6.3730 PSet 3 Part 2

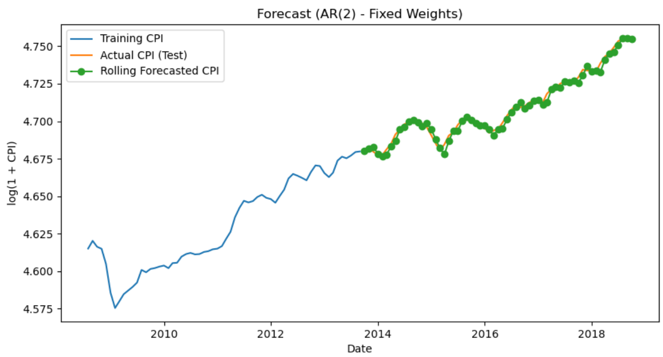
Stephen Andrews

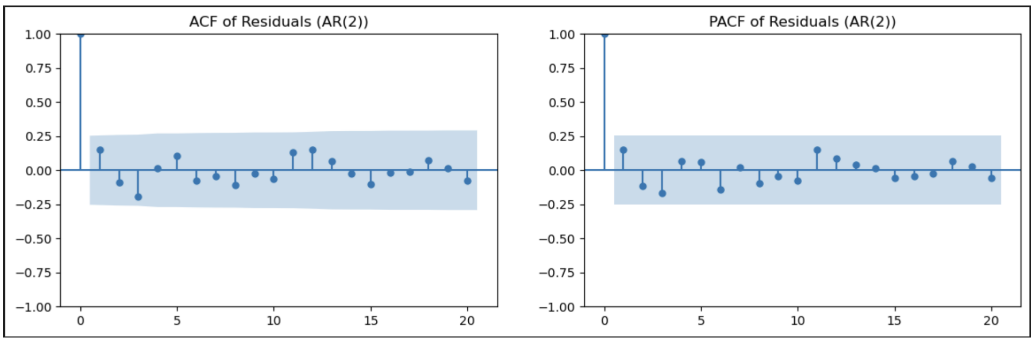
April 5, 2025

3.2:

(a):

After processing the CPI data to use the first CPI value of each month, I fit several AR models AR(1), AR(2), AR(3) and compared their forecast performance. We use log transformed data for the entirety of the problem for reasons discussed in part c. I also examined the ACF and PACF of the residuals at each order to see whether additional lags were needed. Based on both the residual diagnostics and the mean squared error of the forecasts, AR(2) provided the best performance which had an MSE of 6.2355×10^{-6}. We see the forecast graph along with the ACF and PACF graph for AR(2).





The ACF shows strong correlations at lags 1 and 2 that decay quickly afterward, while the PACF has significant spikes only at lags 1 and 2 and then cuts off. This pattern indicates that only the first two lags contain most of the predictive information, making an AR(2) model appropriate.

An AR model is a suitable starting point for this CPI forecasting problem for a few reasons. Firstly, CPI often exhibits persistence. Today's value is closely related to its own recent past. AR models directly capture this "memory" by regressing the current value on its previous lags. In addition, AR models are interpretable as the coefficients show how each lag contributes to the current value, making it straightforward to interpret economic inertia.

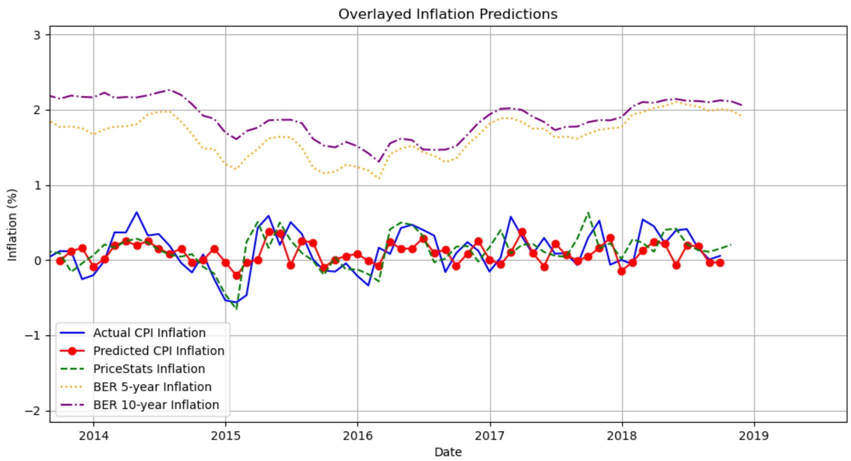
By contrast, a moving average (MA) model focuses on past error terms rather than past observed values. This is less intuitive for CPI data, which usually depends more on its own history (price level momentum) than on random shocks in previous periods. Consequently, starting with an AR model often yields better insights and forecasts for inflation-like processes.

(b):

To compute inflation from the CPI data, you first calculate the percentage change from one month to the next. If you work in log scale, you can approximate inflation for a given month as 100 times the difference between the log of the current CPI and the log of the previous month's CPI. When predicting CPI, you replace the actual CPI value of the current month with the one-month-ahead forecast while still using the actual CPI from the previous month to calculate the inflation rate.

For PriceStats data, since it is recorded daily, you need to aggregate it to a monthly frequency—typically by taking the average over each month—and then compute the inflation using the same percentage change or log difference method as for CPI. The BER data, which are provided as daily figures, should be similarly aggregated to monthly values (either by averaging or by taking the month-end value) so that they can be directly compared with the monthly CPI measures. This approach ensures that all data series are aligned and consistent when assessing the accuracy of the inflation forecasts.

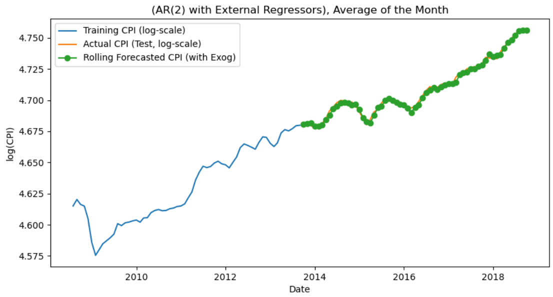
The following overlayed inflation graph is shown below:



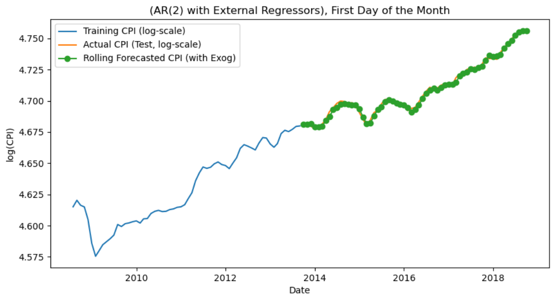
(c):

For both the monthly average and first day ARs, we will be working on the log scale. Working on the log scale is advantageous because differences in logarithms closely approximate percentage changes, which is how inflation is naturally defined. Additionally, the log transformation stabilizes variance and often makes the time series more stationary, improving the reliability and interpretability of our forecasting models.

For the moving average AR, we get the following forecast:



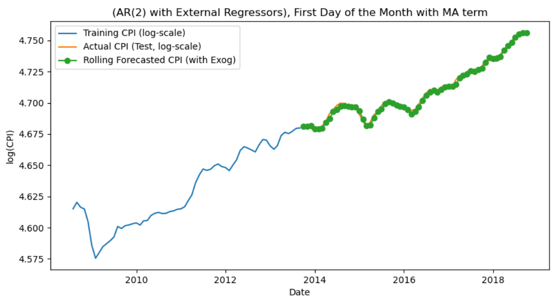
and for the first day AR, we get the following forecast:



Both forecasts seem to perform decently but the first day AR has slightly lower MSE at 3.181×10^{-6} where the average AR has MSE at 4.287×10^{-6}. Using the monthly average tends to increase prediction error because it smooths out the early-month signal that appears to be most predictive of next month's CPI. In other words, averaging incorporates later-day fluctuations that dilute the timely, informative value available on the first day of the month, making the model less responsive to the conditions that drive CPI changes.

(d):

The lowest MSE was achieved by introducing an MA term which got it to 2.887×10^{-6}. The forecast graph is below:



3.3:

(a):

Since $E[W_t] = 0$, it follows that

$$E[X_t] = E[W_t + \theta W_{t-1}] = 0.$$

And because W_t and W_{t-1} are uncorrelated and each has variance σ^2,

$$\mathrm{Var}(X_t) = \mathrm{Var}(W_t + \theta W_{t-1}) = \sigma^2 + \theta^2\sigma^2 = \sigma^2(1 + \theta^2).$$

The for the autocovariance,

- $k = 0$:
$$\gamma_0 = \mathrm{Var}(X_t) = \sigma^2(1 + \theta^2).$$
- $k = 1$:
$$\gamma_1 = \mathrm{Cov}(X_t, X_{t-1}) = \mathrm{Cov}(W_t + \theta W_{t-1},\, W_{t-1} + \theta W_{t-2}) = \theta\sigma^2,$$
since only the term $\theta W_{t-1} \cdot W_{t-1}$ contributes.
- $k > 1$, $\gamma_k = 0$ because non-overlapping W_t terms are uncorrelated.

(b):

A stationary AR(1) process has:

$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2}, \quad \gamma_k = \gamma_0\, \phi^{|k|}.$$

This shows the autocorrelation decays polynomially with lag. An AR(1) process can also be written as:

$$X_t = \sum_{j=0}^{\infty} \phi^j\, W_{t-j},$$

which shows that it is equivalent to an MA(∞) process.