Investigating ARMA Models on Commodity Data

Dario Greco, Kayne LaRose

December 2023

1 Introduction

The Bank of Canada Commodity Price Index (BCPI) is basket of 26 commodities produced in Canada and sold in international markets (Bank of Canada 2023). In this paper, ARMA models are fit from a subset of the BCPI data. The goal of this research is to capture and model the underlying trend of the data to make a reasonable 10 month CPI forecast.

2 BCPI

2.1 Basket Structure

The BCPI is achain Fisher price index measured in USD. The weight of each of the 26 commodities is a function of their impact on Canadian commodity production. This impact is measured every 4 years by Statistics Canada and subsequent weights are then projected forward (Bank of Canada 2012).

2.2 Market Volatility

Given that the BCPI time series is measuring prices, it is worth observing the impact of various economic calamaties and/or significant political events on the data throughout the series. Figure 1 shows us that the basket reacts with extreme volatility to times of financial uncertainty. Furthermore, as seen between 2010 and 2019, the basket can experience large fluctuations even when the stock market is experiencing times of relative stability. This could be a result of the fact that many of the commodities in the basket are historically uncorrelated with the US stock market (Enilov, Mensi, and Stankov 2023). Historically, gold prices have increased in response to market turmoil, producing a perceived negative correlation with the market (Baur and McDermott 2010). We also observed that other precious metals performed very well during the Covid-19 pandemic although it is hard to generalize this behavior further at this moment.

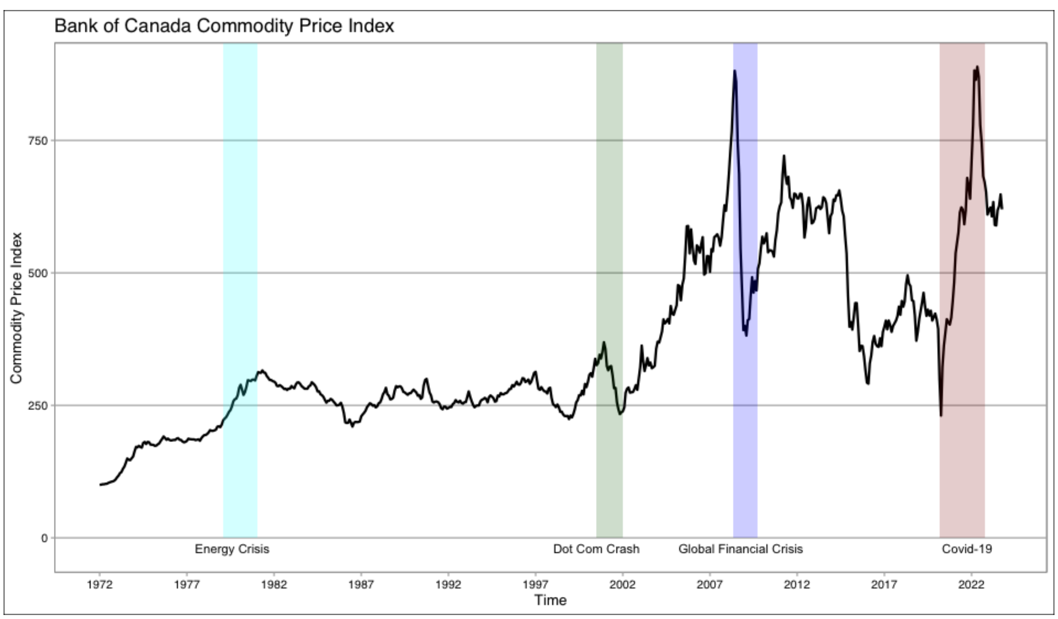
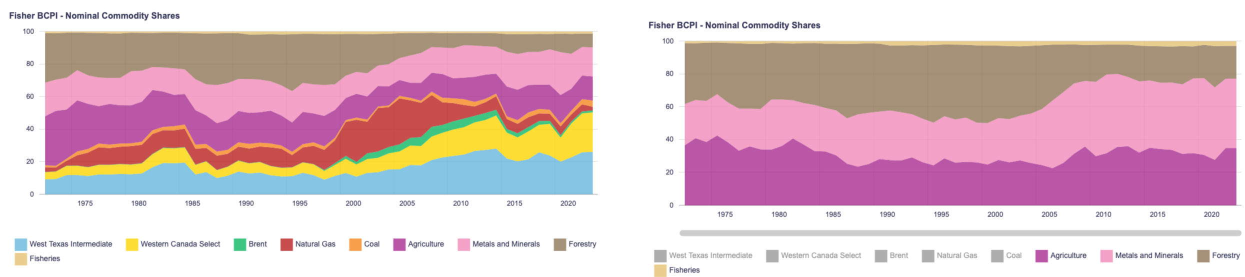


Figure 1: BCPI Time Series

On the other hand, Agriculture commodities are most sensitive to harvesting factors and therefore we can expect them to be influenced less by movements in the stock market. In fact, during the rise of the Covid-19 pandemic, both soybeans and wheat experienced a sharp increase in price (Enilov, Mensi, and Stankov 2023). This leads us to determining whether to exclude energy from the basket for analysis.

2.3 Removal of Energy Component

The Energy component of the BCPI, which includes coal and oil, has historically made up the most significant portion of the basket.



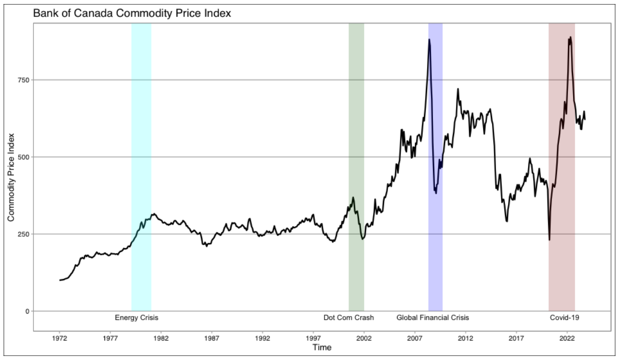
BCPI Weights

Removal of Energy

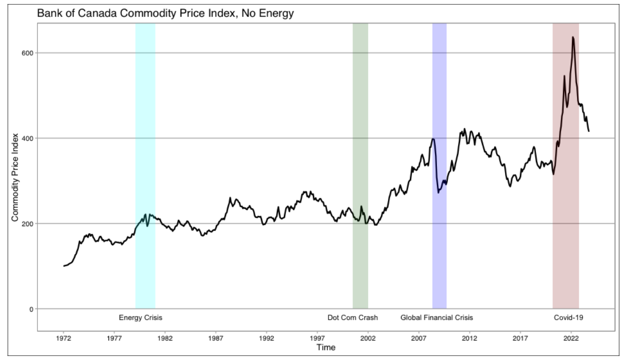
Figure 2: Comparison of CPI Weights (Bank of Canada (2023))

We chose to exclude energy from the time series given that we observe that the magnitude of effect energy had on outlier events was vastly different than the rest of the basket. For example, the wars and political instability of the Middle East in the 1970s had a great effect on energy price as well as the various decisions of OPEC in the 2010s (EIA 2023). Furthermore, it is unclear how ESG (Environment, Social, and Governance) investing strategies will affect oil prices moving forward.

The resulting time series we will analyze will be referred to as BCNE as shown in Figure 3.



BCPI



BCNE (No Energy)

Figure 3: Comparison of Time Series

3 Time Series Analysis

We will proceed with our analysis on the following time series.

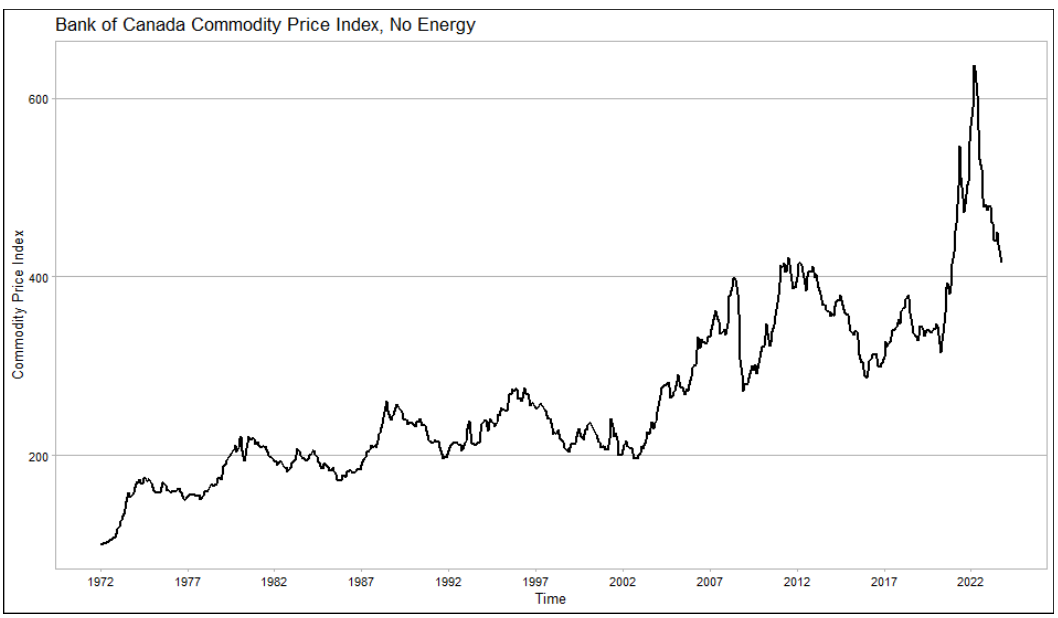


Figure 4: BCNE Time Series

3.1 Transformations

It is immediately observed that the time series is not stationary. A log transformation of the data is taken to address the heteroscedastic behavior of the time series; variance is increasing steadily as time goes on. The first difference is then taken afterwards to address the upward linear trend and to create stationarity. Figure 5 visualizes these results.

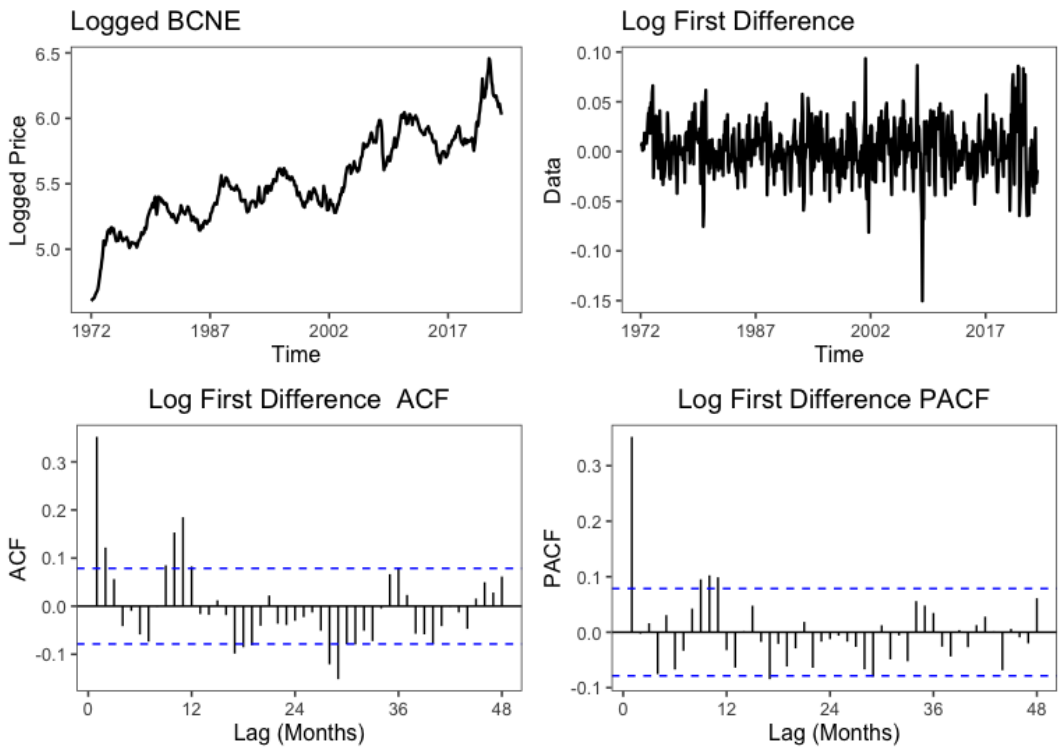


Figure 5: Transformations of BCNE

3.2 Selection

The PACF and ACF plots in Figure 5 show significant correlations in the 10-12 month range in terms of lag. We begin testing ARMA models with AR and MA parameters ranging from 0 to 12. Models that passed diagnostics without being over-fit are shown below in Table 1 along with the relevant criterion scores to assess relative model strength.

Model Selection

Model	AIC	AICc	BIC
ARMA(0,1,12)	-4.582486	-4.581660	-4.489721
ARMA(1,1,11)	-4.579341	-4.578514	-4.486575
ARMA(11,1,0)	-4.580438	-4.579740	-4.494808
ARMA(12,1,0)	-4.578124	-4.577297	-4.485358
ARMA(10,1,3)	-4.586333	-4.585217	-4.479296

Table 1: Model Selection Criterion

ARMA(10,1,3) had the lowest AIC and AICc while ARMA(11,1,0) had the lowest BIC. It should also be noticed that ARMA(0,1,12) had the second lowest AIC, AICc, and BIC. For the sake of more interesting commentary on comparing model results, we will only discuss and analyze the ARMA(10,1,3) and ARMA(0,1,12) models.

ARMA(10,1,3)

	Estimate	SE	t.value	p.value
ar1	1.6207	0.1327	12.2178	0.0000
ar2	-1.7921	0.1612	-11.1176	0.0000
ar3	1.1216	0.1862	6.0227	0.0000
ar4	-0.3588	0.1244	-2.8834	0.0041
ar5	0.2199	0.1154	1.9053	0.0572
ar6	-0.2543	0.1160	-2.1935	0.0287
ar7	0.1328	0.1165	1.1402	0.2547
ar8	-0.0042	0.1077	-0.0388	0.9691
ar9	-0.0326	0.0821	-0.3975	0.6911
ar10	0.0937	0.0442	2.1198	0.0344
ma1	-1.2849	0.1278	-10.0506	0.0000
ma2	1.3725	0.1026	13.3715	0.0000
ma3	-0.6125	0.1244	-4.9221	0.0000
constant	0.0022	0.0018	1.2513	0.2113

ARMA(0,1,12)

	Estimate	SE	t.value	p.value
ma1	0.3392	0.0400	8.4909	0.0000
ma2	0.1153	0.0421	2.7375	0.0064
ma3	0.0897	0.0420	2.1374	0.0330
ma4	-0.0407	0.0418	-0.9754	0.3297
ma5	0.0169	0.0423	0.3990	0.6900
ma6	-0.0304	0.0429	-0.7084	0.4789
ma7	-0.0666	0.0417	-1.5969	0.1108
ma8	0.0038	0.0396	0.0966	0.9231
ma9	0.0647	0.0430	1.5048	0.1329
ma10	0.1042	0.0440	2.3696	0.0181
ma11	0.1874	0.0412	4.5528	0.0000
ma12	0.1064	0.0394	2.7038	0.0070

Table 2: T-tables for ARMA(10,1,3) and ARMA(0,1,12)

Let $y_t = log(x_t)$. The explicit backshift equations for both chosen models are

ARMA(10,1,3):

$(1-B)(1-1.6207_{0.13}B+1.7921_{0.16}B^2-1.1216_{0.19}B^3+0.3588_{0.12}B^4-0.2199_{0.12}B^5+0.2543_{0.12}B^6-0.1328_{0.12}B^7+0.0042_{0.11}B^8+0.0326_{0.08}B^9-0.0937_{0.04}B^{10})(\hat{y}_t-0.0022_{0.002}) = (1-1.2849_{0.13}B+1.3725_{0.10}B^2-0.6125_{0.12}B^3)w_t$

ARMA(0,1,12):

$(1-B)\hat{y}_t = (1+0.3392_{0.040}B+0.1153_{0.042}B^2+0.0897_{0.042}B^3-0.0407_{0.042}B^4+0.0169_{0.042}B^5-0.0304_{0.043}B^6-0.0666_{0.042}B^7+0.0038_{0.040}B^8+0.0647_{0.043}B^9+0.1042_{0.044}B^{10}+0.1874_{0.041}B^{11}+0.1064_{0.039}B^{12})w_t$

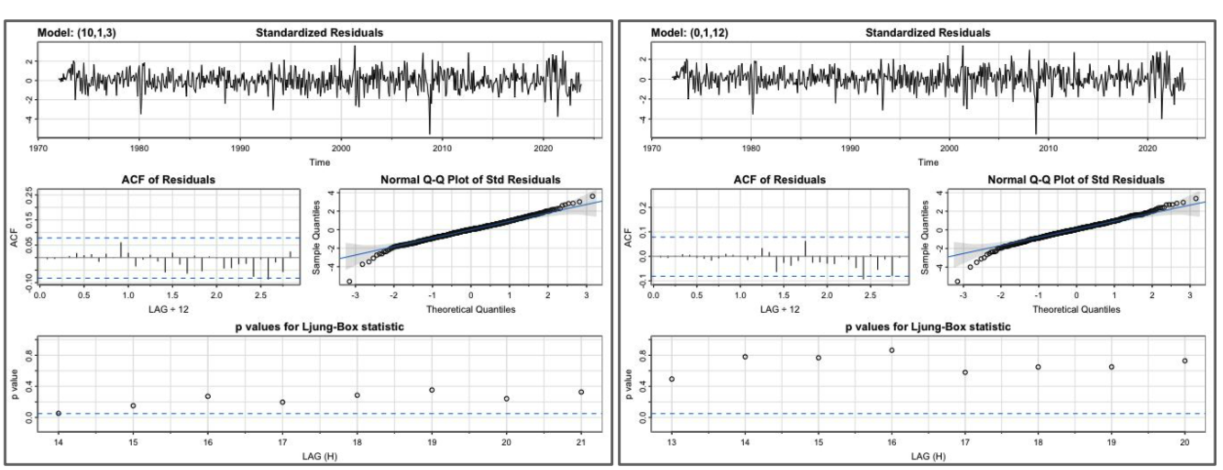


Figure 6: Diagnostics of both models

Figure 6 demonstrates that the selected models both have passable model diagnostics. However, normality is most definitely a concern that we unfortunately must concede if we wish to move forward. We argue that any observable trend in the data is the result of a rather small collection of significant outlier events, attempting to achieve normality within a model for this data would be a naive endeavor. Thus, we move on with forecasting our two chosen models.

4 Forecasting

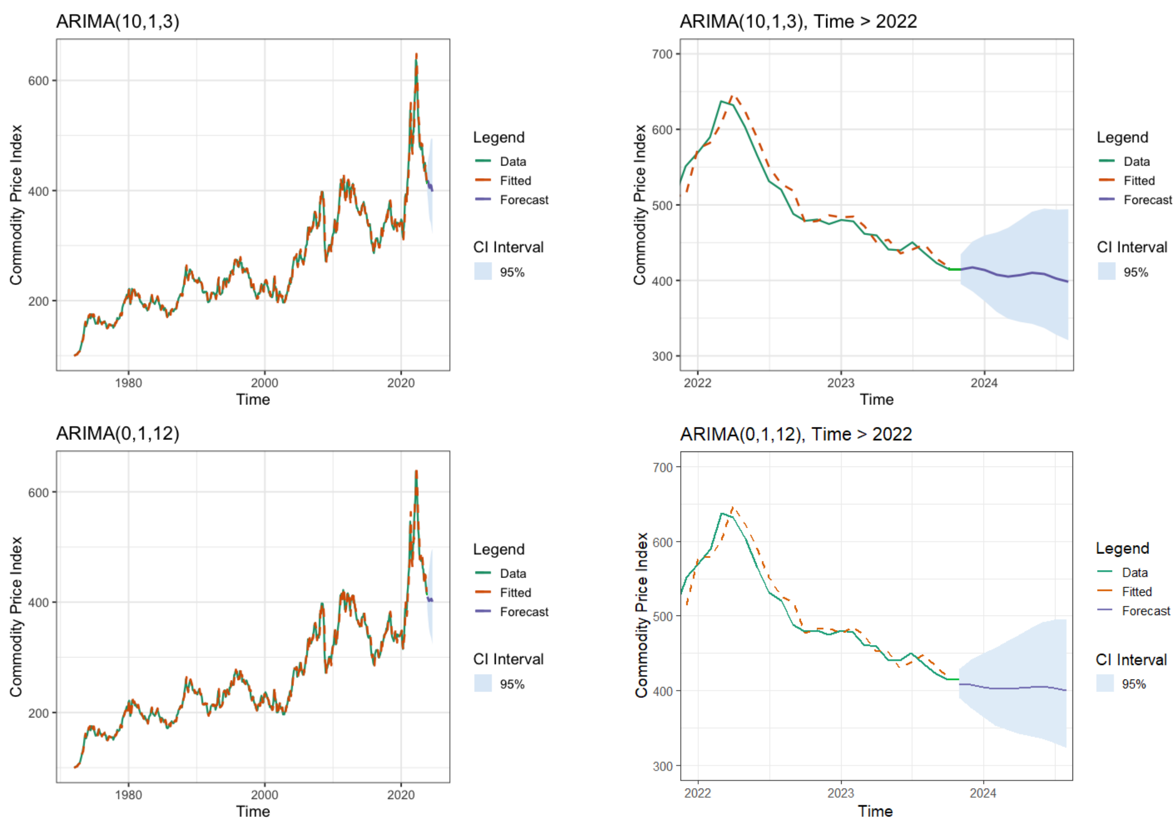


Figure 7: Forecast Models

Our models did surprisingly well on capturing the trend of the data given the complex nature of both models. We observe that, given the drastic mathematical difference between the two models, their performances are nearly identical with marginal differences which can be observed on the zoomed-in plots in Figure 8. This suggests an interesting hypothesis that these models converge to the same fit as p and q become larger.

However, it is clear that the model strength is questionable given our confidence interval has a variation of roughly ± 175 CPI when forecasting 10 months into the future. It is uncertain that these results are solid enough to be actionable in most contexts.

5 Conclusion

Commodities are notoriously hard to model. After removing energy from the BCPI basket, our time series analysis on the data became much smoother. Normality is the main concern given how the behavior of the time series is dictated by a handful of outlier events. Regardless, the results of the model forecast fit with the general trend of the price index. That is, both models predict a downward trend in commodity prices for the near future. In subsequent studies, it would be useful to look for underlying seasonality trends or implement tools to capture super cycle structures within the time series in regards to creating stronger models for this particular data set.

References

Baur, Dirk G and Thomas K McDermott (2010). “Is gold a safe haven? International evidence”. In: *Journal of Banking & Finance* 34.8, pp. 1886–1898.

Bank of Canada (2012). *Monetary Policy Report*. https://www.bankofcanada.ca/wp-content/uploads/2012/07/mpr-july2012.pdf. pg, 17.

– (2023). *Commodity Price Index*. https://www.bankofcanada.ca/rates/price-indexes/bcpi. Accessed : 2023-12-05.

EIA (2023). *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*. https://www.eia.gov/finance/markets/crudeoil/spot_prices.php. Accessed : 2023-12-05.

Enilov, Martin, Walid Mensi, and Petar Stankov (2023). “Does safe haven exist? Tail risks of commodity markets during COVID-19 pandemic”. In: *Journal of Commodity Markets* 29, p. 100307.

Appendix: R Code

```r
packages <- c("astsa", "dplyr", "ggplot2",
              "timeSeries", "forecast",
              "ggfortify", "grid",
              "ggthemes", "reshape2", "ggforce",
              "scales", "lubridate", "tidyquant",
              "cowplot", "knitr",  "kableExtra")
invisible(lapply(packages, library, character.only = TRUE))
# Need to install individually if not installed
#The following functions mainly used for decomposition to make the code
# more readable.


#Template for time series plots
graph_aes <- function(data, title, x, y){
  if(missing(x)){
    x = "Time"
  }
  if(missing(y)){
    y = "Commodity␣Price␣Index"
  }

  autoplot(data, data.colour = "black", size = 0.8,
           data.linetype = "solid", xlab = x, ylab = y) +
    ggtitle(title) +
    theme_calc() + #Theme to make it look pretty
    scale_x_date(breaks = seq(as.Date("1972-01-01"),
                              as.Date("2023-10-01"), by="5␣years"),
                 labels=label_date("%Y"))# more frequent time intervals
}

#Template for ACF/PACF plots
acf_aes <- function(ts, title, p = FALSE){
  x <- ggAcf
  y <- '␣ACF'
  if(p == TRUE){
    x <- ggPacf
    y <- 'PACF'
  }
  return(
    x(ts, lag = 48) + xlab("Lag␣(Months)") +
      ggtitle(paste(title,y)) +
      theme_few() +
      theme(plot.title = element_text(hjust = 0.5))
  )
}
#Template for forecast plots. Includes the function
#to zoom around the forecast period.
forecast_plot_aes <- function(fs, title, zoom = FALSE){
```

```r
  #Allows me to group data,fitted,forecast for Legend
  forecast <- rename(fortify(fs), "Forecast" = "Point␣Forecast",
                     "date" = "Index")

  fsplot <- ggplot(forecast, aes(x = date)) +
    xlab("Time") +
    ylab("Commodity␣Price␣Index") +
    ggtitle(title) +
    geom_ribbon(aes(ymin = `Lo 95`, ymax = `Hi 95`, fill = "95%")) +
    geom_line(aes(y = Data, group = 1, colour = "Data"),
              linetype = "solid", size = 0.8) +
    geom_line(aes(y = Fitted, group = 2, colour = "Fitted",),
              linetype = "dashed", size = 0.9) +
    geom_line(aes(y = Forecast, group = 3, colour = "Forecast"),
              size = 1) +

    #Connects empty space between data and forecast.
    geom_segment(aes(x = as.Date('2023-10-01'),
                     xend = as.Date("2023-11-01"),
                     y = 415.01, yend = 414.3796),
                 size = 0.8, colour ="#00BA38" ) +
    #Themes and Legend color themes to make it prettier.
    guides(colour = guide_legend(order = 1),
           fill = guide_legend(order = 2)) +
    scale_colour_brewer(name = "Legend",
                        type = "qual", palette = "Dark2") +
    scale_fill_brewer(name = "CI␣Interval") +
    theme_bw(base_size = 14)

  # Zoom
  if(zoom == TRUE){
    fsplot <- fsplot +
      coord_x_date(xlim = c(as.Date('2022-01-01'),
                            as.Date('2024-07-01'))) +
      scale_y_continuous(limits = c(300, 700))
  }
  fsplot

}


###Following ggplot objects to combine into annotated graph###

#Covid19
covid_19_obj <- annotate("text", x = as.Date('2020-03-01') + 580,
                         y = -20,
                         label = "Covid-19", color = "black",
                         size = 3, fontface = 'plain')
covid_rect <- annotate("rect", xmin=as.Date('2020-03-01'),
                       xmax = as.Date('2022-10-01'),
                       ymin = 0, ymax = Inf,
```

```r
                       fill = "darkred", alpha = 0.2)

#2008 Housing Market Crash
glob_cris_obj <- annotate("text", x = as.Date('2008-03-24') + 200,
                          y = -20,
                          label = "Global␣Financial␣Crisis",
                          color = "black",
                          size = 3, fontface = 'plain')
glob_rect <- annotate("rect", xmin=as.Date('2008-04-24'),
                      xmax = as.Date('2009-10-01'),
                      ymin = 0, ymax = Inf, fill = "blue",
                      alpha = 0.2)

#Dot Com Bubble
dot_com_obj <- annotate("text", x = as.Date('2000-03-24') + 100,
                        y = -20,
                        label = "Dot␣Com␣Crash", color = "black",
                        size = 3, fontface = 'plain')
dot_com_rect <- annotate("rect", xmin=as.Date('2000-06-24'),
                         xmax = as.Date('2002-01-01'),
                         ymin = 0, ymax = Inf, fill = "darkgreen",
                         alpha = 0.2)


#Energy Crisis 1979
ener_obj <- annotate("text", x = as.Date('1979-01-24') + 200, y
                     = -20,
                     label = "Energy␣Crisis", color = "black",
                     size = 3, fontface = 'plain')
ener_rect <- annotate("rect", xmin=as.Date('1979-01-24'),
                      xmax = as.Date('1981-01-24'),
                      ymin = 0, ymax = Inf, fill = "cyan",
                      alpha = 0.2)

world_events <- c(covid_19_obj, covid_rect, glob_cris_obj, glob_rect,
                  dot_com_rect, dot_com_obj, ener_rect, ener_obj)




#Dataset and respective indexes for isolated commodity data
BCPI_full <- as.data.frame(read.csv("M.BCPI.csv"))
BCPI <- 2 # Entire basket
BCNE <- 3 # No Energy <- Chosen data
ENER <- 4 # Only Energy
METAL <- 5 # Only metals/minerals
FORSTRY <- 6 # Only forestry
AGRI <- 7 # Only agriculture
FISH <- 8 # Only fish
```

```r
BCPI_ts <- ts(BCPI_full[,BCPI], start = 1972, frequency = 12)
BCNE_ts <- ts(BCPI_full[,BCNE], start = 1972, frequency = 12)

#Time series plots

graph_aes(BCPI_ts, "Bank of Canada Commodity Price Index")
graph_aes(BCNE_ts, "Bank of Canada Commodity Price Index, No Energy")

#Annotated graphs highlighting world events
graph_aes(BCPI_ts, "Bank of Canada Commodity Price Index") +
  world_events
graph_aes(BCNE_ts, "Bank of Canada Commodity Price Index, No Energy") +
  world_events

#Transformations on BCNE
log_plot <- graph_aes(log(BCNE_ts), "Logged BCPI", y = "Logged CPI") +
 theme_clean() +
  #Fix time scaling so that it doesn't get squished.
 scale_x_date(breaks = seq(as.Date("1972-01-01"),
                           as.Date("2023-10-01"),
                           by="15 years"),
              labels=label_date("%Y"))

log_diff <- graph_aes(diff(log(BCNE_ts)),
                      "First Difference",
                      y = "Data") +
  theme_clean() +
  scale_x_date(breaks = seq(as.Date("1972-01-01"),
                            as.Date("2023-10-01"),
                            by="15 years"),
               labels=label_date("%Y"))

acf_ld <- acf_aes(diff(log(BCNE_ts)),
                  "Log First Difference")
pacf_ld <- acf_aes(diff(log(BCNE_ts)),
                   "Log First Difference", p = TRUE)

#Plots all four
plot_grid(log_plot, log_diff, acf_ld,pacf_ld,
          labels = c('','', '', ''))

#Diagnostics / Potential Models

#Logged BCNE Data will be our chosen data to model
main_ts <- log(BCNE_ts)
```

```
#Handful of models that were able to pass diagnostics
#Note: ARIMA(11,1,2), ARIMA(12,1,1) are both OVERFIT


M1 <- sarima(main_ts, 0,1,12, no.constant = TRUE)
M2 <- sarima(main_ts, 0,1,11, no.constant = TRUE)
M3 <- sarima(main_ts, 11,1,0, no.constant = TRUE)
M4 <- sarima(main_ts, 12,1,0, no.constant = TRUE)
M5 <- sarima(main_ts, 10,1,3, no.constant = TRUE)

#Differentiate model performance by comparing criterion
AICv <- c(M1$AIC, M2$AIC, M3$AIC, M4$AIC, M5$AIC)
BestAIC <- order(AICv)[1:3]
## Models M5,M1,M3 are best (in order)

#Some models had Ljung-Box statistics plots with suspicious plots near p=0.

# Extracts Ljung - Box statistic p -values for an ARMA(p,d,q)
LJB_pvalues <- function(Model, max.lag, df){
  resd = Model[["fit"]][["residuals"]]
  pv <- c()
  for(i in df+1: max.lag){
    p <- Box.test(resd, i, type = "Ljung-Box", fitdf = df)$p.value
    pv = append(pv, p)
  }
  return(pv)
}

# Extract p-values up to lag 35
p5  = LJB_pvalues(M5, 35, 13)
p1 = LJB_pvalues(M1, 35, 14)
p3 = LJB_pvalues(M3, 35, 11)
## p -value >0.05 at higher lags for all of Model 5, 1, 3


#Forecasting
#ARIMA(10,1,3)
m5_forecast <- Arima(BCNE_ts, c(10,1,3), lambda = 0) %>%
                forecast(., level= c(95), h = 10)
#MA(12)
m1_forecast <- Arima(BCNE_ts, c(0,1,12), lambda = 0) %>%
                forecast(.,level = c(95), h=10)
#AR(11)
m3_forecast <- Arima(BCNE_ts, c(11,1,0), lambda = 0) %>%
               forecast(.,level = c(95), h=10)

#Sample Forecasts
forecast_plot_aes(m5_forecast, "ARIMA(10,1,3)")
forecast_plot_aes(m5_forecast, "ARIMA(10,1,3),␣Time␣>␣2022", zoom = TRUE)
```

```
forecast_plot_aes(m1_forecast, "ARIMA(0,1,12)")
forecast_plot_aes(m1_forecast, "ARIMA(0,1,12),␣Time␣>␣2022", zoom = TRUE)

forecast_plot_aes(m3_forecast, "ARIMA(11,1,0)")
forecast_plot_aes(m3_forecast, "ARIMA(11,1,0),␣Time␣>␣2022", zoom = TRUE)



#Code for tables
AICv <- c(M1$AIC, M2$AIC, M3$AIC, M4$AIC, M5$AIC)
AICCv <- c(M1$AICc, M2$AICc, M3$AICc, M4$AICc, M5$AICc)
BICv <- c(M1$BIC, M2$BIC, M3$BIC, M4$BIC, M5$BIC)
BestAIC <- order(AICv)[1:3] ## Models 5,1,3 are best (in order)
BestAICC <- order(AICCv)[1:3]
BestBIC <- order(BICv)[1:3]
names <- c("ARMA(0,1,12)", "ARMA(1,1,11)",
           "ARMA(11,1,0)", "ARMA(12,1,0)", "ARMA(10,1,3)")
crit<- data.frame(Model = names, AIC = AICv, AICc = AICCv, BIC = BICv)

table <- kable(crit, format = "html",
               caption = "<b␣style='color:black;'>Model␣Selection</b>") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                html_font = "arial")
tablet <- kable(M5$ttable, format = "html",
                caption = "<b␣style='color:black;'>ARMA(10,1,3)</b>") %>%
  kable_styling(bootstrap_options = c("striped", "hover"),
                html_font = "arial")
tablet2 <- kable(M1$ttable, format = "html",
                 caption = "<b␣style='color:black;'>ARMA(0,1,12)</b>") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```