6.3730 Pset 4

Problem 1

Part A

The data was initially processed to remove missing entries and then processed through python to create the graphs above. We have t in months and out coefficients are as follows:

α_1 : 307.29304118712025 ppm α_2 : 0.12669573235362713 ppm/month

Our residuals take on a clear quadratic trend which indicates that is convex and that our linear fit might not be the best to fit the data.

Pasted image 20230419142555.png

Figure 1: Linear Fit

Pasted image 20230419142630.png

Figure 2: Linear Residuals

Part B

Beta values: $[3.14436078 * 10^2, 6.55570405 * 10^{-02}, 8.70921537 * 10^{-05}]$ High-frequency oscillation stays the same. The quadratic fit describes data better than linear fit, but it does not appear to match completely

Pasted image 20230419224047.png

Figure 3: Quadratic Fit

Pasted image 20230419224101.png

Figure 4: Residuals

Part C

Pasted image 20230419224117.png

Figure 5: Quartic Fit

Pasted image 20230419224130.png Gamma values: $[3.15774455*10^2, 2.72002845*10^{-2}, 3.28912255*10^{-4}, -5.30108402*10^{-7}, 3.74378873*10^{-10}]$ Between all three the quatric and the quadratic seemed to perform the best and their residuals look fairly similar. It is not always better to use the higher-order fit because you can over-fit your data and lose valuable information by capturing even more noise. We can use techniques like cross-validation and comparing model evaluation metrics (e.g., AIC, BIC, or validation set error). You can also use regularization techniques like Lasso or Ridge regression to control the complexity of the model. MSE for Linear fit: 15.167430269390964 MSE for Quadratic fit: 4.875366661489615 MSE for Quartic fit: 4.685182947296229

Part D

When compared to the seasonally adjusted data our residuals seem comparable when comparing the difference between the original data set and adjusted data.

Part E

Our resulting residuals from adding pi to our quadratic model show that the overall variation in CO2 levels has remained relatively constant.

Pasted image 20230419224819.png

Figure 6: Monthly Residuals

Pasted image 20230419224209.png

Figure 7: Seasonally Adjusted Data

Pasted image 20230419224758.png

Figure 8: Residuals

Problem 2

Part A

Pasted image 20230420171704.png

Figure 9: AR model Order

This figure represents the mean squared error resulting from AR models with orders up from 0 to 24. The Best AR order was 18 followed up by 4. I choose to stick with order 4 in order to reduce the complexity and potential overfitting of the model. By plotting the predictions alongside the original graph we can see a fairly good prediction.

Pasted image 20230420173013.png

Figure 10: Predicted Values vs Actual values

An AR model may be a more suitable starting point for our problem than other types of models because it assumes that the current value of a time series is a linear combination of its past values. It is reasonable to assume that the current CPI value depends on its recent history, making the AR model a good starting point for our analysis.

Part B

Calculate the percentage change in the CPI data from the previous month to the current month. Do the same for your 1-month ahead predictions.

To calculate the monthly inflation from CPI and for Pricestats, I took the month-to-month percent change from the predicted values. BER data was used as an approximation of the expected average annual inflation rate over the next 5 or 10 years and converted the annual inflation rate to a monthly rate.

Part C

In this part I took the average value for each month for the PriceStats and the BER datasets and used them as regressors and did the same using the first value

Pasted image 20230420184546.png

Figure 11: Monthly Inflation Rates

Pasted image 20230421222026.png

Figure 12: CPI with different Regressors

of each month. In addition to these predictions I also included the log(CPI) data as well as the original AR fit. I used SARIMAX in order to compute these two predictions. Afterwards I calculated the MSE for both variations. The difference in the resulting models and error comes from the aggregation method used for external regressors. Using the average value can smooth out short-term fluctuations and capture the overall trend, while using the first day of each month might be more sensitive to short-term changes.

Part D

Pasted image 20230421225007.png

I attempted to get improved results by searching for the best ARIMA parameters and seasonal parameters and then use them for the SARIMAX model. Although in practice it seemed not be be super helpful compared to the other models

Problem 3

Part A

Given the MA(1) model, $X_t = W_t + \theta W_{t-1}$ White noise: E(W_t)= 0 Cov(W_s, W_t)= E(W_s, W_t)= $\delta_{s,t}\sigma^2$

Auto-covariance is

$$\begin{aligned}\gamma(h) = E(X_t X_{t+h}) &= E[(W_t + \theta W_{t-1})(W_{t+H} + \theta W_{t-1+h})] \\ &= \sigma^2(\delta_{0,h} + \theta\delta_{-1,h} + \theta\delta_{1,h} + \theta^2\delta_{0,h})\end{aligned}$$

$$y = \begin{cases}\sigma^2(1+\theta^2), & h = 0 \\ \sigma^2(\theta^2), & |h| = 1 \\ 0, & |h| > 1\end{cases}$$

Part B

Given the AR(1) model, $X_t = \phi X_{t-1} + W_t$

$$\begin{aligned}\gamma(h) = E[X_t X_{t+h}] &= E[X_t(\phi X_{t-1+h} + W_{t+h})] \\ &= E[\phi X_t X_{t-1+h} + X_t W_{t+h}] \\ &= \phi\gamma(h-1) + E[X_t W_{t+h}]\end{aligned}$$

So for h $>$ 0 the second term goes to 0

$\gamma(h) = \phi^h \gamma(0)$

Now using the stationary condition we can say that $\gamma(0) = Var(X_t)$

$$\begin{aligned}\gamma(0) &= Var(\phi X_{t-1}) + Var(W_t) \\ &= \phi^2 Var(X_t) + Var(W_t) \\ &= \phi^2\gamma(0) + \sigma^2 \\ \gamma(0) &= \frac{\sigma^2}{1-\phi^2}\end{aligned}$$

$$\gamma(h) = \phi^h \frac{\sigma^2}{1-\phi^2}$$

As for the equivalence to an MA model, the AR(1) model is equivalent to an infinite-order MA model. The AR(1) process has memory that decays exponentially with the lag, so it cannot be fully captured by a finite-order MA model.