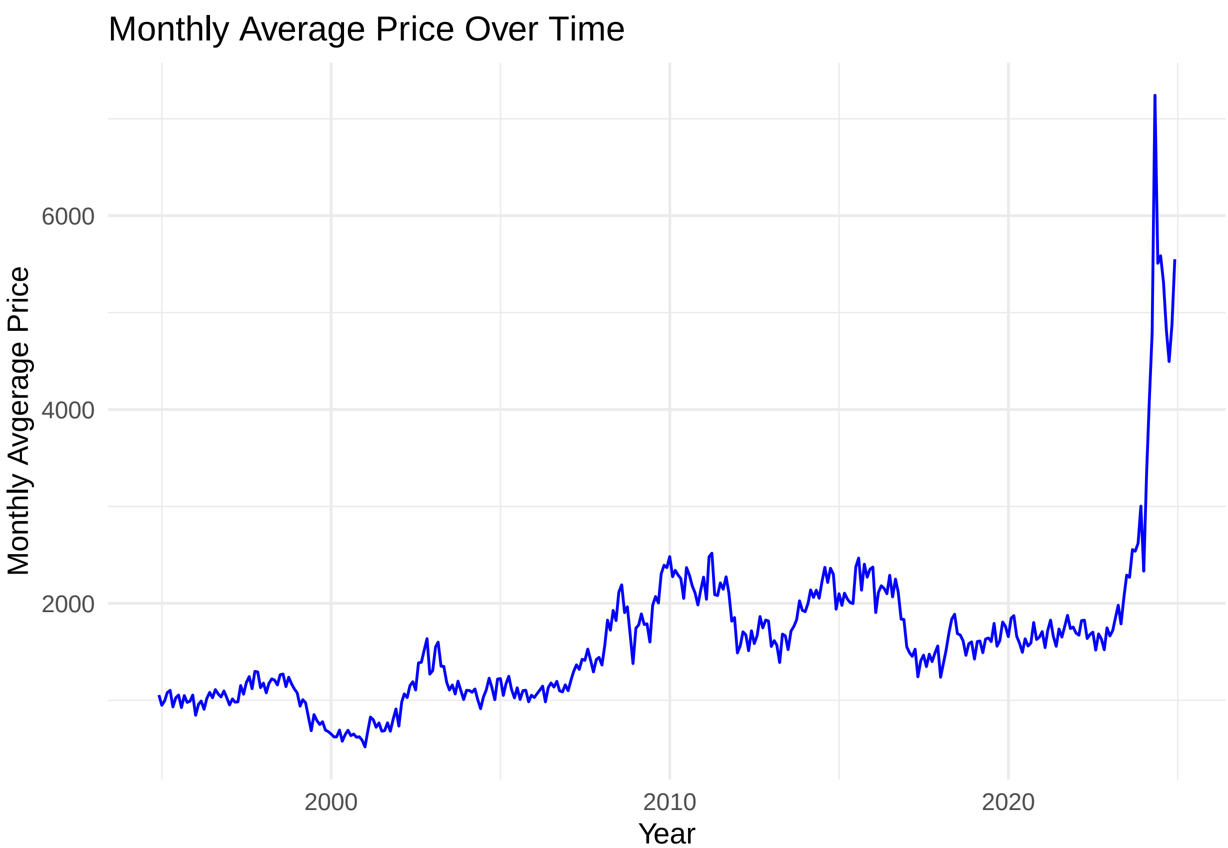
STA457 Final Project - ETS Model

Andrew Hu

```r
cocoa_data <- read.csv("cocoa data ets.csv")
cocoa_data$date <- as.Date(cocoa_data$date, format = "%m/%d/%Y")
cocoa_data$Price <- as.numeric(gsub(",", "", cocoa_data$Price_Monthly_Avg))

cocoa_data <- cocoa_data %>%
  mutate(log_price = log(Price),
         diff_log_price = c(NA, diff(log_price))) %>%
  drop_na()
```

```r
ggplot(cocoa_data, aes(x = date, y = Price_Monthly_Avg)) +
  geom_line(color = "blue") +
  ggtitle("Monthly Average Price Over Time") +
  xlab("Year") +
  ylab("Monthly Avgerage Price") +
  theme_minimal()
```

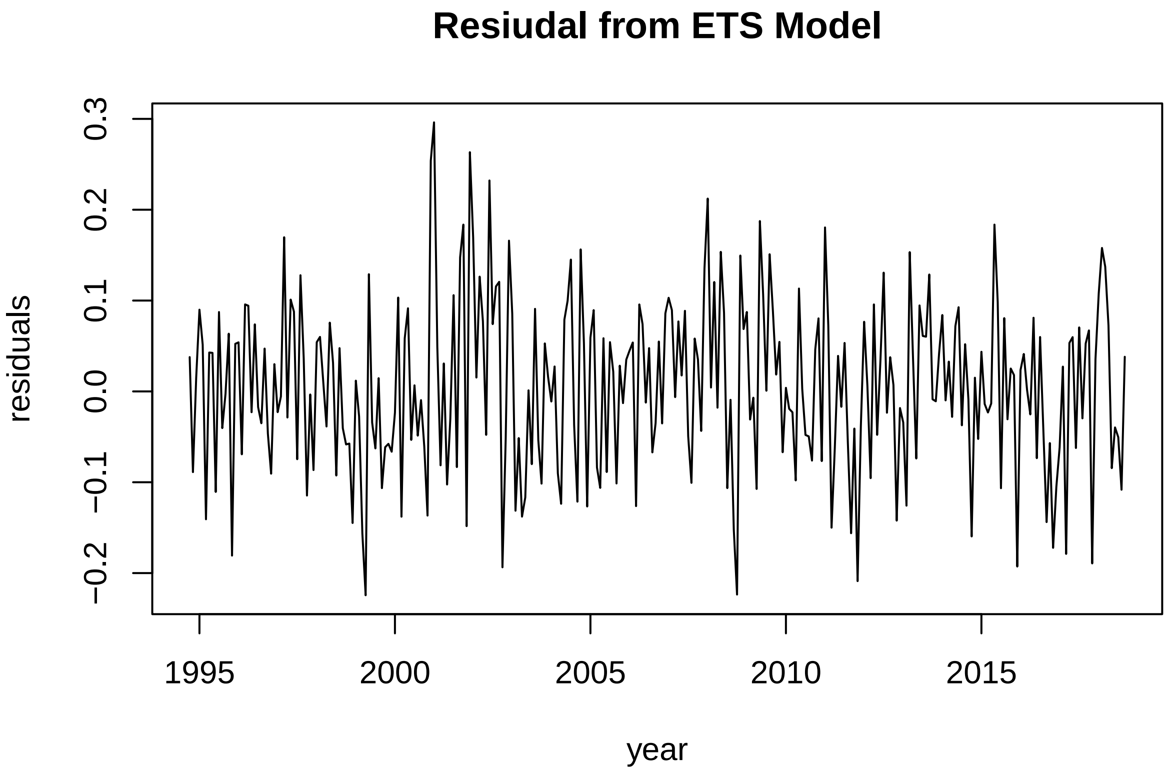



```r
price_ts <- ts(cocoa_data$Price_Monthly_Avg, start = c(1994, 10), frequency = 12)
```

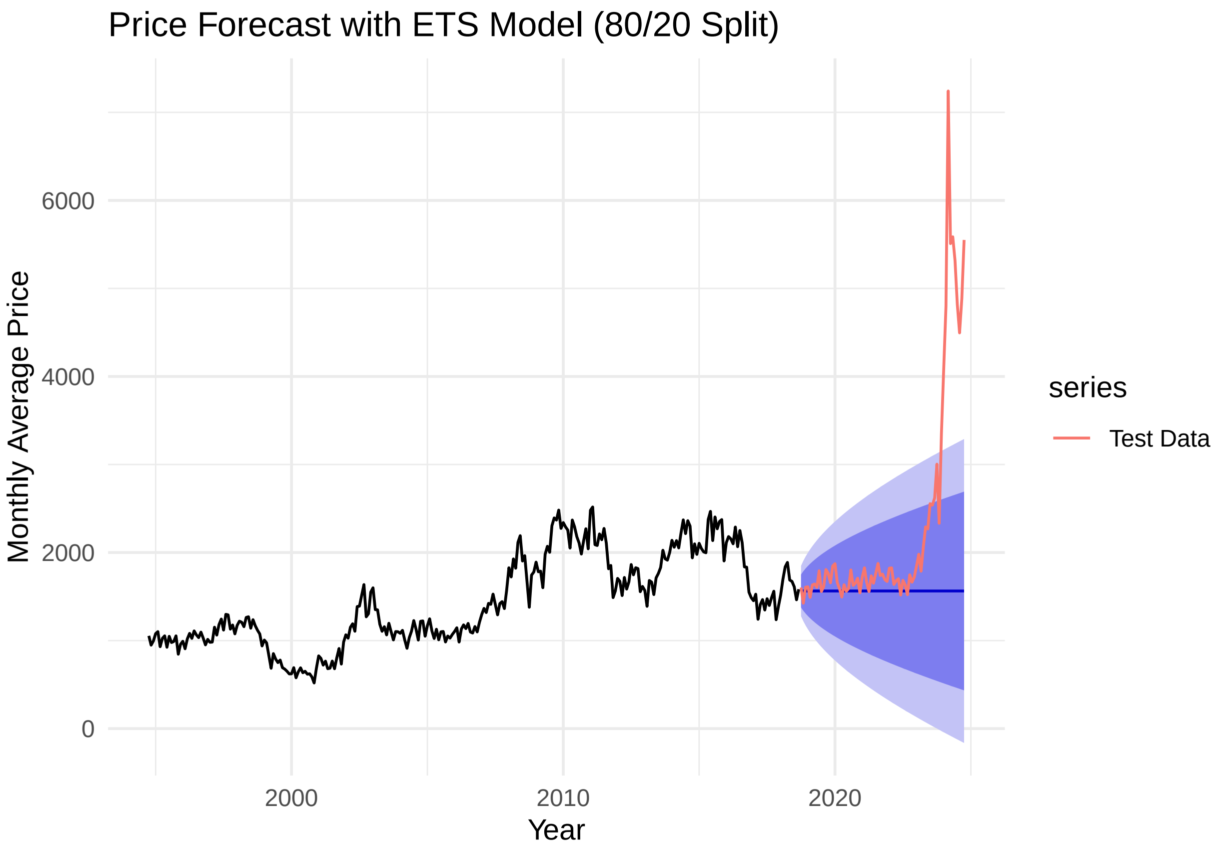
```r
n <- length(price_ts)
split_index <- floor(0.8 * n)
train_data <- window(price_ts, end = time(price_ts)[split_index])
test_data <- window(price_ts, start = time(price_ts)[split_index + 1])

ets_model <- ets(train_data, model = "ZZZ")
summary(ets_model)
```

```
## ETS(M,N,N)
##
## Call:
##  ets(y = train_data, model = "ZZZ")
##
##   Smoothing parameters:
##     alpha = 0.6515
##
##   Initial states:
##     l = 1015.4246
##
##   sigma:  0.0936
##
##      AIC     AICc      BIC
## 4421.590 4421.675 4432.579
##
## Training set error measures:
##                    ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 2.919422 136.2439 106.3791 -0.3609434 7.638507 0.4127703
##                    ACF1
## Training set 0.03449358
```

```r
plot(residuals(ets_model), main = "Resiudal from ETS Model", xlab = "year", ylab = "residuals")
```




```r
forecast1 <- forecast(ets_model, h = length(test_data))

autoplot(forecast1) +
  autolayer(test_data, series = "Test Data") +
  ggtitle("Price Forecast with ETS Model (80/20 Split)") +
  xlab("Year") +
  ylab("Monthly Average Price") +
  theme_minimal()
```




```r
accuracy(forecast1, test_data)
```

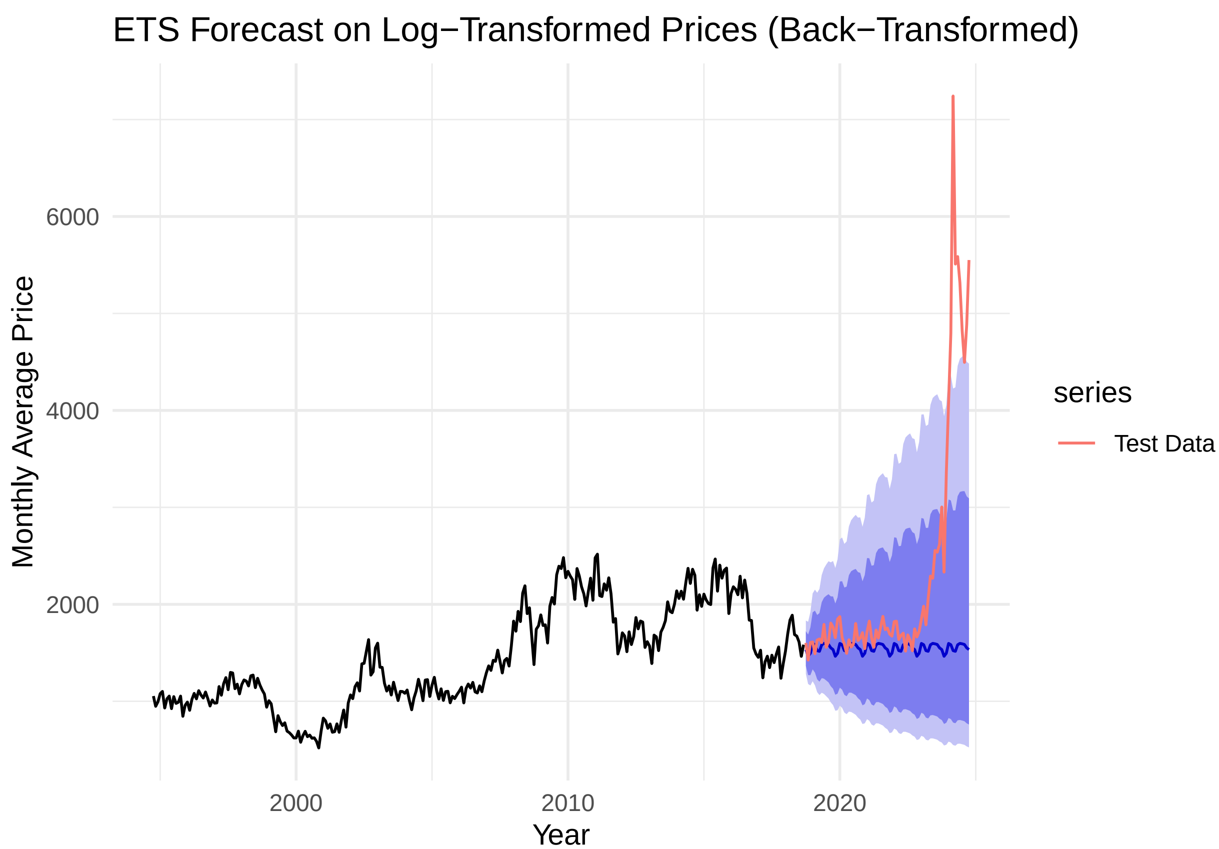
```
##                      ME      RMSE      MAE        MPE      MAPE      MASE
## Training set   2.919422  136.2439 106.3791 -0.3609434  7.638507 0.4127703
## Test set     712.279961 1446.2536 723.3655 19.0506738 19.800260 2.8067899
##                    ACF1 Theil's U
## Training set 0.03449358        NA
## Test set     0.89187094  2.731099
```

```r
log_price_ts <- ts(cocoa_data$log_price, start = c(1994, 10), frequency = 12)


n <- length(price_ts)
split_index <- floor(0.8 * n)
log_train_data <- window(log_price_ts, end = time(log_price_ts)[split_index])
log_test_data <- window(log_price_ts, start = time(log_price_ts)[split_index + 1])

log_ets_model <- ets(log_train_data, model = "ZZZ")
summary(log_ets_model)
```

```
## ETS(A,N,A)
##
## Call:
##  ets(y = log_train_data, model = "ZZZ")
##
##   Smoothing parameters:
##     alpha = 0.6995
##     gamma = 1e-04
```

```
##
##   Initial states:
##     l = 6.9154
##     s = 0 0.0221 0.0265 0.0292 0.0201 -0.0232
##            -0.0186 0.0206 0.0289 -0.0366 -0.0576 -0.0115
##
##   sigma:  0.0909
##
##      AIC     AICc      BIC
## 265.5331 267.2978 320.4775
##
## Training set error measures:
##                      ME       RMSE        MAE        MPE     MAPE      MASE
## Training set 0.00214092 0.08869351 0.07231002 0.01998717 1.009136 0.3876736
##                     ACF1
## Training set -0.07071379
```

```r
log_forecast <- forecast(log_ets_model, h = length(log_test_data))

# Back-transform
log_forecast$mean <- exp(log_forecast$mean)
log_forecast$lower <- exp(log_forecast$lower)
log_forecast$upper <- exp(log_forecast$upper)
log_forecast$x <- exp(log_forecast$x)

autoplot(log_forecast) +
  autolayer(test_data, series = "Test Data") +
  ggtitle("ETS Forecast on Log-Transformed Prices (Back-Transformed)") +
  xlab("Year") +
  ylab("Monthly Average Price") +
  theme_minimal()
```




```r
# Compare to test data

accuracy(log_forecast, test_data)
```

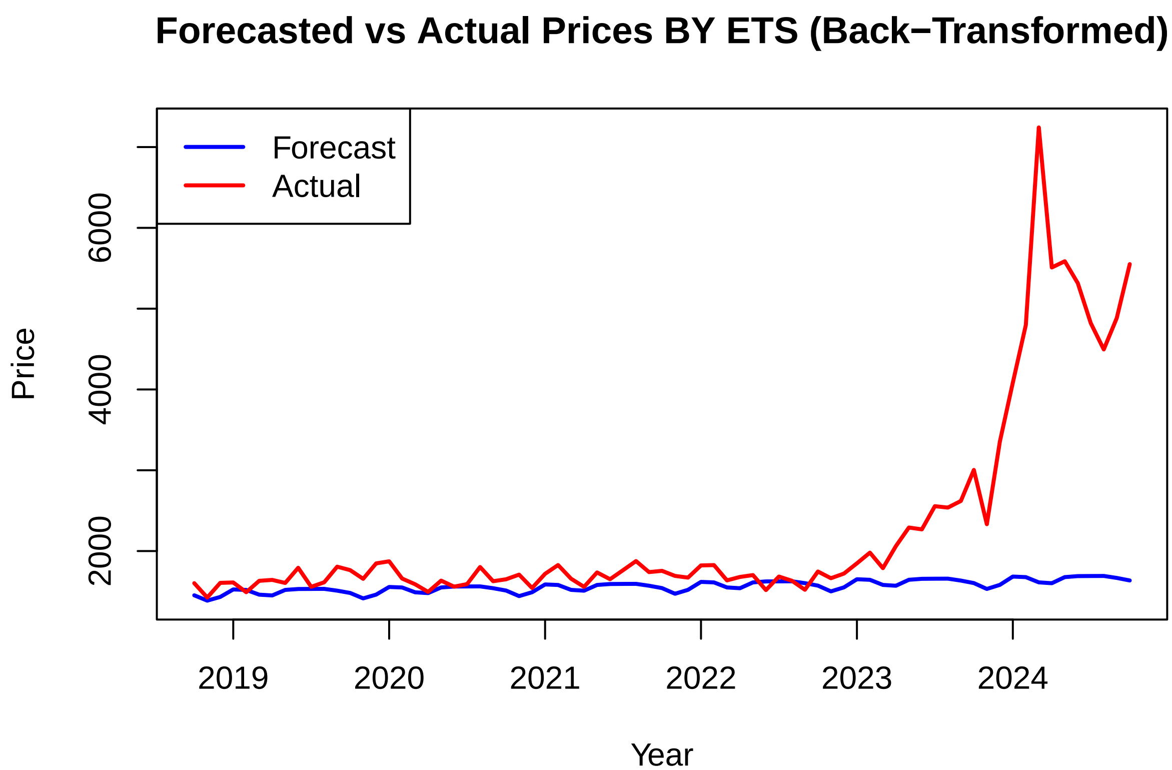
```
##                    ME     RMSE       MAE      MPE     MAPE     MASE      ACF1
## Training set 1433.911 1521.199 1433.9110 99.43856 99.43856 5.563835 0.9593853
## Test set      724.049 1449.538  733.3112 19.74428 20.35921 2.845381 0.8913310
##              Theil's U
## Training set        NA
## Test set      2.749785
```

```r
diff_log_price <- ts(cocoa_data$diff_log_price, start = c(1994, 10), frequency = 12)

n <- length(diff_log_price)
split_index <- floor(0.8 * n)
diff_log_train <- window(diff_log_price, end = time(price_ts)[split_index])
diff_log_test <- window(diff_log_price, start = time(price_ts)[split_index + 1])

diff_log_ets_model <- ets(diff_log_train, model = "ZZZ")
summary(diff_log_ets_model)
```

```
## ETS(A,N,A)
##
## Call:
##  ets(y = diff_log_train, model = "ZZZ")
##
##   Smoothing parameters:
##     alpha = 1e-04
```

```
##     gamma = 1e-04
##
##   Initial states:
##     l = 0.0017
##     s = -0.0161 -0.0013 -9e-04 0.0056 0.0443 -0.0081
##            -0.0411 -0.0058 0.0611 0.031 -0.0481 -0.0207
##
##   sigma:  0.0967
##
##      AIC     AICc      BIC
## 301.0174 302.7821 355.9618
##
## Training set error measures:
##                         ME       RMSE        MAE      MPE     MAPE      MASE
## Training set -0.0002304408 0.09432926 0.07645199 58.53721 162.9805 0.7983175
##                    ACF1
## Training set -0.3195361
```

```r
diff_log_forecast <- forecast(diff_log_ets_model, h = length(diff_log_test))

forecasted_diffs <- diff_log_forecast$mean

last_log_price <- tail(cocoa_data$log_price, 1)

# Reconstruct & back-transform
log_price_forecast <- ts(cumsum(forecasted_diffs) + 7.3,
                         start = time(diff_log_test)[1], frequency = 12)

price_forecast <- exp(log_price_forecast)
actual_price_test <- window(price_ts, start = time(diff_log_test)[1])

# Plot forecast vs actual
plot(price_forecast, col = "blue", lwd = 2, ylim = range(c(price_forecast, actual_price_test)),
     main = "Forecasted vs Actual Prices BY ETS (Back-Transformed)", ylab = "Price", xlab = "Year")
lines(actual_price_test, col = "red", lwd = 2)
legend("topleft", legend = c("Forecast", "Actual"), col = c("blue", "red"), lty = 1, lwd = 2)
```




```
accuracy(price_forecast, actual_price_test)
```

```
##                ME     RMSE      MAE      MPE     MAPE      ACF1 Theil's U
## Test set 710.5973 1410.485 716.5764 19.69213 20.08671 0.8878682  2.666692
```