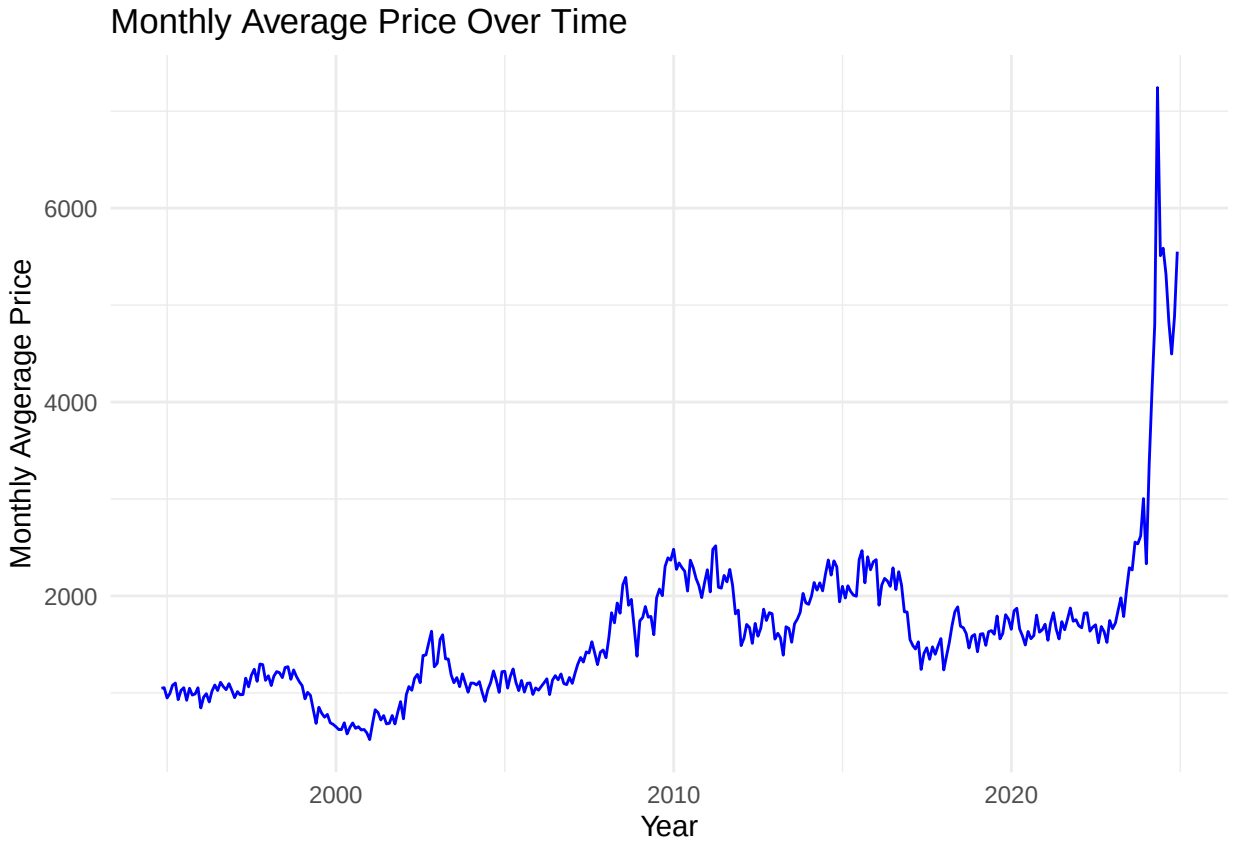
STA457 Final Project - ETS Model

Andrew Hu

```
cocoa_data <- read.csv("merged_df.csv")
cocoa_data$Date <- as.Date(cocoa_data$Date)

cocoa_data <- cocoa_data %>%
  mutate(across(everything(), ~replace_na(., 0))) %>%
  mutate(log_price = log(Price_Monthly_Avg),
         diff_log_price = c(NA, diff(log_price)))
```

```
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()
```




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

# Split into training and testing
```

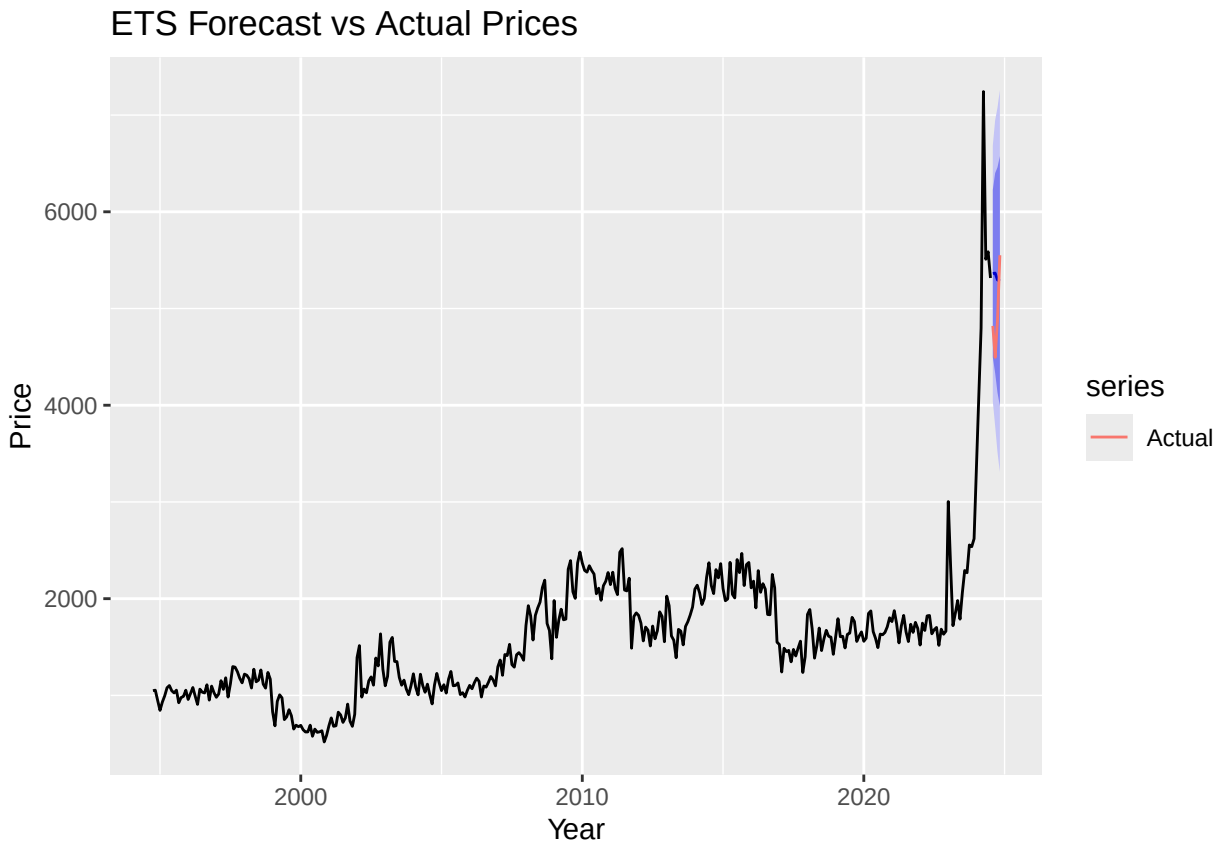
```r
train_ts <- window(price_ts, end = c(2024, 7))
test_ts <- window(price_ts, start = c(2024, 8))

# Fit ETS model
ets_model <- ets(train_ts)
summary(ets_model)
```

```
## ETS(M,N,M)
##
## Call:
##  ets(y = train_ts)
##
##   Smoothing parameters:
##     alpha = 0.6505
##     gamma = 1e-04
##
##   Initial states:
##     l = 1011.9858
##     s = 0.9889 0.9883 1.0002 0.9913 1.0193 1.0423
##            1.0054 1.026 1.0525 0.9363 0.9738 0.9759
##
##   sigma:  0.1258
##
##      AIC     AICc      BIC
## 5845.026 5846.430 5903.234
##
## Training set error measures:
##                    ME     RMSE      MAE        MPE     MAPE      MASE      ACF1
## Training set 18.63841 247.8444 147.1088 -0.2438118 9.038622 0.4735246 0.1452572
```

```r
# Forecast next 4 months
forecast_ets <- forecast(ets_model, h = 4)

# Optional: plot forecast vs actuals
autoplot(forecast_ets) +
  autolayer(test_ts, series = "Actual") +
  ggtitle("ETS Forecast vs Actual Prices") +
  ylab("Price") +
  xlab("Year")
```




```
accuracy(forecast_ets, test_ts)
```

```
##                       ME     RMSE      MAE        MPE      MAPE      MASE
## Training set   18.63841 247.8444 147.1088 -0.2438118  9.038622 0.4735246
## Test set     -389.22173 568.4565 522.8038 -8.5558544 10.962267 1.6828388
##                   ACF1 Theil's U
## Training set 0.1452572        NA
## Test set     0.1018738  1.197426
```

```
# previous attempt with 80/20 train test set proportion
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)
```

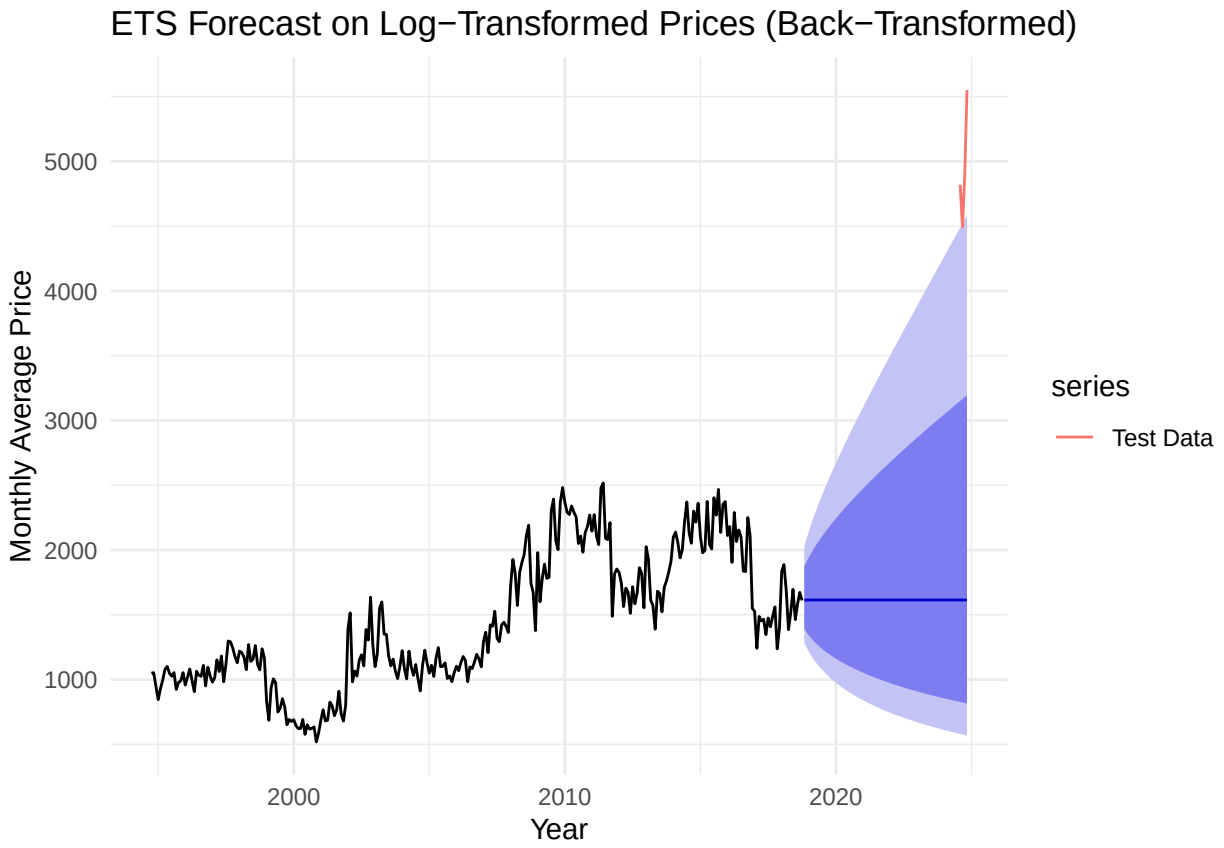
```
## ETS(A,N,N)
##
## Call:
##  ets(y = log_train_data, model = "ZZZ")
##
##   Smoothing parameters:
##     alpha = 0.5272
```

```
##
##   Initial states:
##     l = 6.9288
##
##   sigma:  0.1162
##
##      AIC     AICc      BIC
## 397.3856 397.4698 408.3848
##
## Training set error measures:
##                       ME      RMSE        MAE        MPE     MAPE      MASE
## Training set 0.003005167 0.1157786 0.08701451 0.02268668 1.211463 0.4618273
##                    ACF1
## Training set 0.1211796
```

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




```
accuracy(log_forecast, test_ts)
```

```
##                     ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
## Training set 1432.554 1519.805 1432.554 99.43789 99.43789  5.476696 0.93708346
## Test set     3323.246 3345.316 3323.246 67.11449 67.11449 12.704865 0.07274434
##              Theil's U
## Training set        NA
## Test set      7.076345
```

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

diff_log_train <- window(diff_log_price, end = c(2024, 7))
diff_log_test <- window(diff_log_price, start = c(2024, 8))

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

```
## Warning in ets(diff_log_train, model = "ZZZ"): Missing values encountered.
## Using longest contiguous portion of time series
```

```
summary(diff_log_ets_model)
```

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

```
##   Initial states:
##     l = 0.0045
##
##   sigma:  0.1249
##
##      AIC     AICc      BIC
## 616.8982 616.9662 628.5314
##
## Training set error measures:
##                       ME      RMSE        MAE      MPE    MAPE      MASE
## Training set 5.495675e-05 0.1245226 0.09376588 98.83632 102.804 0.6867329
##                     ACF1
## Training set -0.1693585
```

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

# Get forecasted differenced log prices
forecasted_diffs <- diff_log_forecast$mean

# reconstruct & back transform
last_log_price <- tail(window(cocoa_data$log_price, end = c(2024, 7)), 1)
```

```
## Warning in window.default(cocoa_data$log_price, end = c(2024, 7)): 'end' value
## not changed
```

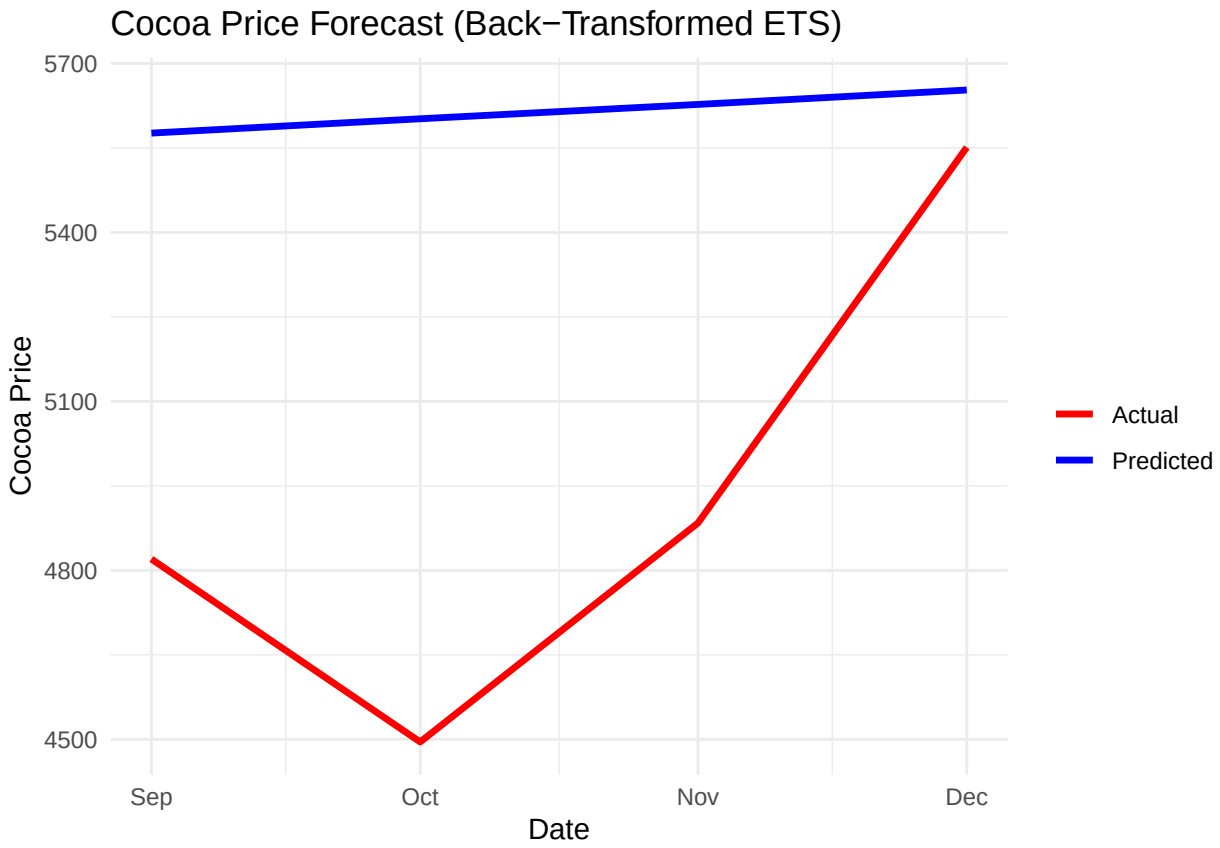
```r
log_price_forecast <- ts(cumsum(forecasted_diffs) + last_log_price,
                         start = time(diff_log_test)[1], frequency = 12)
price_forecast <- exp(log_price_forecast)

# Actual prices
actual_price_test <- window(price_ts, start = time(diff_log_test)[1])

# Forecast vs actual price
plot_df <- data.frame(
  Date = seq.Date(from = as.Date("2024-09-01"), by = "month", length.out = 4),
  Actual = as.numeric(actual_price_test),
  Predicted = as.numeric(price_forecast)
)

ggplot(plot_df, aes(x = Date)) +
  geom_line(aes(y = Actual, color = "Actual"), size = 1.2) +
  geom_line(aes(y = Predicted, color = "Predicted"), size = 1.2) +
  labs(title = "Cocoa Price Forecast (Back-Transformed ETS)",
       y = "Cocoa Price",
       x = "Date",
       color = "") +
  scale_color_manual(values = c("Actual" = "red", "Predicted" = "blue")) +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

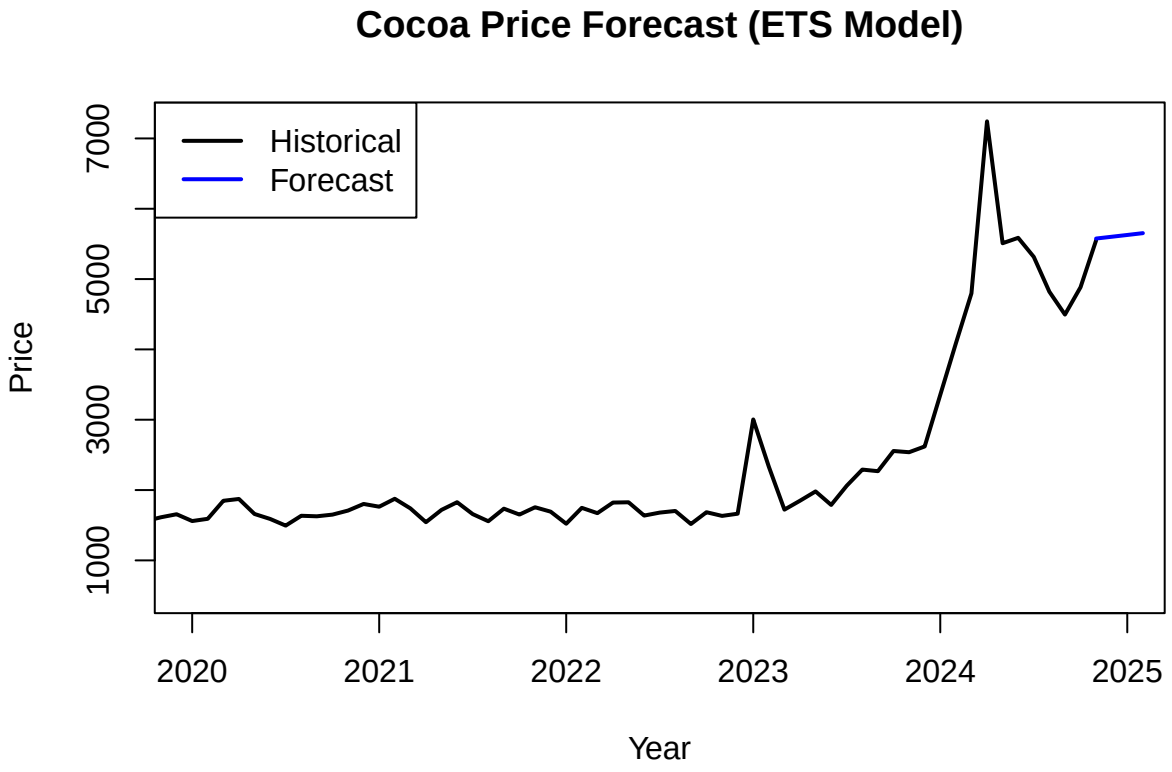



```
accuracy(price_forecast, actual_price_test)
```

```
##                  ME     RMSE      MAE       MPE     MAPE       ACF1 Theil's U
## Test set -676.9251 767.9354 676.9251 -14.33849 14.33849 0.04644613  1.619625
```

```
# Create a time series for the forecasted prices
forecast_start <- c(2024, 11) #data ends at nov
price_forecast_ts <- ts(price_forecast, start = forecast_start, frequency = 12)

# Plot the forecast
plot(price_ts, xlim = c(2020, 2025), ylim = range(c(price_ts, price_forecast_ts)),
     main = "Cocoa Price Forecast (ETS Model)", ylab = "Price", xlab = "Year", col = "black", lwd = 2)
lines(price_forecast_ts, col = "blue", lwd = 2)
legend("topleft", legend = c("Historical", "Forecast"), col = c("black", "blue"), lty = 1, lwd = 2)
```


Cocoa Price Forecast (ETS Model)
Historical
Forecast
Price
1000
3000
5000
7000
2020
2021
2022
2023
2024
2025
Year