**DS480G**

Michelle Golz

**ARIMA, SARIMA, ARCH, and GARCH Final Project**

Data Description

When searching for data, one of my primary interests was in how time series models can provide forecasting for price data. The Bureau of Labor Statistics (BLS) compiles datasets about certain consumer goods. Initially, I selected the price of grade A large eggs by the dozen as my price forecast. During my data exploration and initial modelling, I found that the recent price trend on eggs made the data set very non-stationary and was difficult to model well. So, I chose milk prices instead and included my data for this report in the attached milk\_price.xlsx file.

Consumer goods like milk don’t seem to have a high degree of seasonal price variation, so I wanted to choose another dataset for my seasonal model. Weather is very seasonal, so I selected average temperature in the United States for my SARIMA model. I gathered the data through the National Center for Environmental Information’s website using the Climate at a Glance tool. That data is attached in the file temperature\_US.csv.

Autoregressive Integrated Moving Average Model (ARIMA)

ARIMA models integrate the functionality of autoregressive and moving average models together. Some of their benefits include the flexibility of handling a wide variety of time series data and are relatively simple to interpret. ARIMA cannot handle seasonal data without some modification, but that modification is relatively simple to implement [1]. The drawbacks of the prediction model are that it can be computationally intensive, is not suitable for long-term predictions, and can be poor at predicting turning points, something that I experienced when trying to model egg prices [1].

For my ARIMA model, I used US Milk Prices, which were collected as described above. The first figure is of egg prices over time, and based on the graph it appears that variance is not constant over the time series. As such, I used log transformation on my data to make it more useful for my ARIMA model.

A graph showing a graph of a number of years

AI-generated content may be incorrect.

Figure : Milk Price Over Time

My initial plots including autocorrelation function (ACF) and partial autocorrelation function (PACF) values already suggested that my data was non-stationary (Figure 2). After visualizing ACF and PACF, I used the Augmented Dickey-Fuller test, which produced a p-value consistent with non-stationarity.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

AI-generated content may be incorrect.

Figure : ACF values suggest non-stationarity

I did apply differencing to the milk price data (Figure 2) and performed an Augmented Dickey-Fuller test, which produced a p-value consistent with stationarity.

A graph of milk prices

AI-generated content may be incorrect.

Figure : Differenced Milk Prices

I used the auto.arima function to choose my model and R found that the ARIMA(2,0,2) provided the best fit. The residuals seem normally distributed (Figure 4) upon further analysis, and the autoplot provides evidence of model fit (Figure 5)

A graph of a graph of a graph

AI-generated content may be incorrect.

Figure

A graph of a graph of a graph

AI-generated content may be incorrect.

Figure

The mathematical formula for the ARIMA model is:

yt = -0.024yt-1 – 0.4747yt-1 + 0.3184εt-1 + 0.3184εt-2 + εt

A screenshot of a computer

AI-generated content may be incorrect.

Figure : Forecasts

A graph of a wave

AI-generated content may be incorrect.

Figure

Generally speaking the model produced forecasts close to the log-prices that were then transformed by differencing. This does make the model harder to interpret from a practical aspect, but comparing my data to the forecasts shows that the numbers the forecast produced are very close to the data points in 2024 and 2025.

Seasonal Autoregressive Integrated Moving Average Model (SARIMA)

The SARIMA model has many of the same advantages and disadvantages of the ARIMA model, with the added advantage of handling seasonal data well. My temperature data seemed to have constant variance and both the ACF, PACF, and ADH test results supported stationarity, so I didn’t need to perform any of the transformations that I did for my milk price data.

A graph of different types of data

AI-generated content may be incorrect.

Figure

Figure 9

Using auto.arima, I determined that the best-fitting model was ARIMA(3,0,0)(2,1,0)[12]. The corresponding mathematical formula is yt = 0.1267yt-1 + 0.1062yt-2 + 0.2121yt-3 - 0.5567εt-1 - 0.4785εt-2 + εt

Analyzing the residuals supports goodness of fit for this model.

A graph of a graph of a graph

AI-generated content may be incorrect.

Figure

A graph of a function

AI-generated content may be incorrect.

Figure

The forecast information is included below. The forecast followed the same seasonal trend as the rest of my data.

A screenshot of a computer

AI-generated content may be incorrect.

Figure

A graph with lines and numbers

AI-generated content may be incorrect.

Figure

Autoregressive Conditional Heteroscedasticity Model (ARCH)

Autoregressive conditional heteroscedasticity models allow for data with changing variance. This provides an advantage when modelling volatile data. However, with non-volatile data, predictions may be less accurate [2].

As with my ARIMA model, I log-transformed the data for ARCH and dealt with the problem of stationarity by differencing the data. This produced a model with the following equations:

yt = xt – 0.089

σt = √(0.999y2t-1

The forecasts are as follows:

A screenshot of a computer

AI-generated content may be incorrect.

To comment, there is little variation in the ARCH forecasts. The ARIMA model seems to have predicted the prices better.

Generalized Autoregressive Controlled Heteroscedasticity Model (GARCH)

GARCH incorporates past error terms, which provides an advantage over ARCH. It has the same benefits as managing heteroscedasticity, but the disadvantages include not performing well when there is stable variance in a sample.

The GARCH model provides the mathematical equations:

stationarity by differencing the data. This produced a model with the following equations:

yt = xt – 4.22

σt = √(0.03+0.034y2t-1+0.065ε2t-1

The forecasts are much better with GARCH.

A screenshot of a computer

AI-generated content may be incorrect.

References

[1] Hayes, A. 2024. “Autoregressive Integrated Moving Average Prediction model”. Investopedia.com. <https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp>

[2] Kenton, W. 2024. “Autoregressive Conditional Heteroskedasticity (ARCH) Explained”. Investopedia.com. <https://www.investopedia.com/terms/a/autoregressive-conditional-heteroskedasticity.asp>