Time Series

Analysis 4 for MIT 6.419x Data Analysis: Modeling and Applications

Evan Rushton

May 17, 2021

1 The Mauna Loa CO_2 Concentration

1. (3 points) Plot the periodic signal P_i. (Your plot should have 1 data point for each month, so 12 in total.) Clearly state the definition of P_i, and make sure your plot is clearly labeled.

Let the model C_i be the average CO_2 concentration in month $i (i = 1, 2, ...,$ counting from March 1958). This model can be decomposed into a quadratic trend $F_2(t_i)$, a seasonal periodic signal P_i and the remaining residual R_i. This can be written, $C_i = F_2(t_i) + P_i + R_i$. Figure 1 shows the predicted periodic signal for the training data. These values are the monthly average residuals of CO_2 concentrations after subtracting the quadratic fit of the training data from the training data, $C_i - F_2(t_i) = P_i + R_i$.

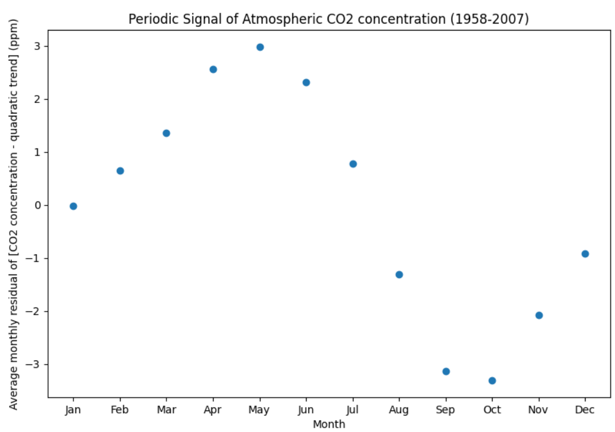

Figure 1: Periodic signal P_i are the monthly average residuals of CO_2 concentrations after subtracting the fitted quadratic $F_2(t_i)$ from the training data C_i.

2. (2 points) Plot the final fit $F_n(t_i) + P_i$. Your plot should clearly show the final model on top of the entire time series, while indicating the split between the training and testing data.

Figure 2 shows the quadratic fit plus the monthly periodic trend of the training data, $F_2(t_i) + P_i$, on top of the entire time series.

3. (4 points) Report the root mean squared prediction error RMSE and the mean absolute percentage error MAPE with respect to the test set for this final model. Is this an improvement over the previous model $F_n(t_i)$ without the periodic signal? (Maximum 200 words.)

The root mean squared prediction error (RMSE) for this final model with respect to the test set is 1.14936 and the mean absolute percentage error (MAPE) is 0.20859%, both of which are better than the quadratic fit $F_2(t_i)$ with an RMSE of 2.50281 and MAPE of 0.53228%. This suggests that this final model $F_n(t_i) + P_i$ is an improvement over the quadratic fit alone. (62 words)

4. (3 points) What is the ratio of the range of values of F to the amplitude of P_i and the ratio of the amplitude of P to the range of the residual R_i (from removing both the trend and the periodic signal)? Is

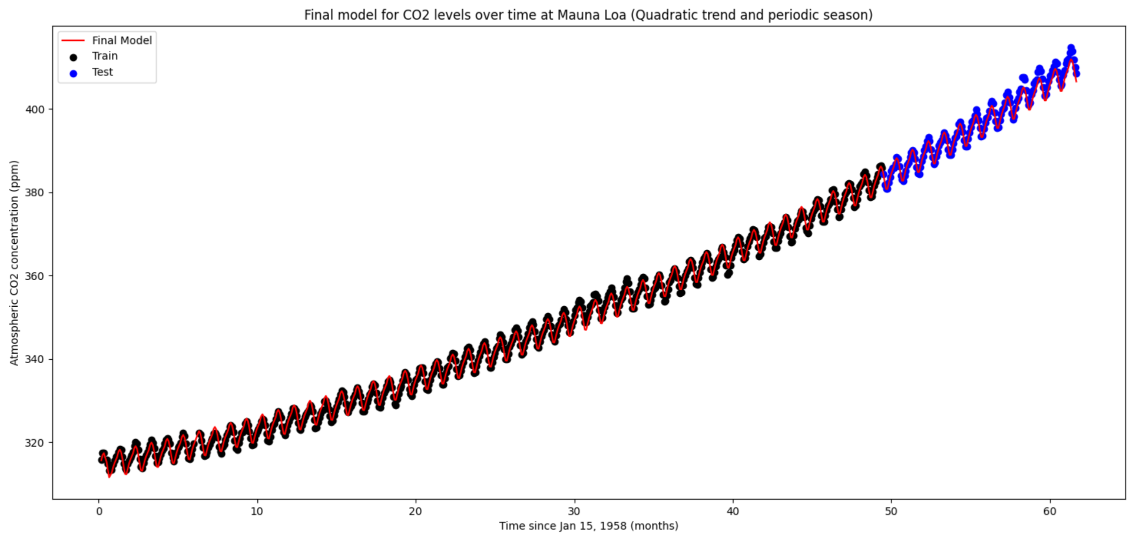

Figure 2: Final model $F_n(t_i) + P_i$ displayed in red on top of the CO_2 time series. The training set are colored black and the test set are colored blue.

this decomposition of the variation of the CO_2 concentration meaningful? (Maximum 200 words.)

As stated in the problem statement, "The decomposition is meaningful only if the range of F is much larger than the amplitude of the P_i and this amplitude in turn is substantially larger than that of R_i." The range of $F_2(t_i)$ over the training set is 69.14 (`model2(X_train).max()-model2(X_train).min()`), the amplitude of P_i is 3.146 (`(p_hat.max()+np.abs(p_hat.min()))/2`, and the range of R_i over the training set is 1.9182 (`(residual["CO2 (ppm)"].max()+np.abs(residual["CO2 (ppm)"].min()))/2`). The ratio $\frac{\text{Range of } F}{\text{Amplitude of } P_i} = \frac{69.14}{3.146} = 21.98$ suggests the decomposition is meaningful because the range of F is much larger than the amplitude of P_i. However, the ratio $\frac{\text{Amplitude of } P_i}{\text{Amplitude of } R_i} = \frac{3.146}{1.9182} = 1.64$ suggests that the decomposition is not meaningful because the amplitude of P_i is not substantially larger than that of R_i. (122 words)

2 Autocovariance Functions

1. (4 points) Consider the $MA(1)$ model, $X_t = W_t + \theta W_{t-1}$, where $W_t \sim W \sim N(0, \sigma^2)$. Find the autocovariance function of X_t. Include all important steps of your computations in your report.

$E[X_t] = 0$ since it is a sum of white noise variables with expectation 0. The autocovariance for this moving average process is

$$\begin{aligned}
\gamma_X(t+h,t) &= \text{Cov}(X_{t+h}, X_t) \\
&= E[(W_{t+h} + \theta W_{t+h-1})(W_t + \theta W_{t-1})] \\
&= E[W_{t+h}W_t + \theta W_{t+h-1}W_t + W_{t+h}\theta W_{t-1} + \theta W_{t+h-1}\theta W_{t-1}] \\
&= E[W_{t+h}W_t] + E[\theta W_{t+h-1}W_t] + E[W_{t+h}\theta W_{t-1}] + E[\theta^2 W_{t+h-1}W_{t-1}] \\
&= \begin{cases} \sigma^2(1+\theta^2) & \text{if } h = 0, \\ \sigma^2\theta & \text{if } h = \pm 1, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Since $E[W_t W_s] = \sigma^2$ when $s = t$ and 0 otherwise for white noise.

2. (4 points) Consider the AR(1) model, $X_t = \phi X_{t-1} + W_t$, where $W_t \sim W \sim N(0, \sigma^2)$. Suppose $|\phi| \leq 1$. Find the autocovariance function of X_t. (You may use, without proving, the fact that X_t is stationary if $|\phi| \leq 1$.) Include all important steps of your computations in your report.

$E[X_t] = \phi E[X_{t-1}] = 0$ since it is a sum of a white noise variable with expectation 0 and X_t is stationary. $E[X_t^2] =$

$\sigma_{AR}^2 = \phi^2 E[X_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1-\phi^2}$. The autocovariance for this autoregressive process is

$$\begin{aligned}
\gamma_X(t+h,t) &= \mathrm{Cov}(X_{t+h}, X_t) \\
&= E[(W_t + \phi X_{t-1})(X_{t-1})] \\
&= E[W_t X_{t-1} + \phi X_{t-1}^2] \\
&= \phi \sigma_{AR}^2 \\
&= \phi \frac{\sigma^2}{(1-\phi^2)}
\end{aligned}$$

Since $E[W_t W_s] = \frac{\sigma^2}{1-\phi^2}$ when $s = t$ and 0 otherwise for white noise.

3 Converting to Inflation Rates

1. Repeat the model fitting and evaluation procedure from the previous page for the monthly inflation rate computed from CPI. Your response should include:

(1 point) Description of how you compute the monthly inflation rate from CPI and a plot of the monthly inflation rate. (You may choose to work with log of the CPI.)

The monthly inflation rate was computed from CPI as the percentage change in CPI per month, $IR_t = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}}$. The first CPI value for each month was used as the representative for the month, CPI_t. The monthly inflation rate from CPI is shown in Figure 3.

(2 points) Description of how the data has been detrended and a plot of the detrended data.

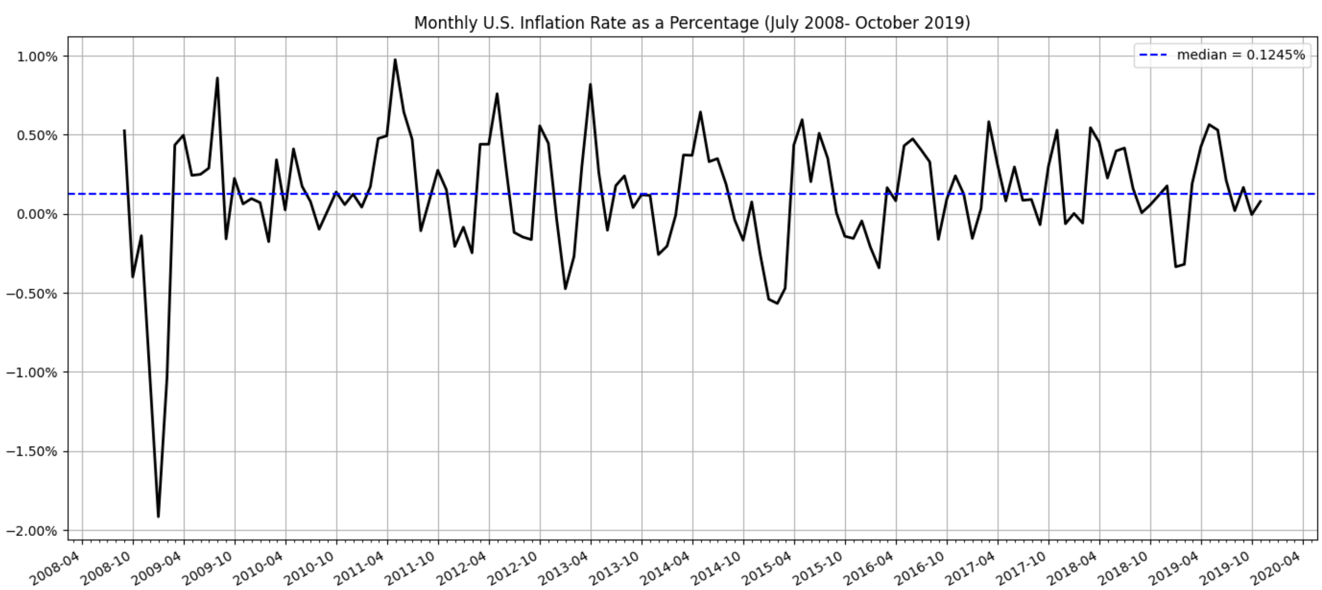

Figure 3: Monthly Inflation Rate from CPI as a percentage (2008-2019)

The monthly inflation rate looks relatively stationary in Figure 3. A linear trend T_t was fit to the training data with a train/test split at September 2013. The detrended data is modeled by $IR_t = T_t + R_t$ where R_t is the residual. The linear trend $T_t = \alpha_1 t + \alpha_0$ has coefficent $\alpha_1 = 5.4937e-05$ and intercept $\alpha_0 = -5.6562e-04$ and is plotted over the data as a red line in Figure 4. The detrended residual R_t is plotted in Figure 7. **(3 points) Statement of and justification for the chosen AR(p) model. Include plots and reasoning.**

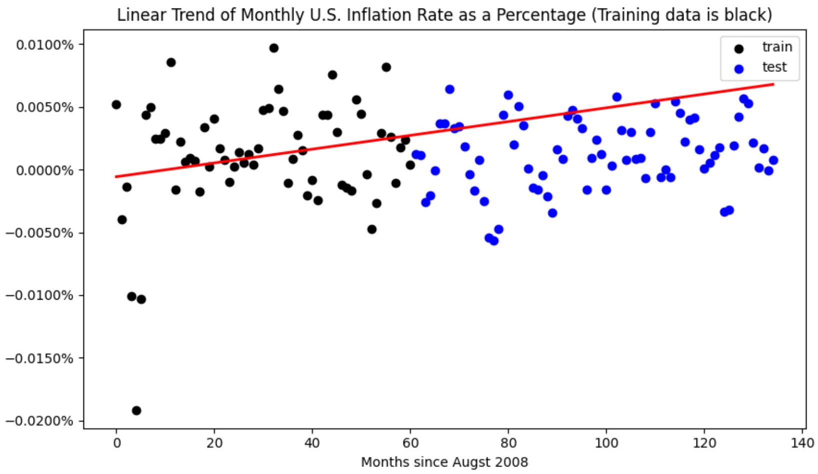

Figure 4: Linear fit T_t in red over the monthly inflation rate computed from CPI IR_t

Figure 5: Detrended residual R_t after subtracting linear fit, $IR_t - T_t$

The remaining residual R_t may be modeled by an autoregressive (AR) or moving average (MA) process. To determine if so and which the autocorrelation and partial autocorrelation functions were plotted. These are shown in Figure 6 and Figure **??**. These plots suggest that these residuals can be modeled by an MA(1) process from the ACF and a AR(1) process from the PACF. Although at lag k = 2 the PACF is right at the cutoff, so it may be better modeled by AR(2).

(3 points) Description of the final model; computation and plots of the 1 month-ahead forecasts for the validation data. In your plot, overlay predictions on top of the data.

Unfortunately, I will not have time to complete. Good luck with taxes all.

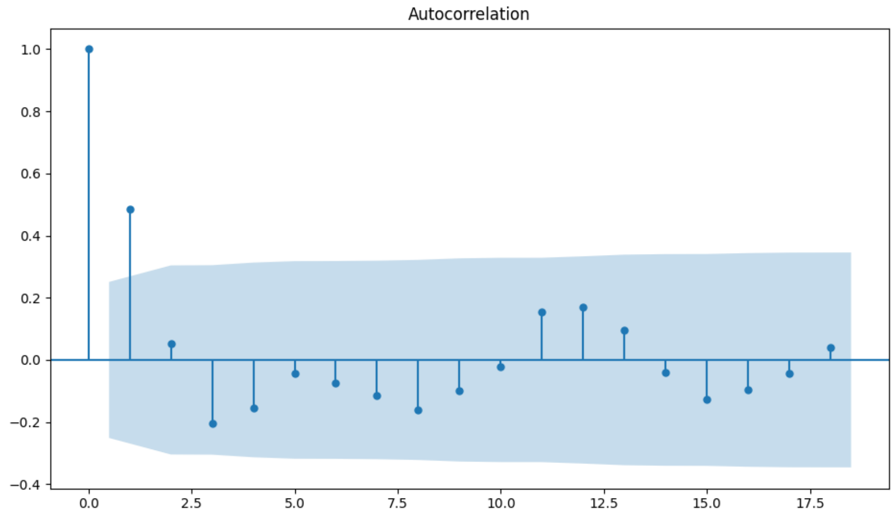

Figure 6: Autocorrelation function plotted for R_t

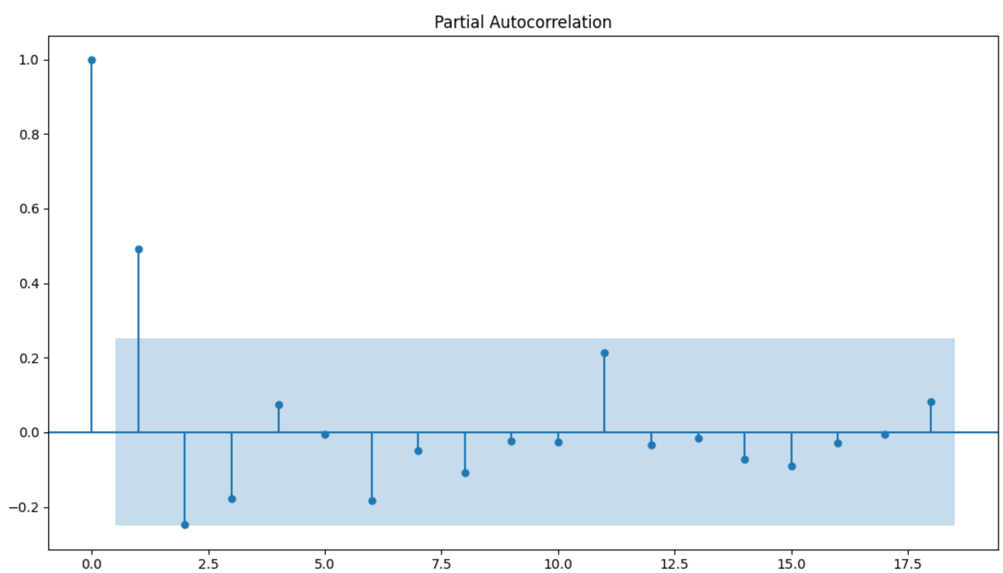

Figure 7: Partial Autocorrelation function plotted for R_t