Ken Marciel
Statistical Consulting (STAT 684), Spring 2022
Report of Final Project

Forecasting the prices of beef and chicken in the United States

Introduction

Beef and chicken are two of the most common conventional protein sources consumed by Americans. In recent years, the United States has experienced escalating beef prices along with growing popularity of chicken. These trends have been particularly acute since the beginning of the global COVID-19 pandemic in early 2020, and the overlapping period of high inflation that continues to the present. Therefore, forecasting these trends would be valuable for consumers, as well as the food and restaurant industries. Furthermore, since these price trends reflect consumption trends, they may also help inform public policy regarding implications for public health and the environment.

Methods

Economic Research data from the Federal Reserve Bank of St. Louis (FRED) was obtained for the historical prices of beef and chicken in the United States. The time series data set for beef currently consists of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022. This data set was downloaded as an Excel file from the following URL: https://fred.stlouisfed.org/series/APU0000703112. The time series data set for chicken currently consists of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022. This data set was downloaded as an Excel file from the following URL: https://fred.stlouisfed.org/series/APU0000706111.

For each data series, three types of forecasting models were trained: linear regression, exponential smoothing (Holt-Winters method), and ARIMA (autoregressive integrated moving average). The evaluation metric used to compare the models is mean absolute percentage error (MAPE).

$$MAPE = \frac{1}{n} \bullet \sum_{t=1}^{n} \left| \frac{Actual\ Price_t - Predicted\ Price_t}{Actual\ Price_t} \right| \cdot 100\%$$

- t, month in time series
- n, number of months in time series
- ∑, sum
- | |, absolute value

Data analysis was performed using the R statistical computing software. The following nine software packages for R were utilized for the analysis: dplyr, dygraphs, forecast, imputeTS, lubridate, plotly, readxl, stats, and TSstudio.

Results

<u>Price of beef</u>

Exploratory data analysis

This univariate time series data set contains the average monthly values for the price of 100% ground beef in the United States, measured in dollars per pound, from January 1, 1984, to March 1, 2022, for a total of 459 months. There was one missing value, for October 1, 2012, which was imputed using linear interpolation. This time series is plotted in Figure 1.

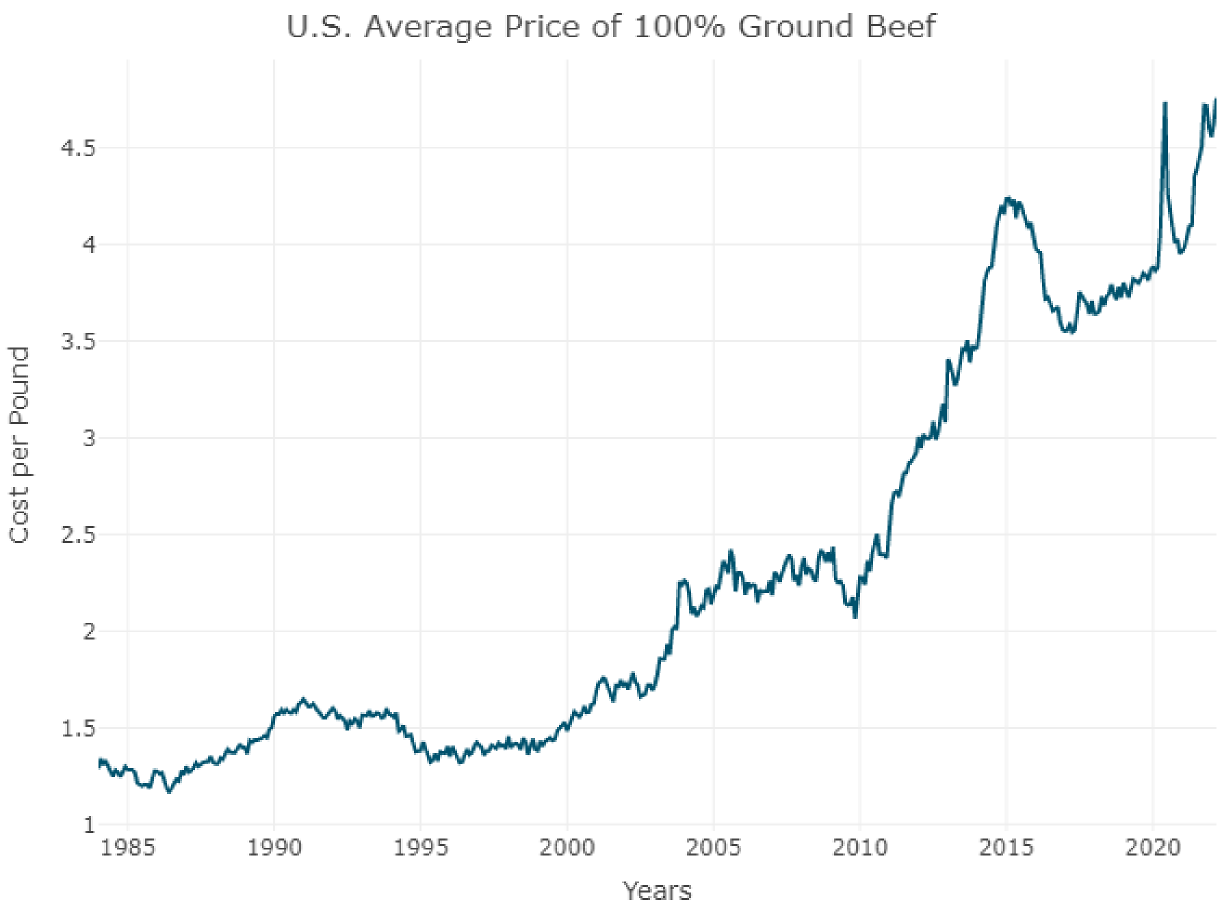

Figure 1. Time series plot of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

The time series has a growing trend with an embedded cycle, which are both apparent in the observed series. The most recent cycle started just before 2010, near the end of the [Great Recession](#) that began in 2008. There is no seasonal component apparent in the observed series. The time series plot can be decomposed to show the trend (including cycle), seasonal, and random components separately. These are plotted in Figure 2. The impact of the COVID-19 pandemic from 2020 to 2022 is conspicuous in both the observed series and the random component.

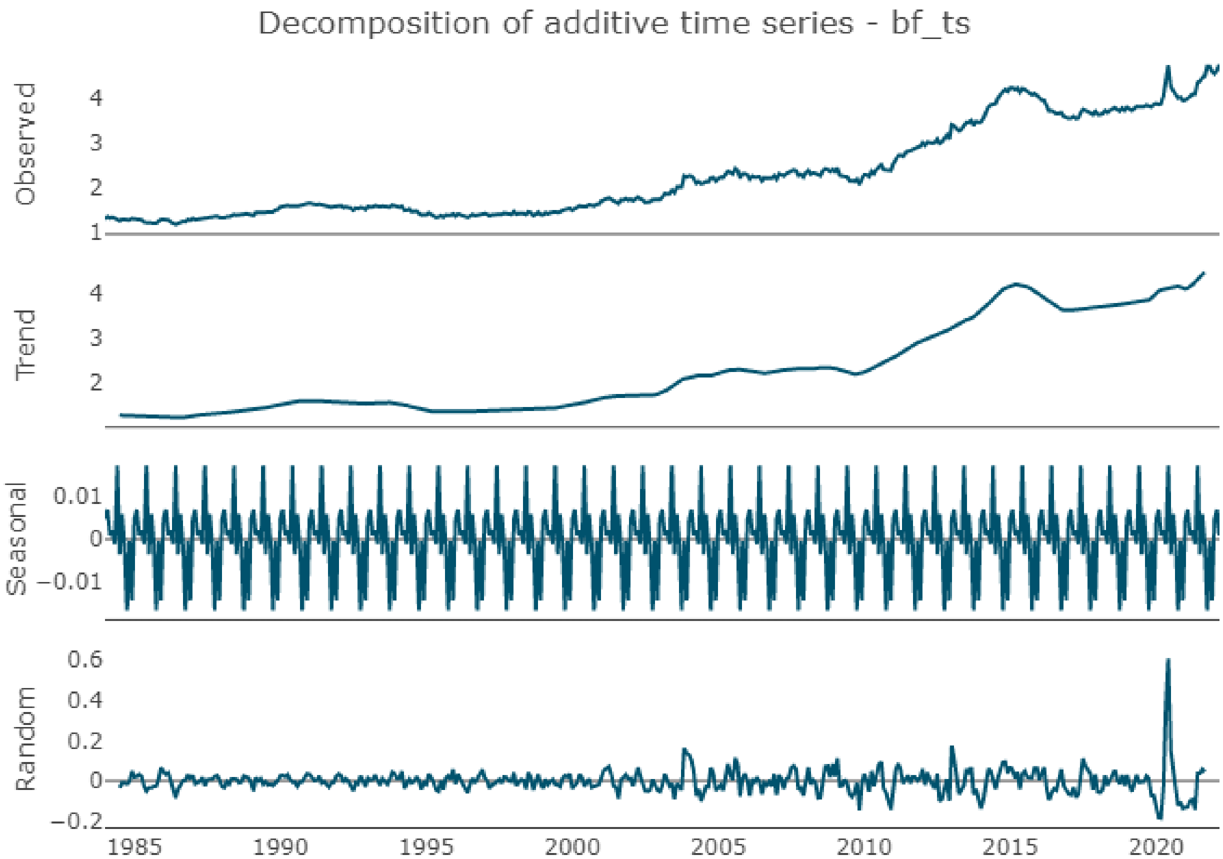

Figure 2. Classical decomposition of additive time series for the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

As plotted in Figure 3, a heatmap of the time series data shows evidence of cyclic behavior (across the vertical bars), but not seasonal behavior (along the horizontal bars). Four additional seasonality plots also reveal a lack of evidence for a seasonal pattern in the time series based on the following behavior: horizontal lines in the standard plot; rope appearance in the cycle plot; level pattern across the box plots; and circular spiral pattern in the polar plot. These plots are shown in Figure 4 and Figure 5. The correlation of the series with its lags is decaying gradually over time, with no apparent seasonal component, as shown in the correlation plots of Figure 6. The lack of seasonality in this data series makes sense given that beef is a daily food eaten year-round in the United States.

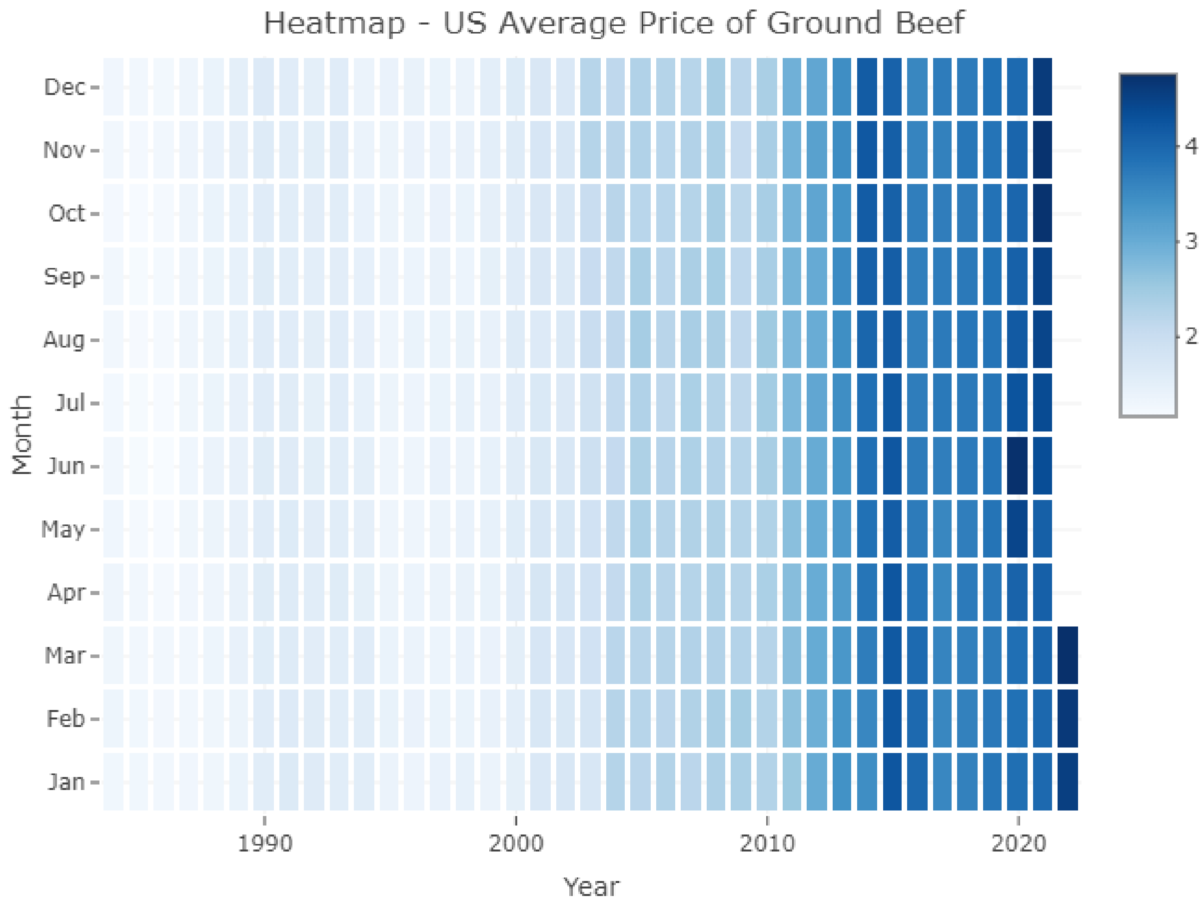

Figure 3. Heatmap of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

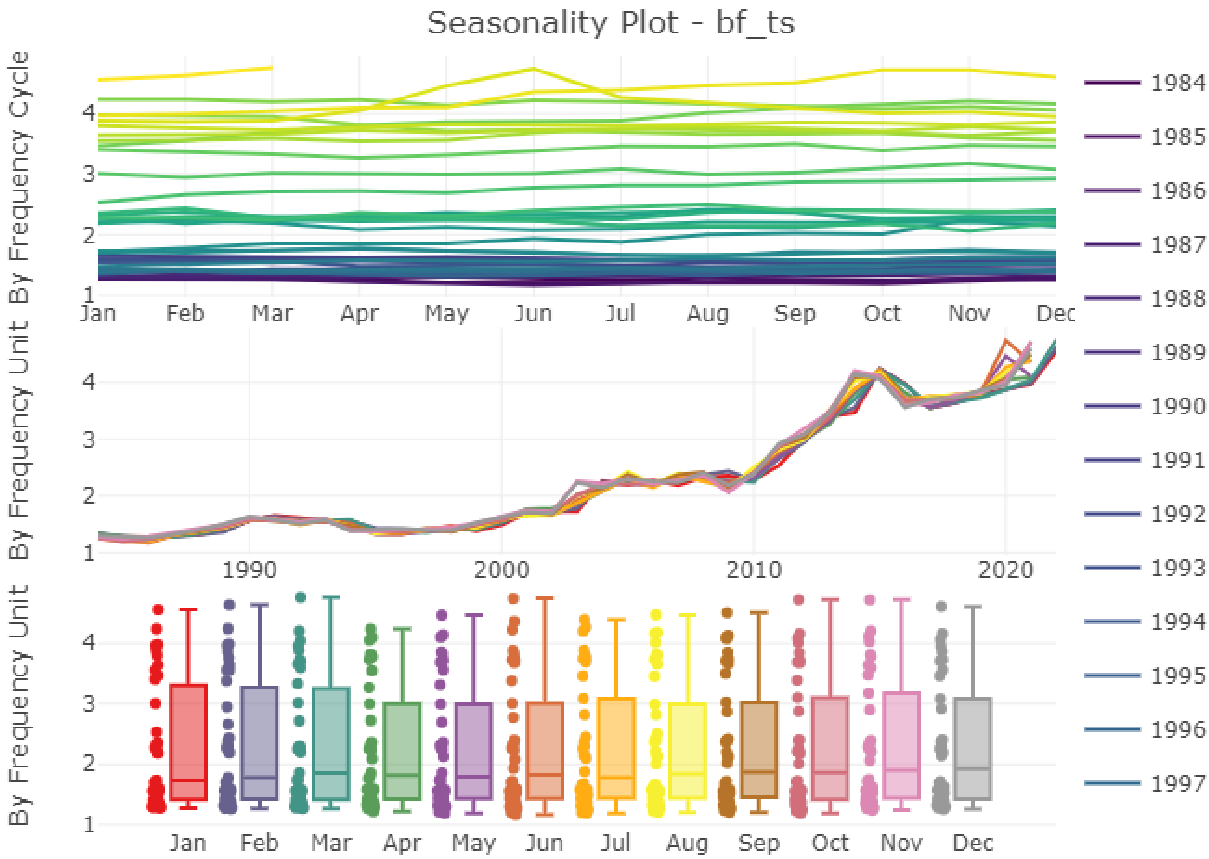

Figure 4. Figure 4. Normal, cycle, and box plots of seasonality of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

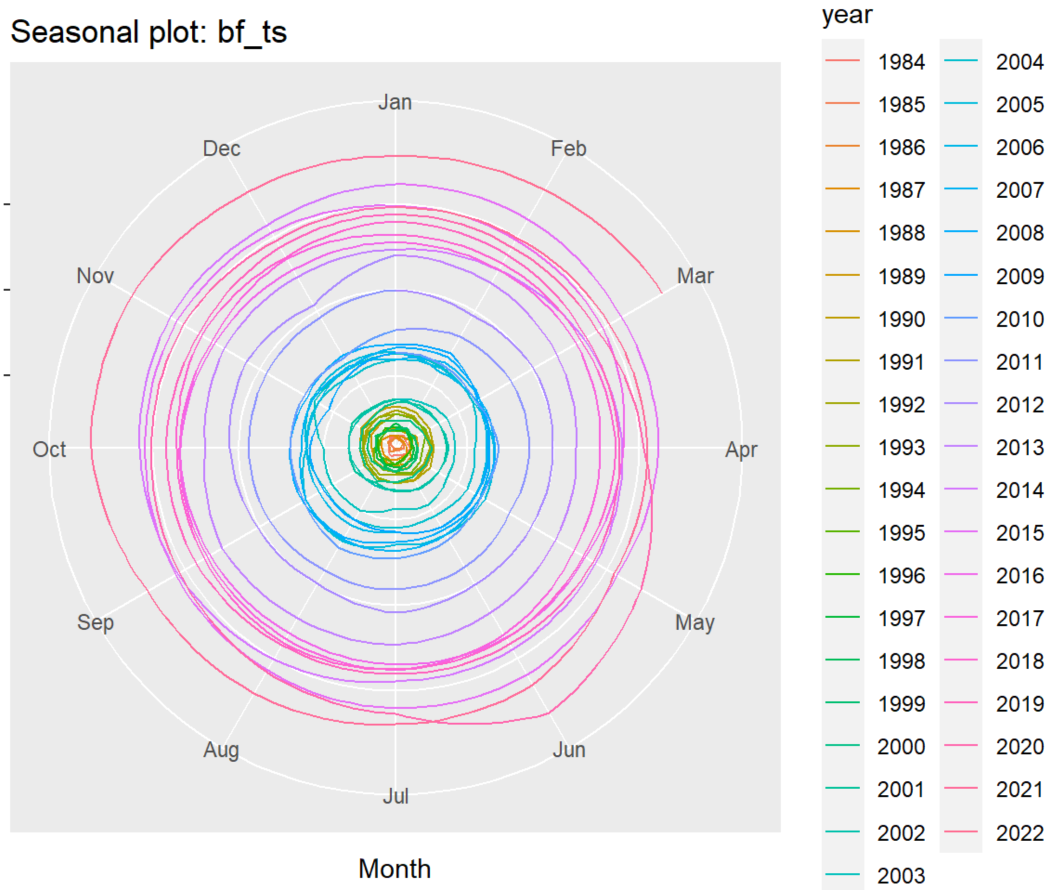

Figure 5. Polar plot of seasonality of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

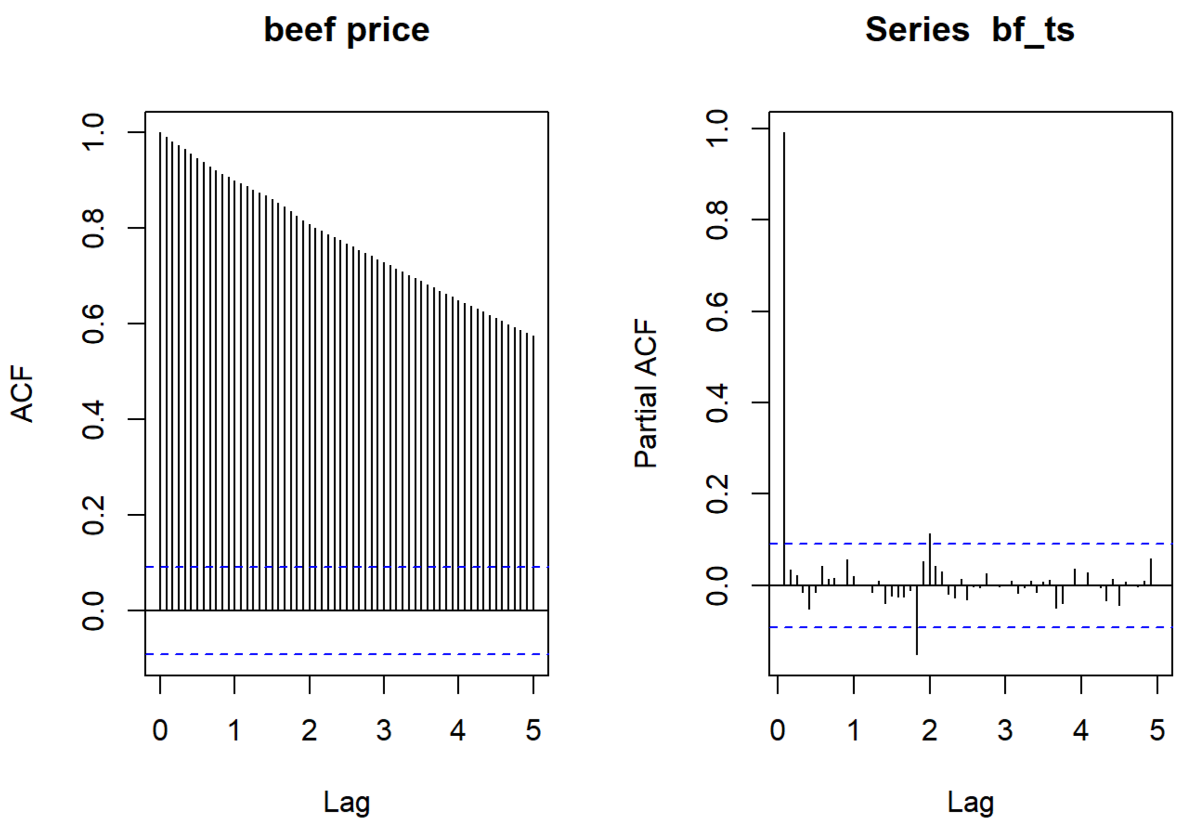

Figure 6. Plot of the autocorrelation function (ACF) for the first 60 lags, and plot of the partial autocorrelation function (PACF) for the first 60 lags, of the time series of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

Forecasting with linear regression

The time series data set was partitioned into a training set consisting of the values of the first 447 months, and a test set consisting of the last 12 months. The following equation was used to train a linear regression model:

$$Price_t = \beta_0 + \beta_1 \bullet Trend_t + \beta_2 \bullet Trend_t^2 + \beta_3 \bullet Season_t + Irregular_t$$

where

- $Price_t$ is the price at month *t* out of *n* total observations
- $Trend_t$ is the trend and cycle component for linear growth of the series over time
- $Trend_t^2$ is the squared value of the trend and cycle component for nonlinear growth of the series over time
- $Season_t$ is the seasonal component for month of the year

- $Irregular_t$ is the irregular component for any other patterns not captured by the trend, cycle, and seasonal components
- β_0, β_1, β_2, and β_3 are the model intercept and coefficients of the trend, squared trend, and seasonal components, respectively

Ordinary least squares (OLS) was the technique used to estimate the coefficients for the linear regression model. The model coefficients for the intercept and the trend components are statistically significant ($p < 0.001$). However, none of the coefficients for the seasonal components are statistically significant, as suspected from the exploratory data analysis.

The Ljung-Box test has a statistically significant p-value < 0.001, rejecting the null hypothesis that the irregular component is white noise. This is confirmed by visual analysis of the residuals, which show significant correlation in the model between the series and its lags (as seen in Figure 7). This means that the model does not capture most of the variation patterns of the series. Therefore, it is not a valid model for consideration. However, I used its MAPE score of 3.07% on the test set as a benchmark to evaluate the performance of the other models that I trained.

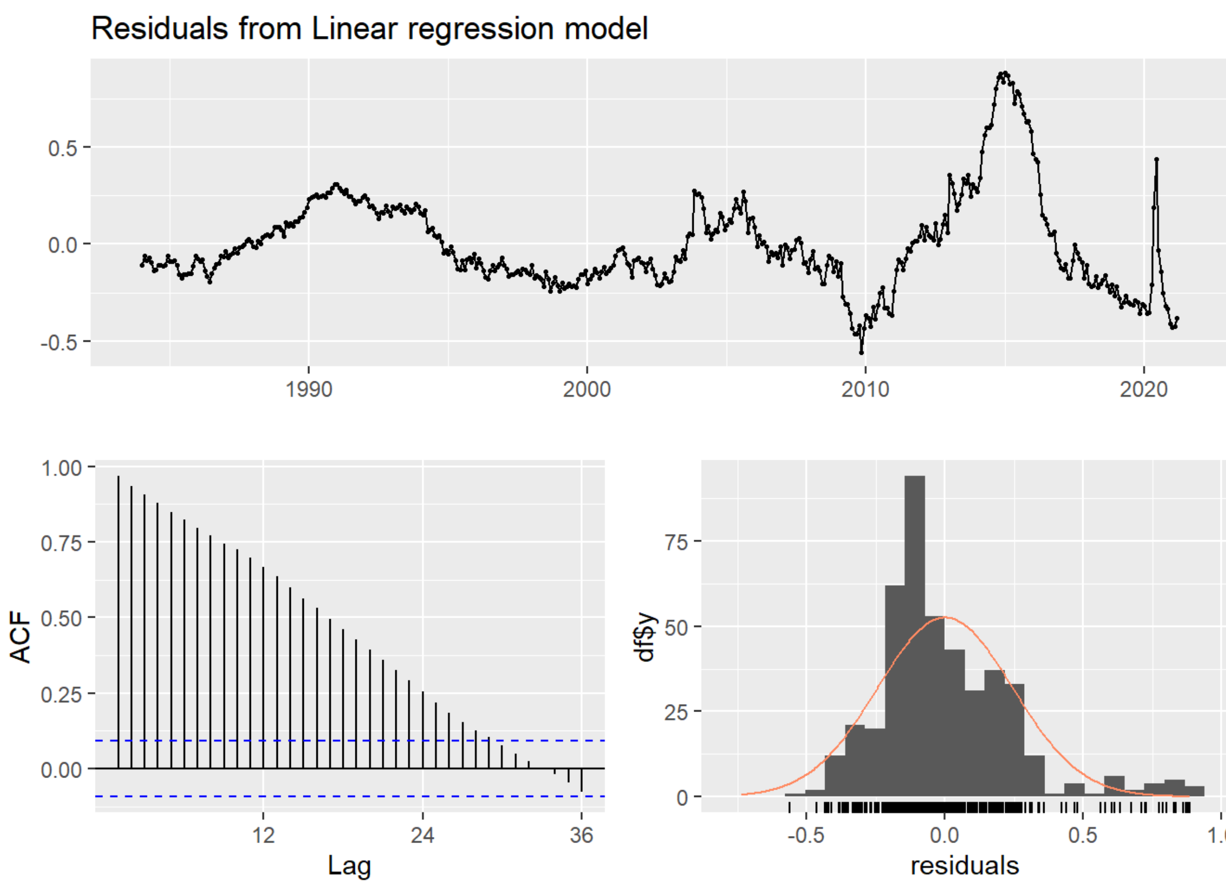

Figure 7. Residual plots of the trained linear regression model for the time series of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

Forecasting with exponential smoothing

The time series data set was partitioned into a training set consisting of the values of the first 447 months, and a test set consisting of the last 12 months. The Holt-Winters method was used to train an exponential smoothing model. Although correlation analysis indicated that the model is valid, it has an MAPE score of 9.32% on the testing set, which is worse than my benchmark value of 3.07%. Therefore, I did not choose this to be my forecasting model.

Forecasting with ARIMA models

To transform the time series to a stationary state and stabilize its variation, a log transformation was applied to the series, followed by first-order differencing. The result is plotted in Figure 8. Correlation plots of the transformed series are shown in Figure 9. The series cuts off after the first lag; the lags do not appear to tail off in either plot; and there is no apparent seasonality pattern.

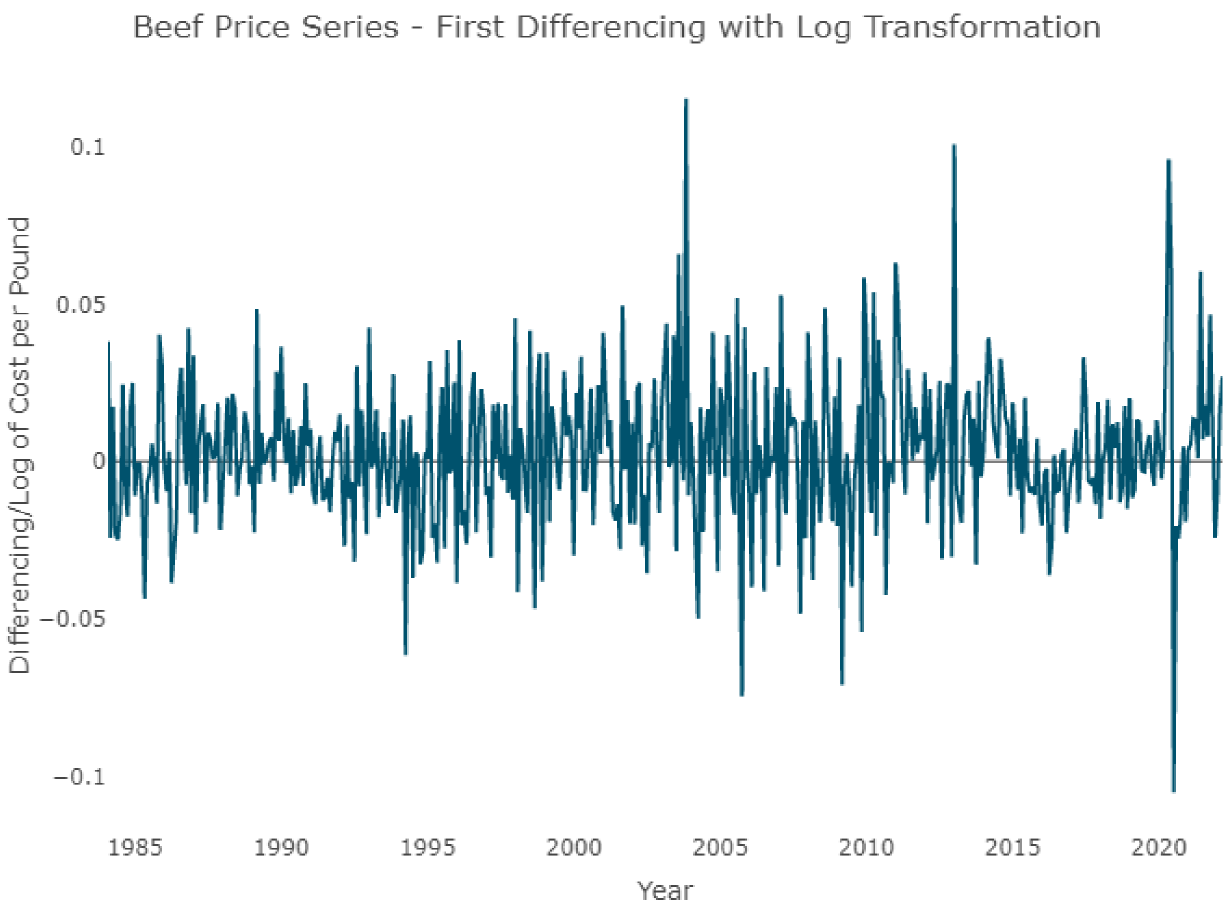

Figure 8. Plot of time series values after log transformation followed by first-order differencing for the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

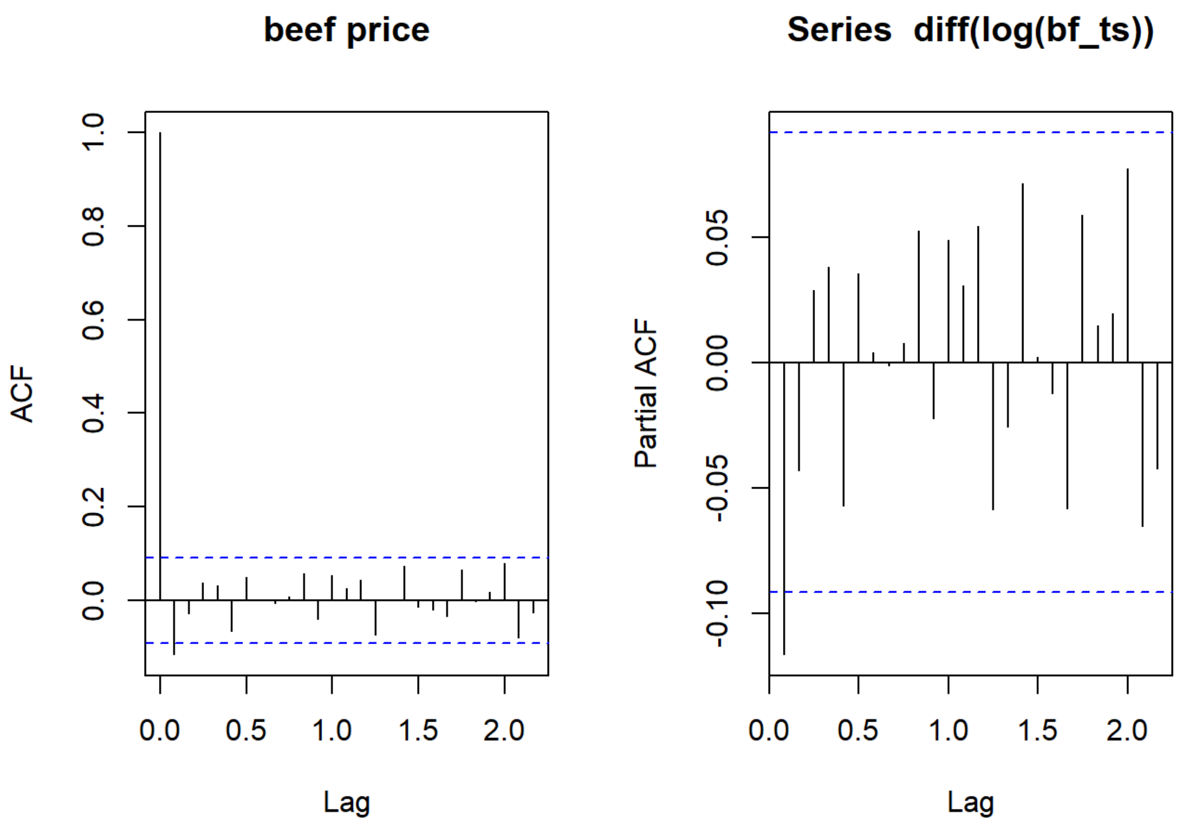

Figure 9. Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots with log transformation and first-order differencing for the time series of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

I fit the correlation pattern of the transformed series with an ARIMA(1,1,1) model. Although correlation analysis indicated that the random component of the model is white noise, it has an MAPE score of 3.31%, which is not an improvement over the benchmark value of 3.07%. Therefore, I did not choose this to be our forecasting model.

I next used an automated tuning process that obtained an ARIMA(0,1,0) model with a drift term. The model has an MAPE score of 1.78% on the test set, which is an improvement over the benchmark value of 3.07%. Furthermore, residual analysis indicates that the errors are uncorrelated with an approximately normal distribution, as shown in Figure 10. Therefore, I chose this as my forecasting model for the beef price time series.

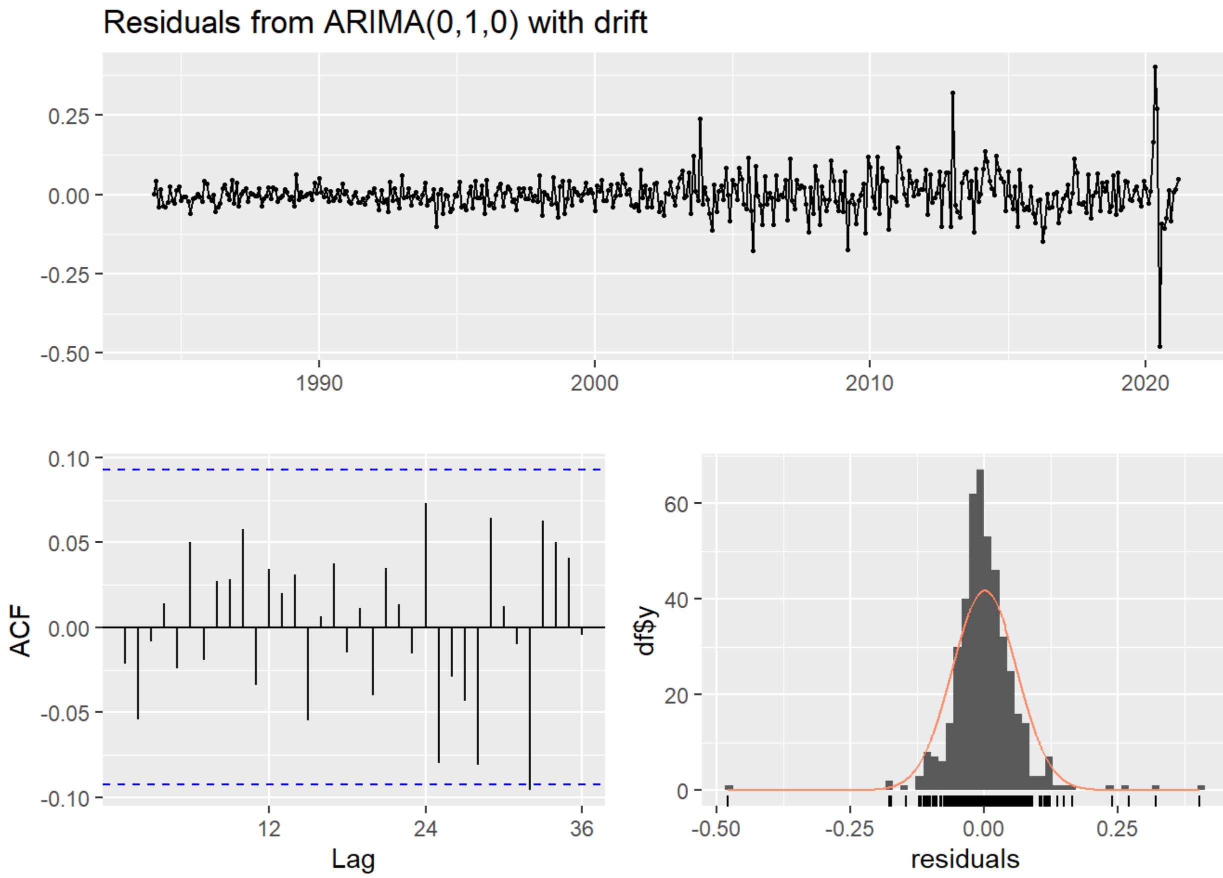

Figure 10. Residual plots of model ARIMA(1,1,1) with drift for the time series of the U.S. average monthly price of 100% ground beef from January 1, 1984, to March 1, 2022

Predicted value for March 1, 2023

Using the ARIMA(0,1,0) with drift term as my forecasting model, the U.S. average cost per pound of 100% ground beef is predicted to be $4.19 on March 1, 2023. This falls within a prediction interval ranging from $3.61 to $4.77, with a 95% level of confidence.

Price of chicken

Exploratory data analysis

This univariate time series data set contains the average monthly values for the price of fresh, whole chicken in the United States, measured in dollars per pound, from January 1st, 1980, to March 1st, 2022, for a total of 507 months. There was one missing value, for May 1st, 2020, which was imputed using linear interpolation. This time series is plotted in Figure 11.

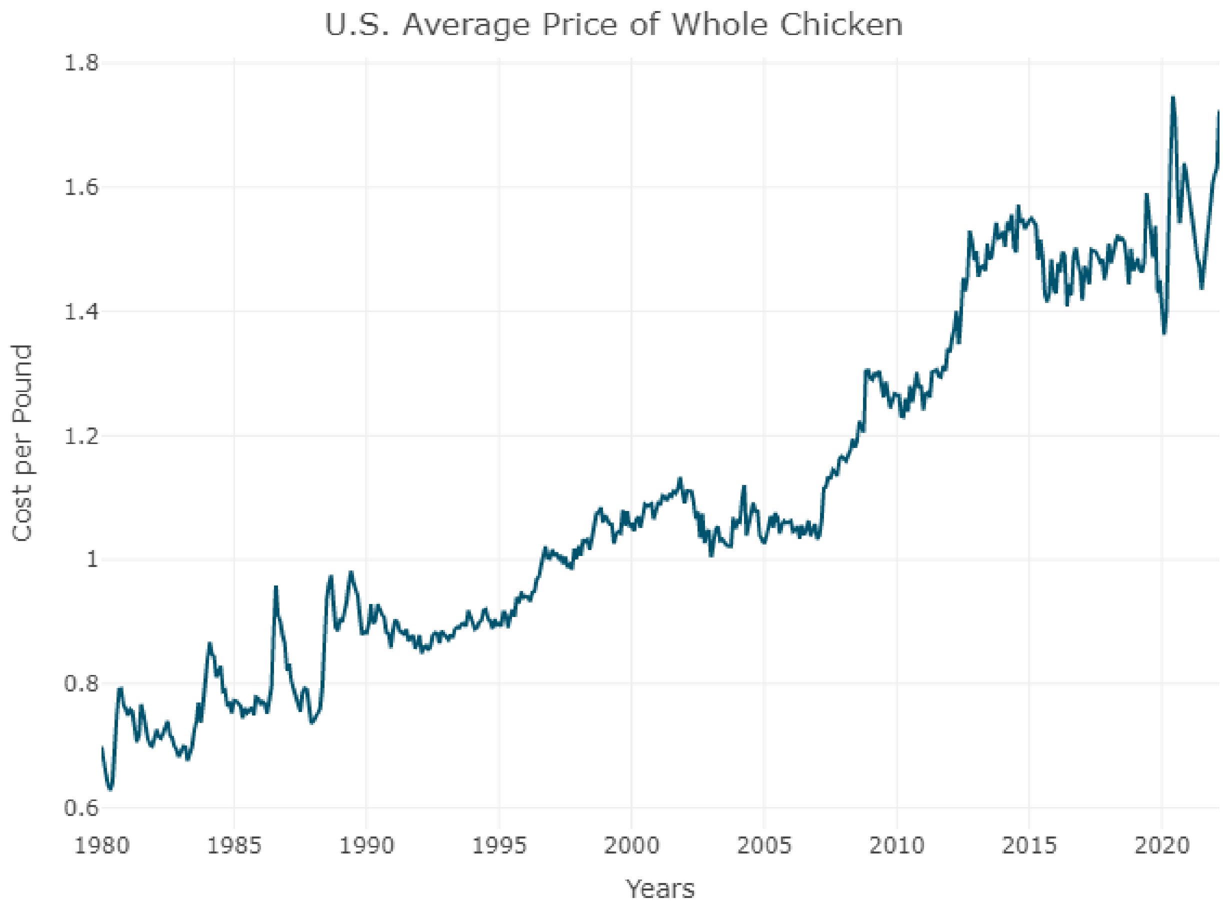

Figure 11. Time series plot of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

The time series has a growing trend with an embedded cycle, which are both apparent in the observed series. The most recent cycle started just before 2010, near the end of the Great Recession that began in 2008. There is no seasonal component apparent in the observed series. The time series plot can be decomposed to show the trend (including cycle), seasonal, and random components separately. These are plotted in Figure 12. The impact of the COVID-19 pandemic from 2020 to 2022 is conspicuous in both the observed series and the random component.

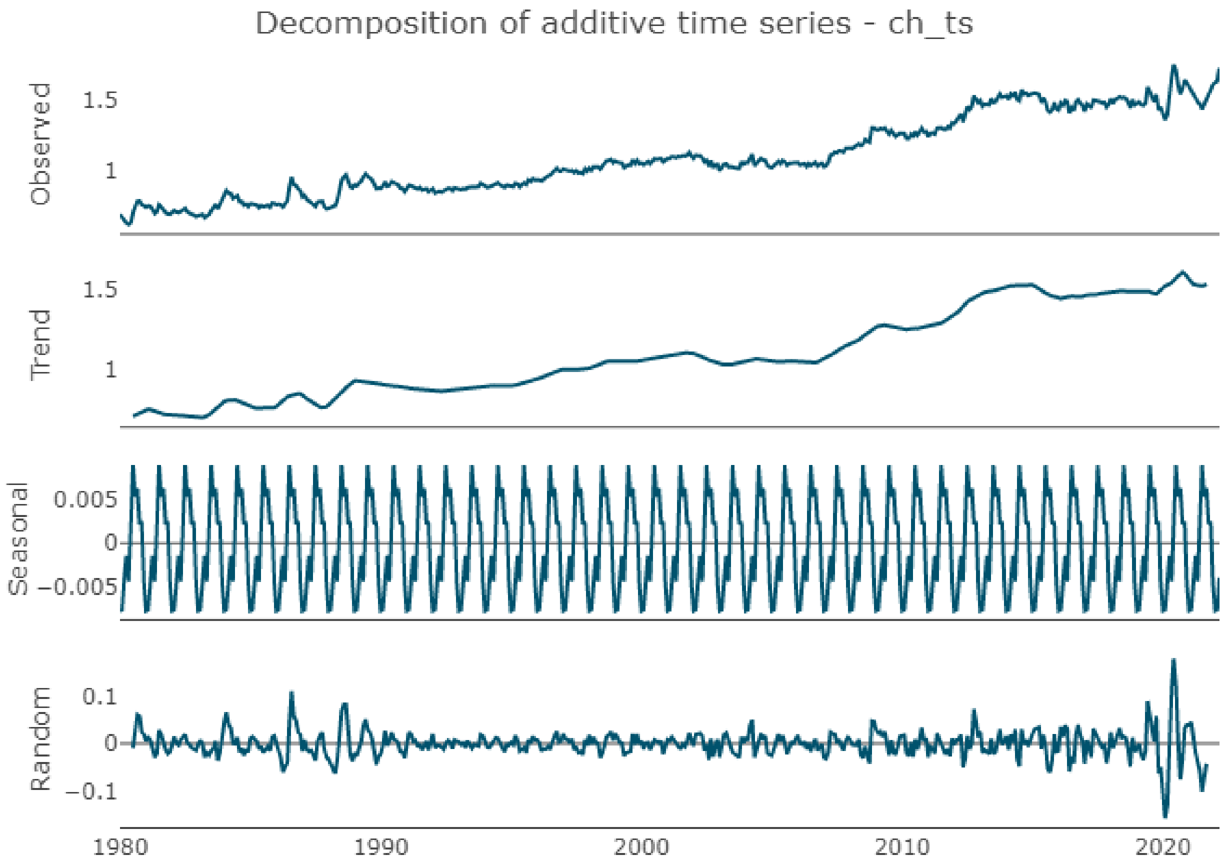

Figure 12. Classical decomposition of additive time series for the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

As plotted in Figure 13, a heatmap of the time series data shows evidence of cyclic behavior (across the vertical bars), but not seasonal behavior (along the horizontal bars). Four additional seasonality plots also reveal a lack of evidence for a seasonal pattern in the time series based on the following behavior: horizontal lines in the standard plot; rope appearance in the cycle plot; level pattern across the box plots; and circular spiral pattern in the polar plot. These plots are shown in Figure 14 and Figure 15. The correlation of the series with its lags is decaying gradually over time, with no apparent seasonal component, as shown in the correlation plots of Figure 16. The lack of seasonality in this data series makes sense given that chicken is a food eaten year-round in the United States.

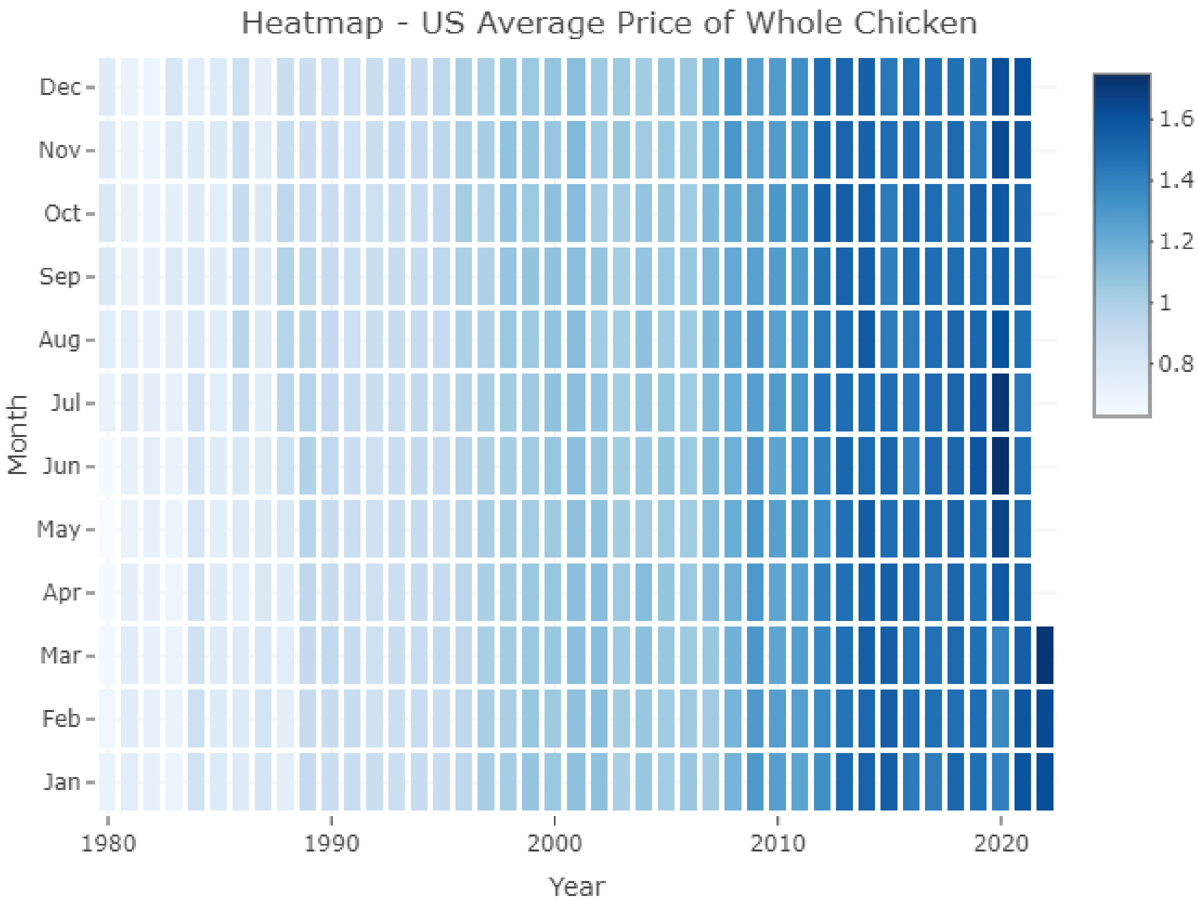

Figure 13. Heatmap of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

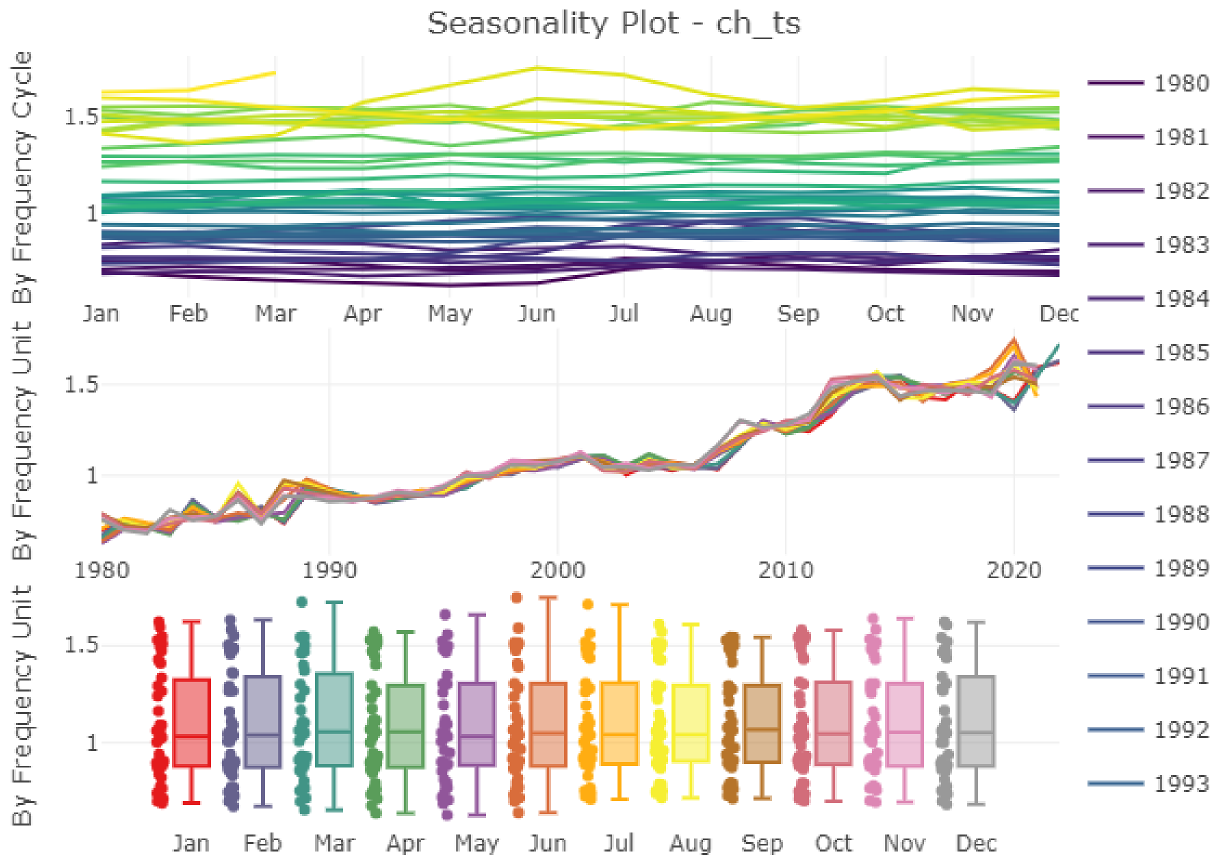

Figure 14. Normal, cycle, and box plots of seasonality of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

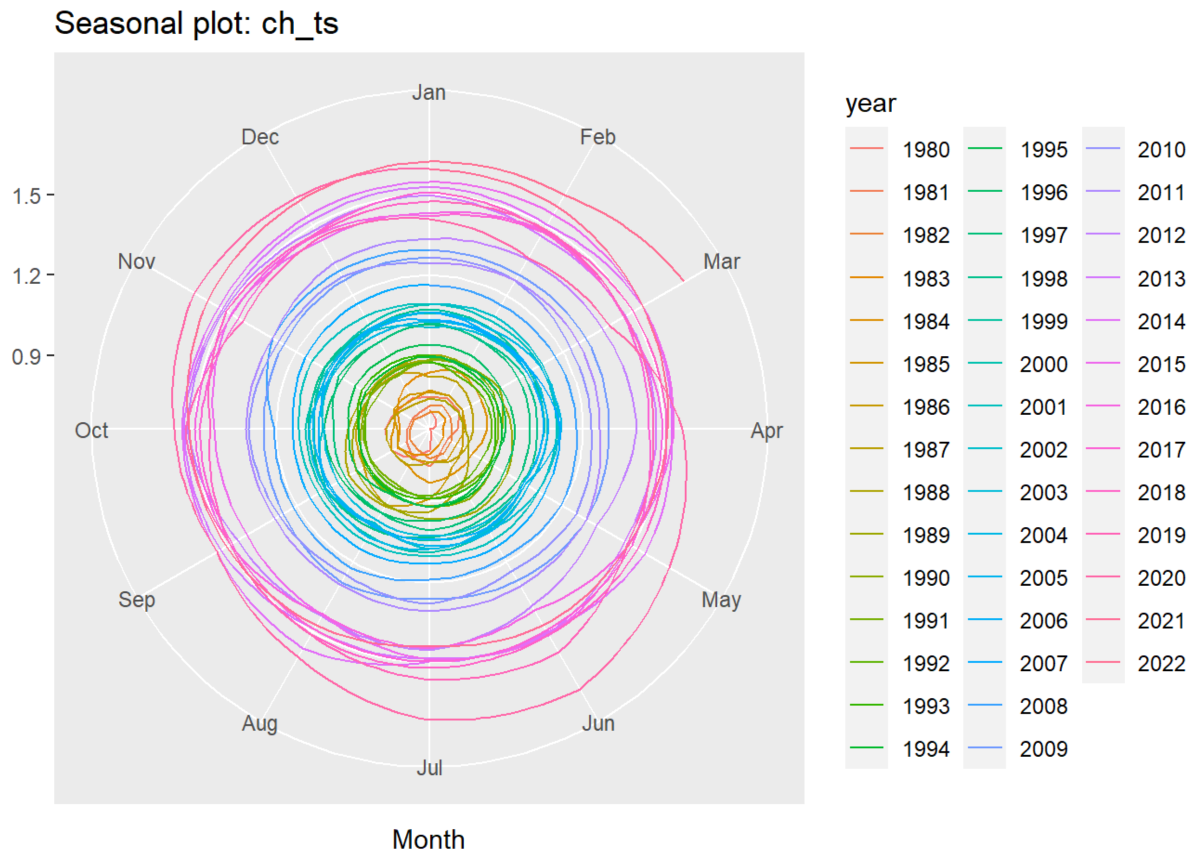

Figure 15. Polar plot of seasonality of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

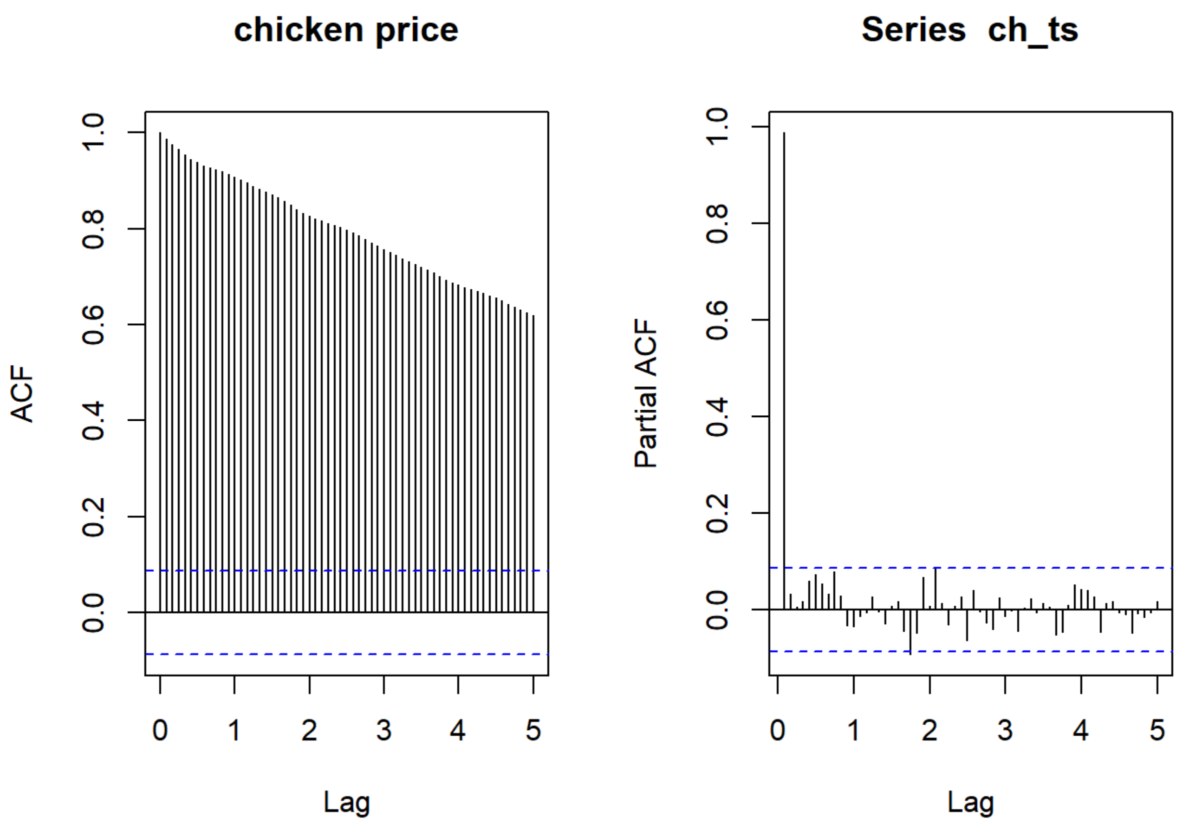

Figure 16. Plot of the autocorrelation function (ACF) for the first 60 lags, and plot of the partial autocorrelation function (PACF) for the first 60 lags, of the time series of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March

Forecasting with linear regression

The time series data set was partitioned into a training set consisting of the values of the first 495 months, and a test set consisting of the last 12 months. The following equation was used to train a linear regression model:

$$Price_t = \beta_0 + \beta_1 \bullet Trend_t + \beta_2 \bullet Trend_t^2 + \beta_3 \bullet Season_t + Irregular_t$$

where

- $Price_t$ is the price at month *t* out of *n* total observations
- $Trend_t$ is the trend and cycle component for linear growth of the series over time
- $Trend_t^2$ is the squared value of the trend and cycle component for nonlinear growth of the series over time
- $Season_t$ is the seasonal component for month of the year

- $Irregular_t$ is the irregular component for any other patterns not captured by the trend, cycle, and seasonal components
- β_0, β_1, β_2, and β_3 are the model intercept and coefficients of the trend, squared trend, and seasonal components, respectively

Ordinary least squares (OLS) was the technique used to estimate the coefficients for the linear regression model. The model coefficients for the intercept and the trend components are statistically significant ($p < 0.001$). However, none of the coefficients for the seasonal components are statistically significant, as suspected from the exploratory data analysis.

The Ljung-Box test has a statistically significant *p*-value < 0.001, rejecting the null hypothesis that the irregular component is white noise. This is confirmed by visual analysis of the residuals, which show significant correlation in the model between the series and its lags (as seen in Figure 17). This means that the model does not capture most of the variation patterns of the series. Therefore, it is not a valid model for consideration. However, I used its MAPE score of 6.51% as a benchmark to evaluate the performance of the other models that I trained.

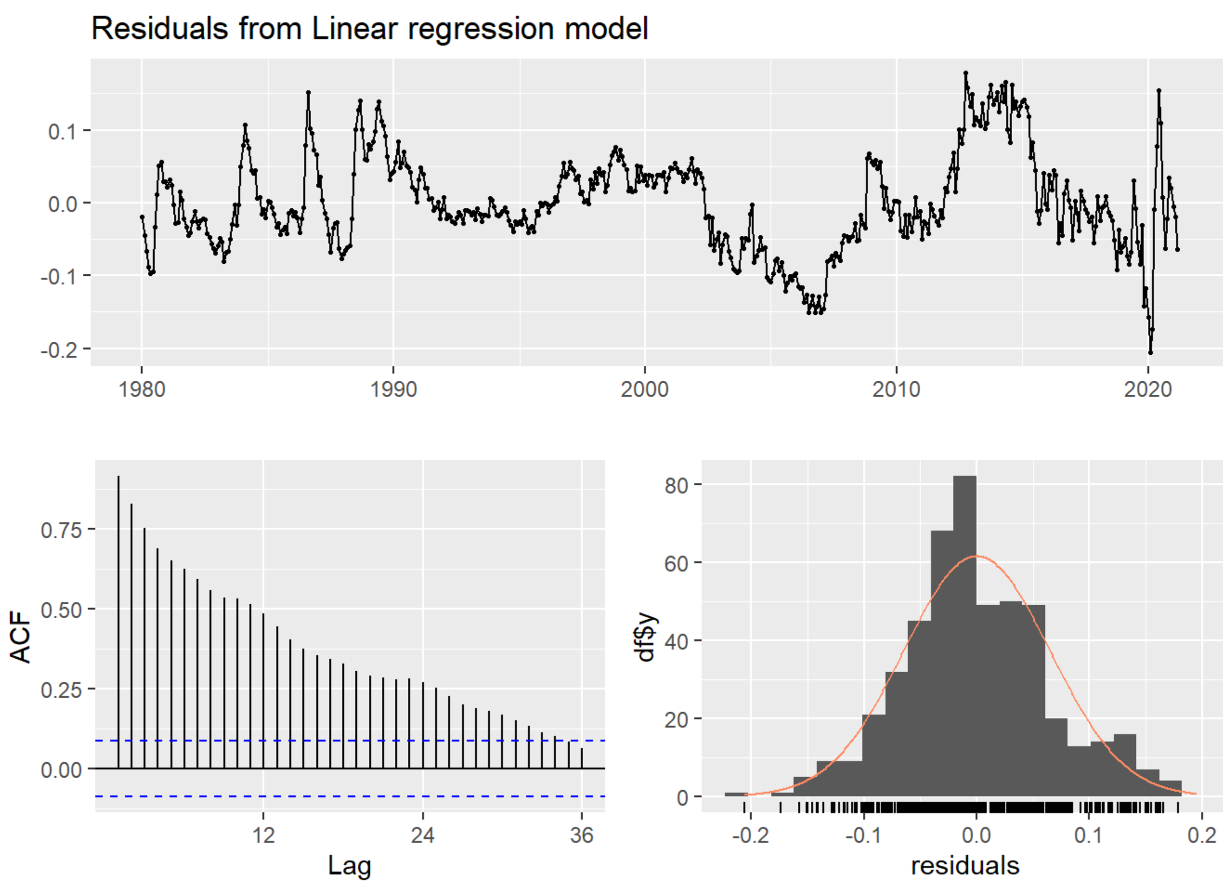

Figure 17. Residual plots of the trained linear regression model for the time series of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

Forecasting with exponential smoothing

The time series data set was partitioned into a training set consisting of the values of the first 447 months, and a test set consisting of the last 12 months. The Holt-Winters method was used to train an exponential smoothing model. The MAPE score is 4.62% in the testing set, which is lower than my benchmark value of 6.51%. However, residual analysis shows significant autocorrelation in the series with its lags, so I conclude that this is not a valid forecasting model. Therefore, I did not choose this to be my forecasting model.

Forecasting with ARIMA models

To transform the time series to a stationary state and stabilize its variation, a log transformation was applied to the series, followed by first-order differencing. The result is plotted in Figure 18. Correlation plots of the transformed series are shown in Figure 19. The transformed series appears to tail off in both correlation plots, and there is no apparent seasonality pattern.

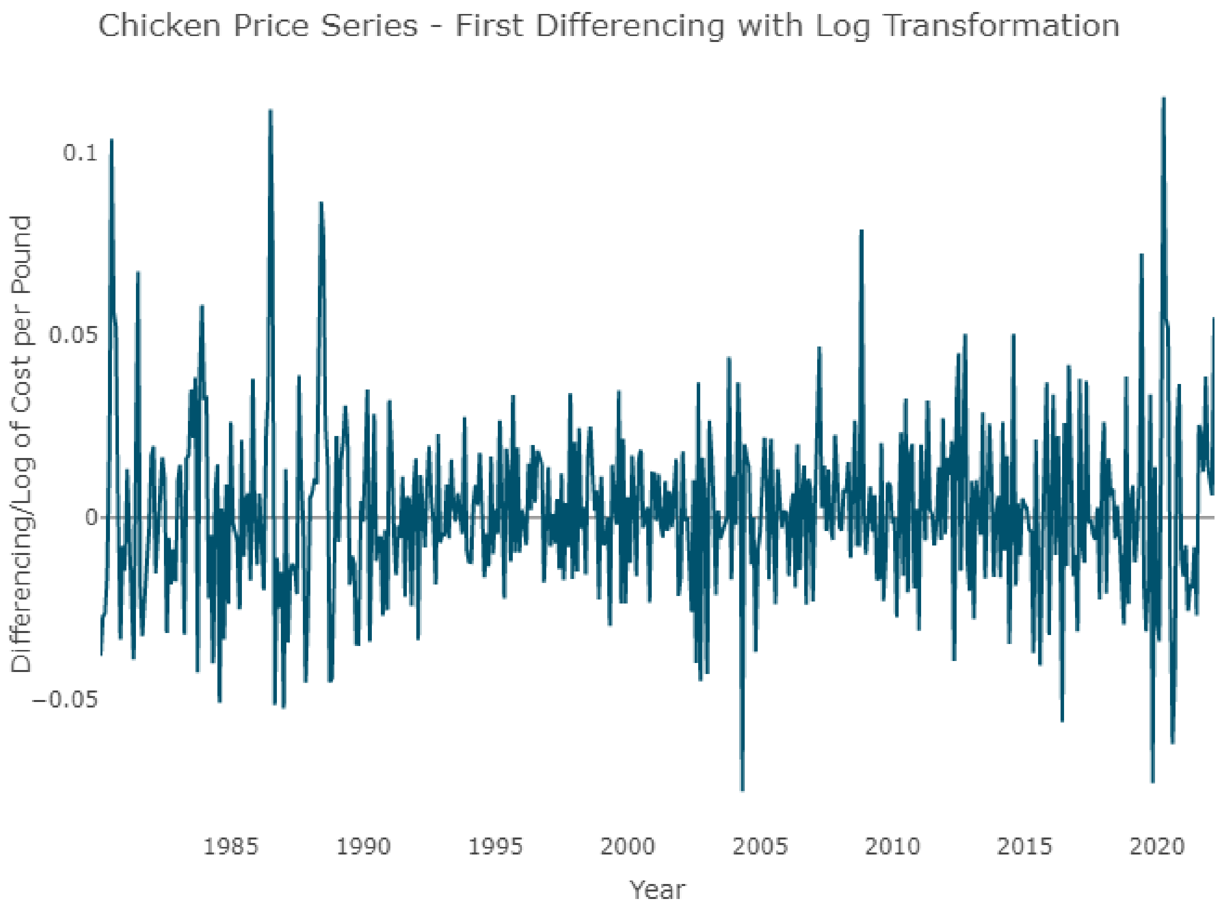

Figure 18. Plot of time series values after log transformation followed by first-order differencing for the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

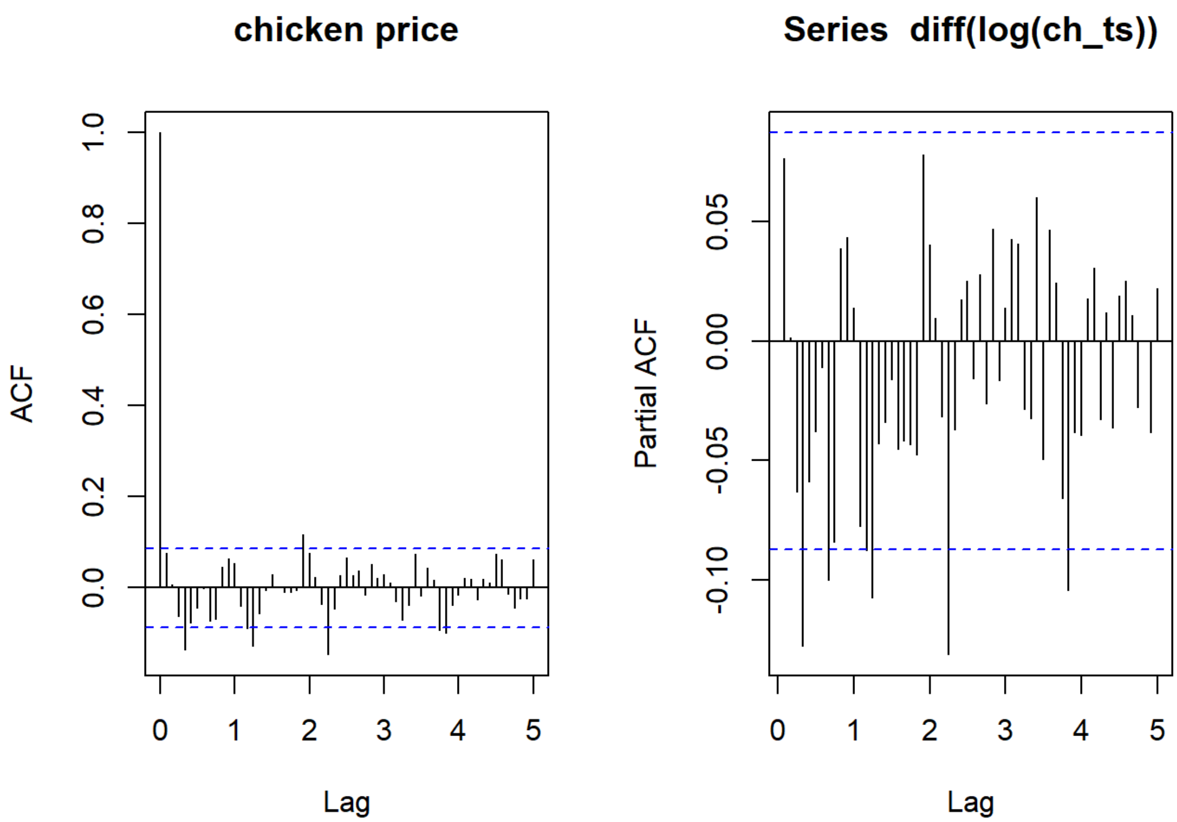

Figure 19. Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots with log transformation and first-order differencing for the time series of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

I fit the correlation pattern of the transformed series with an ARIMA(0,1,0) model, resulting in an MAPE score of 25.2%. Furthermore, residual analysis show that the series has significant autocorrelation with its lags. Therefore, I did not choose this to be my forecasting model.

I next used an automated tuning process that obtained an ARIMA(0,1,5) model with a drift term. The model has an MAPE score of 1.73% on the test set, which is an improvement over the benchmark value of 6.51%. Furthermore, residual analysis indicates that the errors are uncorrelated with an approximately normal distribution, as shown in Figure 20. Therefore, I chose this as my forecasting model for the chicken price time series.

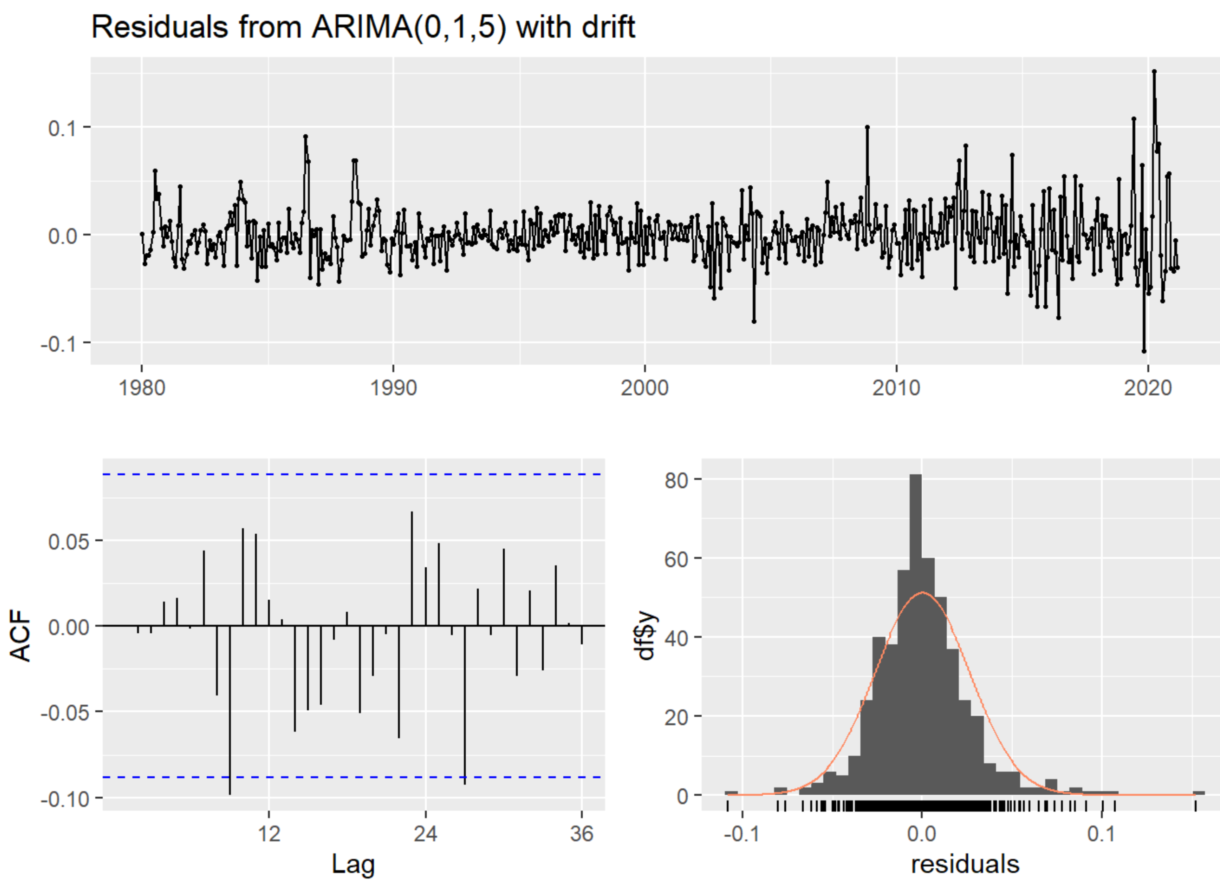

Figure 20. Residual plots of model ARIMA(0,1,5) with drift for the time series of the U.S. average monthly price of fresh, whole chicken from January 1, 1980, to March 1, 2022

Predicted value for March 1, 2023

Using the ARIMA(0,1,5) with drift term as my forecasting model, the U.S. average cost per pound of fresh, whole chicken is predicted to be $1.62 on March 1, 2023. This falls within a prediction interval ranging from $1.46 to $1.78, with a 95% level of confidence.

Price difference between beef and chicken

The time series data sets that I used to investigate beef and chicken prices consist of values that are measured in the same unit of dollars per pound. The observed values of these data sets also occur at the same frequency, which is the first day of each month. Both series have a common endpoint which is currently March 1, 2022. The earliest common starting point is January 1, 1984, which is the beginning of the series for beef prices. Hence, both series can be plotted together, on the same scale, from January 1, 1984, to March 1, 2022, as shown in Figure 21.

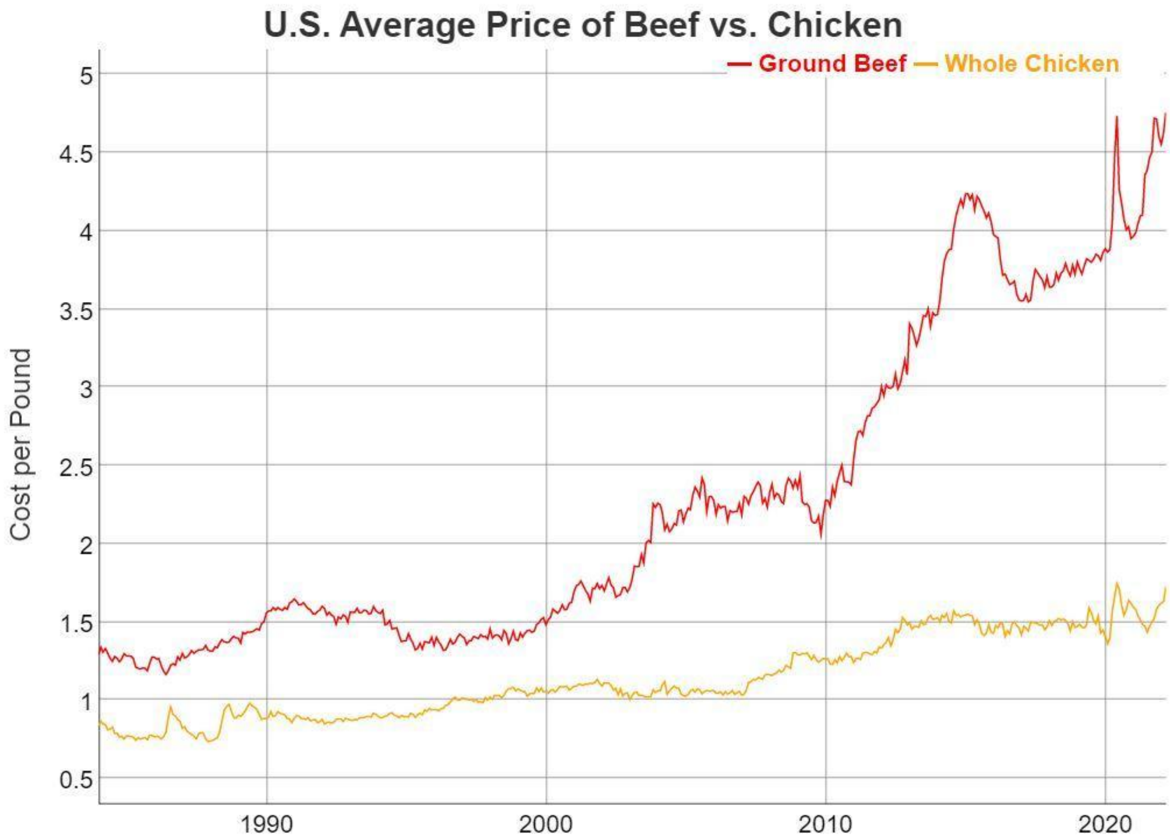

Figure 21. Multivariate time series plot for the U.S. average prices of 100% ground beef and fresh, whole chicken, from January 1, 1984, to March 1, 2022

From January 1, 1984, to January 1, 2000, the price series for beef and chicken each show a mostly horizontal trend, moving in a slightly upward direction. Furthermore, the series are mostly parallel to each other, which suggests a roughly constant difference in price between beef and chicken during this period. However, after January 1, 2000, the price trend for beef shifts to a conspicuously steeper incline compared to chicken. This growing price disparity has continued to the current endpoint of the two series on March 1, 2022.

The price difference between beef and chicken was $3.04 on March 1, 2022. Using the point estimates predicted by the forecasting models that I selected for the two series, the price difference is projected to be $2.57, one year later, on March 1, 2023. Accordingly, this price difference falls within a prediction interval ranging from $2.14 to $2.99, with a 95% level of confidence.

Conclusion

On March 1, 2022, in the United States, the average price for 100% ground beef was $4.76 per pound, and the average price for fresh, whole chicken was $1.72 per pound, for a price difference of $3.04 per pound. I was interested in knowing how those values might change one year later, on March 1, 2023, in the United States. Using the forecasting models that I selected from my investigation, I make the following projections for March 1, 2023, with a 95% level of confidence: the average price of 100% ground beef will be between $3.61 and $4.77 per pound; the average price of fresh, whole chicken will be between $1.46 and $1.78 per pound; and the average price difference between 100% ground beef and fresh, whole chicken will be between $2.14 and $2.99 per pound.

Discussion

Despite public concern over rising meat prices during the current period of high inflation in the United States, my forecasting models suggest, with 95% confidence, that the price per pound for 100% ground beef will not increase by more than $0.01 per pound, and that the price per pound for fresh, whole chicken will not increase by more than $0.06 per pound. Furthermore, despite the growing disparity in prices between beef and chicken over the past two decades in the United States, my forecasting models suggest that this price difference will not be higher on March 1, 2023, than it was on March 1, 2022. Rather, I project with 95% confidence that the price difference will decrease by at least $0.05 per pound.

The findings of my investigation suggest that, in 12 months from March 1, 2022, on March 1, 2023, there should be no major increase in the average prices of beef and chicken in the United States. This may provide welcome relief for consumers when grocery shopping, especially given the current inflationary period. Similarly, consumers may continue to enjoy their favorite dishes with beef and chicken at restaurants without having to dig deeper into their pockets. In turn, restaurants may be able to maintain their current volume of business, as well as work hours for their employees. Likewise, producers of beef and chicken may not have to raise their prices beyond the normal adjustment for inflation. Consequently, this may be a positive indicator of the health of the economy on March 1, 2023, and perhaps a lower risk of entering the next recession by then.

The one-year forecast of my investigation suggests price stability for beef and chicken. In turn, this may reflect stability in the current level of consumption of these two products. Such maintenance of the status quo would not reduce the impact of U.S. meat consumption on climate change, given that livestock produce more emissions than all forms of transportation combined. Likewise, environmental impacts would not lessen in terms of the amount of land, grain and water consumed—and waste produced—by large-scale farming of cattle and chickens. Similarly, antibiotic resistance and zoonotic pandemic risk resulting from industrial animal agriculture would not lessen.

Given the year-over-year price stability of U.S. meat suggested by my forecasting models, future research could look at a longer forecast horizon, such as five years into the future. In addition to the classical time series forecasting methods used for this investigation, future research could include the use of machine learning models (such as Random Forest or Gradient Boosting Machine) to obtain low error rates for longer forecast horizons.

References

Krispin, R. (2019), *Hands-On Time Series Analysis with R*, Birmingham, UK: Packt Publishing Ltd.

Shumway, R. H., and Stoffer, D. S. (2017), *Time Series Analysis and Its Applications* (4th ed.), Cham, CH: Springer International Publishing AG.