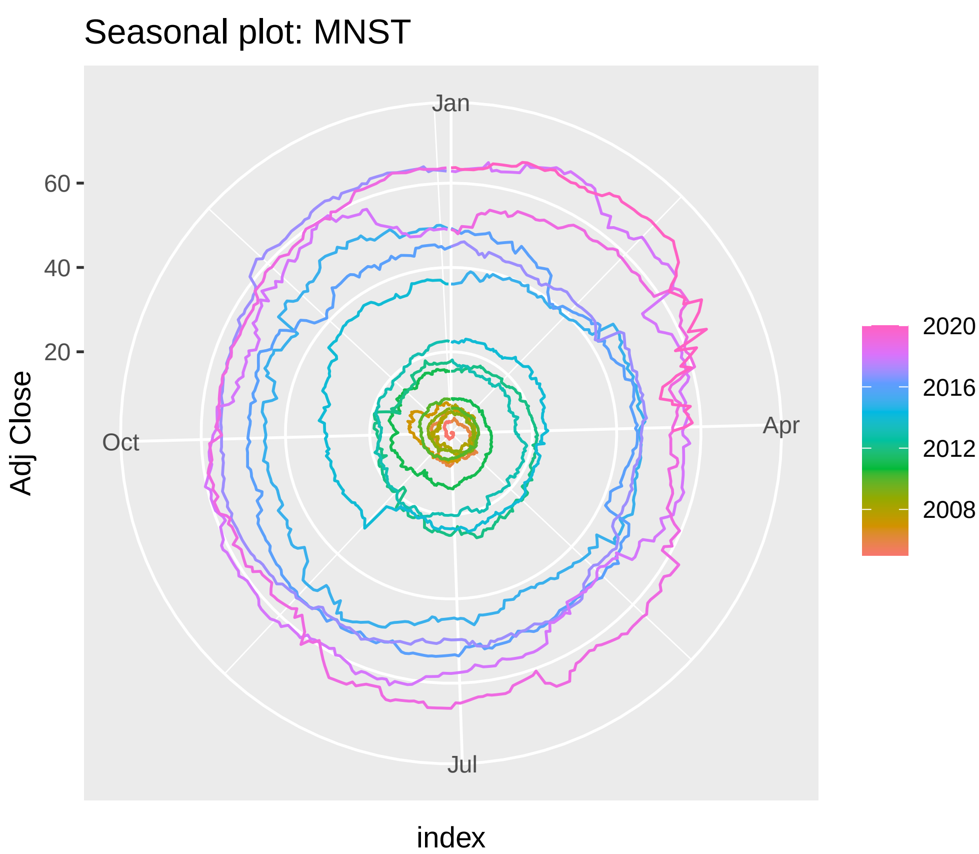

ts
100
75
50
25
0
M
2005
2010
2015
2020
time
isin
SBUX
MNST

Univariate of MNST

```
mnst = ts[,-1]
```

Seasonal Plot

```
mnst %>% gg_season(MNST, polar = T) +
  ggtitle("Seasonal plot: MNST") + ylab("Adj Close")
```

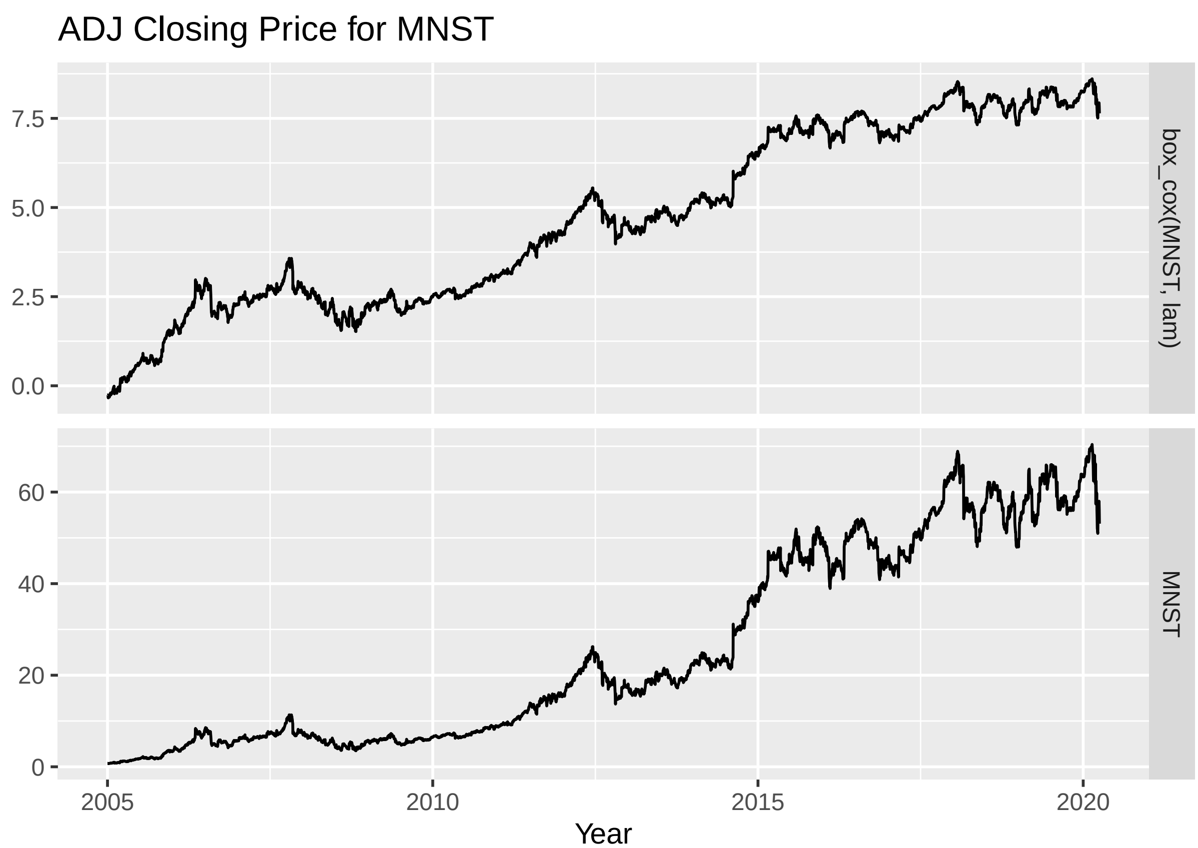


Differencing

$$w_t = \begin{cases} log(y_t) & \lambda = 0 \\ \frac{y_t^\lambda - 1}{\lambda} & \lambda \neq 0 \end{cases}$$

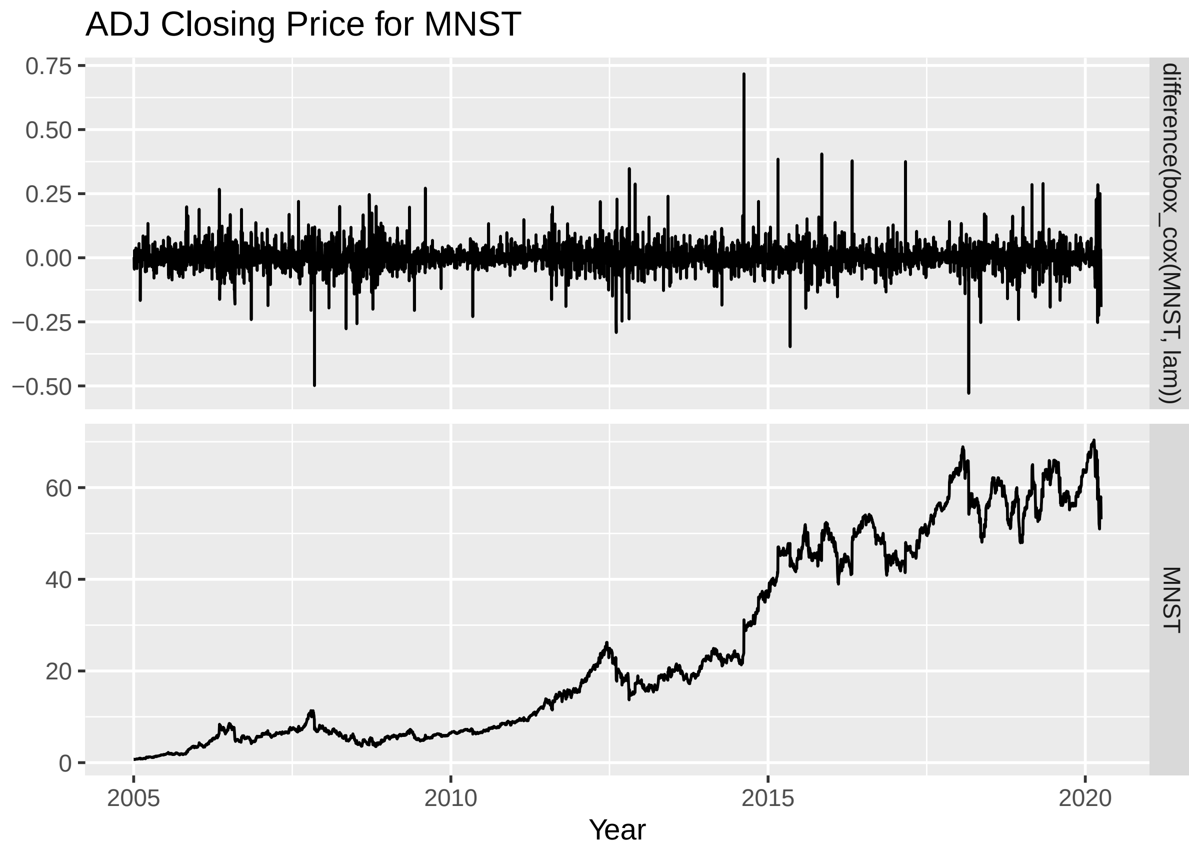
Taking a BoxCox Transformation

```
lam = 0.3

mnst %>%
  mutate(box_cox(MNST, lam)) %>%
  gather() %>%
  ggplot(aes(x = index, y = value)) +
  geom_line() +
  facet_grid(key ~ ., scales = "free_y") +
  xlab("Year") + ylab("") +
  ggtitle("ADJ Closing Price for MNST")
```



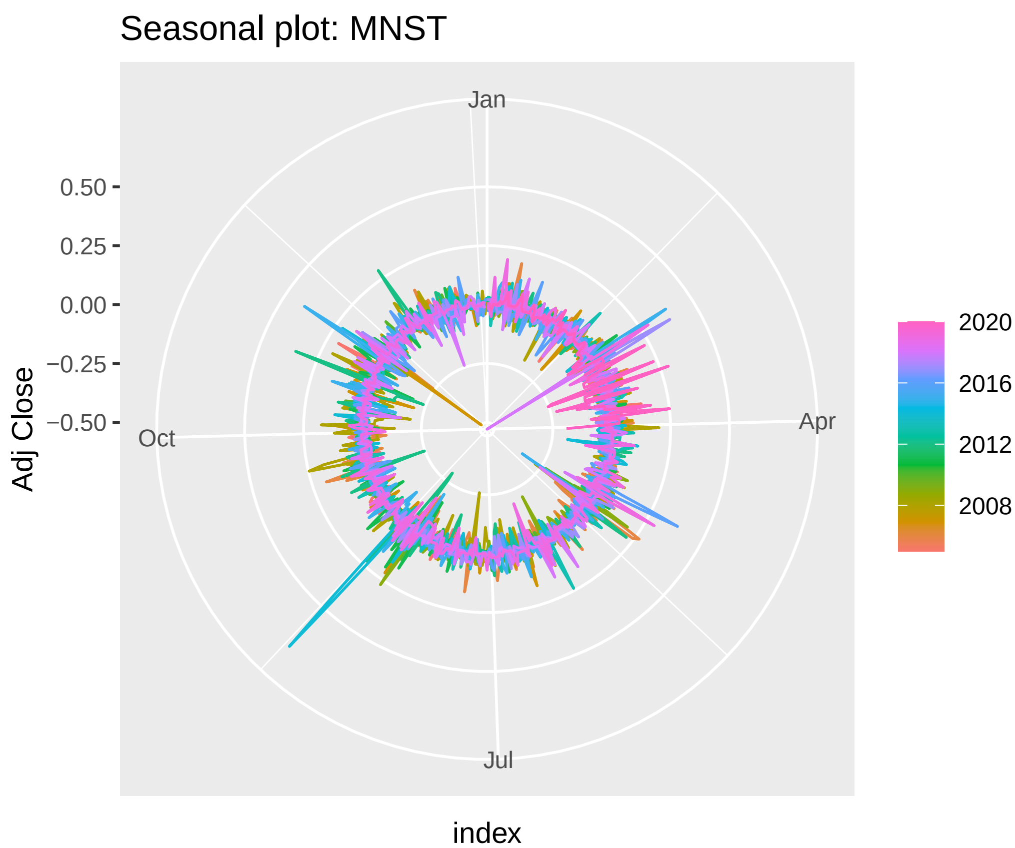

```
mnst %>%
  mutate(difference(box_cox(MNST, lam))) %>%
  gather() %>%
  ggplot(aes(x = index, y = value)) +
  geom_line() +
  facet_grid(key ~ ., scales = "free_y") +
  xlab("Year") + ylab("") +
  ggtitle("ADJ Closing Price for MNST")
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



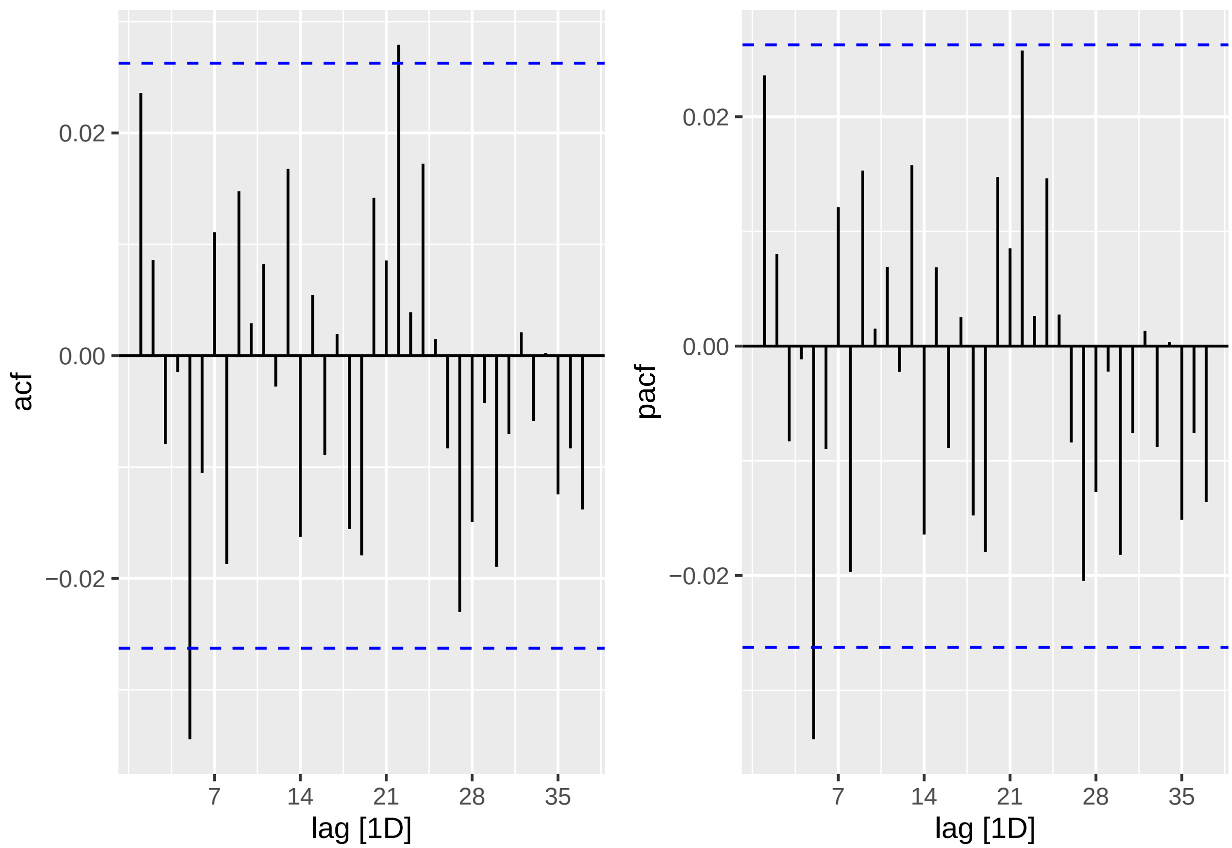

```r
mnst_trans = mnst
mnst_trans$MNST = difference(box_cox(mnst_trans$MNST, lam))
mnst_trans = drop_na(mnst_trans)
```

```r
mnst_trans %>% gg_season(MNST, polar = T) +
  ggtitle("Seasonal plot: MNST") + ylab("Adj Close")
```



Choosing a model

```
plot1 = mnst_trans %>% ACF(MNST) %>% autoplot()
plot2 = mnst_trans %>% PACF(MNST) %>% autoplot()
grid.arrange(plot1, plot2, ncol=2)
```


0.02
0.00
−0.02
acf
7
14
21
28
35
lag [1D]
0.02
0.00
−0.02
pacf
7
14
21
28
35
lag [1D]

Fitting the model

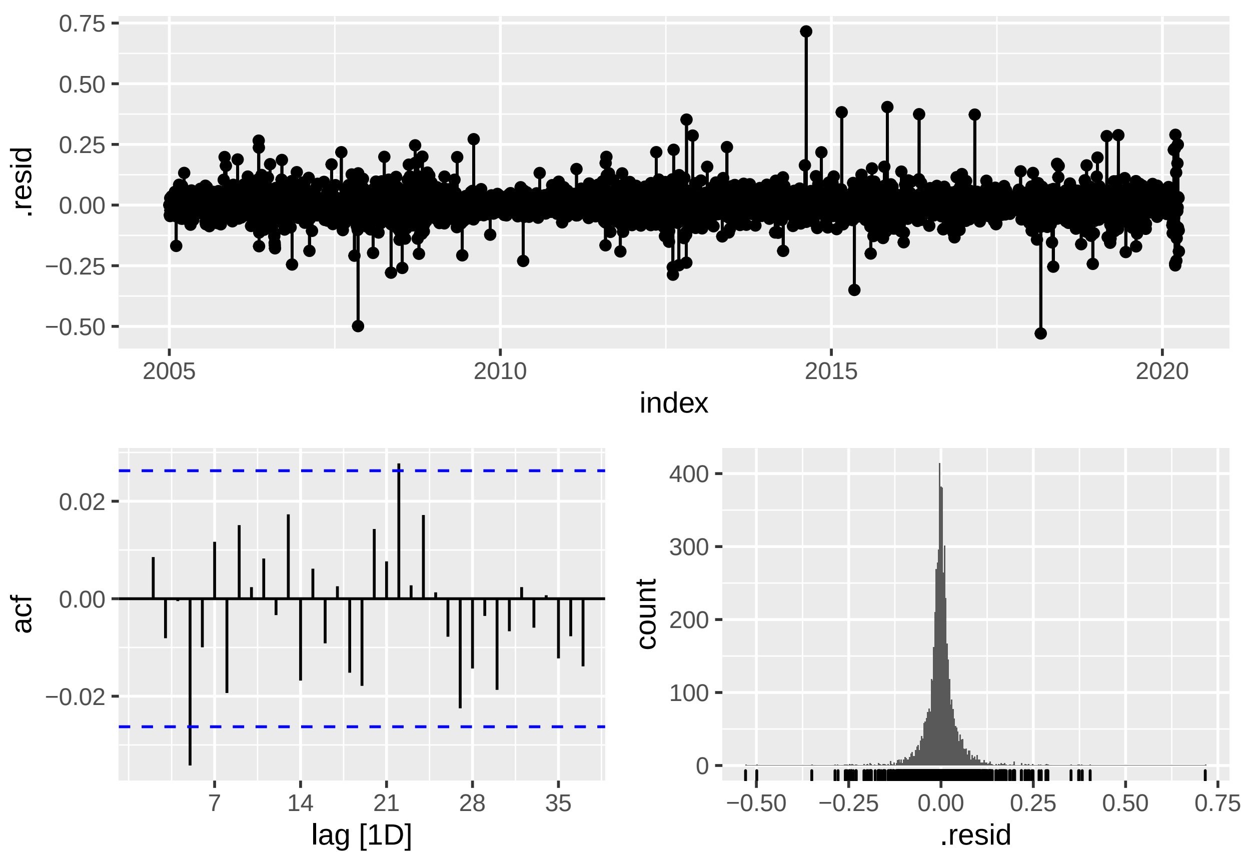
When fitting the original model, $ARIMA(1,1,1)$.

```
fit = mnst %>% model(ARIMA(box_cox(MNST, lam) ~ 1 + pdq(1, 1, 1) + PDQ(0, 0, 0)))
report(fit)
```

```
## Series: MNST
## Model: ARIMA(1,1,1) w/ drift
## Transformation: box_cox(.x, lam)
##
## Coefficients:
##          ar1     ma1  constant
##       0.0121  0.0113    0.0014
## s.e.  0.4600  0.4560    0.0006
##
## sigma^2 estimated as 0.002051:  log likelihood=9334.56
## AIC=-18661.12   AICc=-18661.11   BIC=-18634.62
```

Checking residuals

```
fit %>% gg_tsresiduals()
```



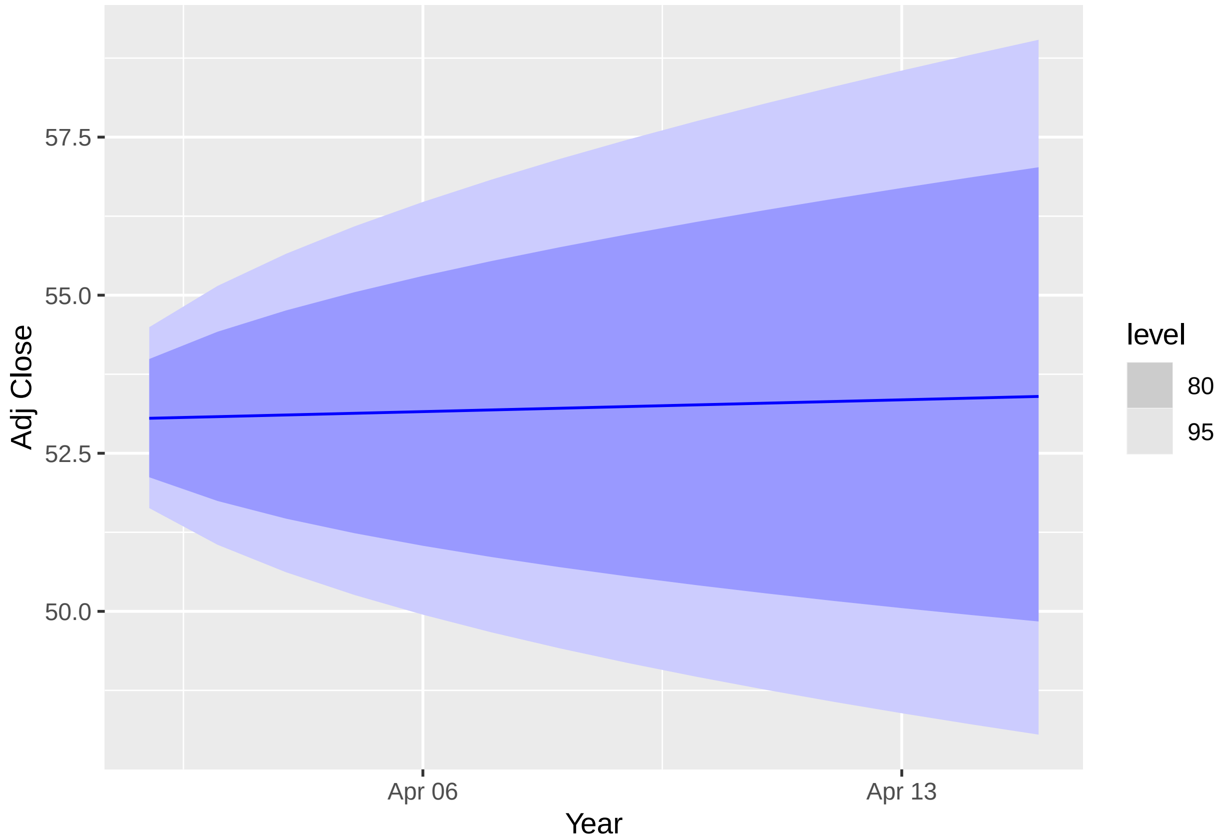

```
augment(fit) %>% features(.resid, ljung_box, lag = 12, dof = 4)
```

```
## # A tibble: 1 x 3
##   .model                                                      lb_stat lb_pvalue
##   <chr>                                                         <dbl>     <dbl>
```

```
## 1 ARIMA(box_cox(MNST, lam) ~ 1 + pdq(1, 1, 1) + PDQ(0, 0, 0))    60.8  3.31e-10
```

Forecasting

```r
fc = fit %>% forecast()

fc %>%
autoplot() +
  ylab("Adj Close") + xlab("Year")
```




```r
accuracy(fit)
```

```
## # A tibble: 1 x 9
##   .model                  .type         ME  RMSE   MAE     MPE  MAPE  MASE    ACF1
##   <chr>                   <chr>      <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl>   <dbl>
## 1 ARIMA(box_cox(MNST, la~ Trai~ -0.00318 0.501 0.235 -0.0119  1.20 0.308 -0.0151
```

The constant c has an important effect on the long-term forecasts obtained from these models.

- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d = 1$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend.

Figure 1: Hi