ECON612 Final Project: The Forecast of the Unemployment Rate in California

Lingzhi Dang

August 7, 2025

Introduction and Description of Variables

In this project, we will forecast the CAUR, which is the seasonally adjusted unemployment rate in California for the next 12 months(hereinafter referred to as the unemployment rate). The unemployment rate data spans from Jan. 1976 to Feb. 2025.

To have the data on the unemployment rate, the government conducts a monthly survey called the Current Population Survey (CPS) to measure the extent of unemployment in the state. The survey includes individuals aged 16 and older who reside in California, excluding those in institutional settings (e.g., penal and mental facilities, homes for the aged) and individuals on active military duty. Briefly, the unemployment rate is defined as the proportion of unemployed individuals within the labor force. People who are jobless, looking for a job, and available for work are unemployed. The labor force is made up of the employed and the unemployed. People who are neither employed nor unemployed are not in the labor force. Notice that students during semesters are usually not considered as labor force. Meanwhile, relatives doing chores in the house are considered employed, which are called unpaid family workers. Further details on how governments measure unemployment, including eligibility criteria, can be found at this link.

In the graph of the unemployment rate data, we can see that the unemployment rate likely follows an AR process with no trend pattern. It also does not have a seasonal pattern since it is seasonally adjusted data. The unemployment rate fluctuates around some positive level with a minimum value 4 percent. In this graph, we can also see some local maxima in 1983, 1993, 2003, 2010, 2020. There may be some evidence to support the increase in the unemployment rate per decade. The unemployment rate is not likely to be a random walk. To see that, we will use some tests later to prove/disprove it.

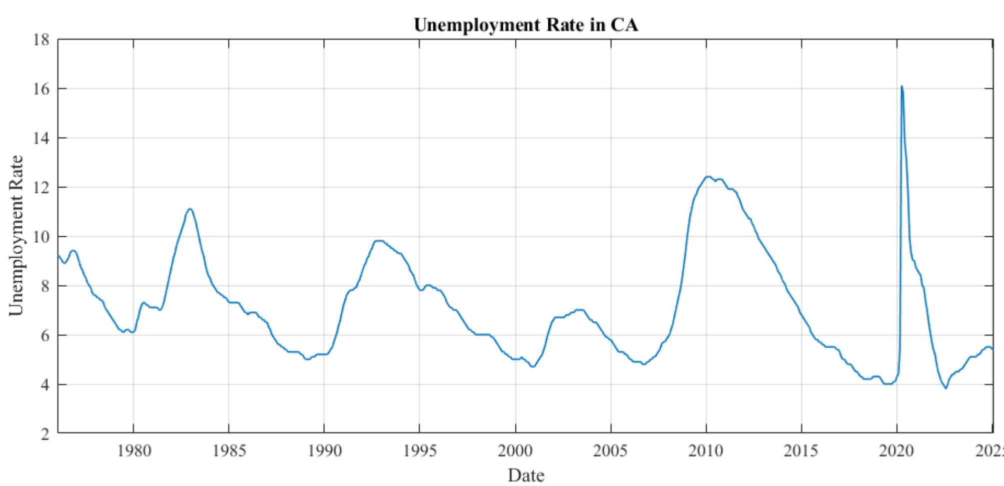


Figure 1: The seasonally adjusted Unemployment rate in CA

General Idea of Selecting Models and Forecasting

In this paper, we will start by checking whether the unemployment rate is stationary. If necessary, we change the data to some stationary data. We found that the unemployment rate is stationary and likely follows an AR model. Then, we estimate an AR model and forecast the unemployment rate based on that AR model. After the baseline AR model, we introduce a strong leading indicator, the junk bond spread, forming an ADL model to do another forecast. We then find that the ADL is a better model in forecasting since it has a lower AIC value. In addition to those 2 models estimated under the full sample size, we then consider excluding some data during the COVID period and re-estimating the ADL model. By comparing the PLS, we finally select the ADL(9,3) model estimated in the sample set, which excludes several COVID observations and choose that model for the last forecast. In the following parts, more details and justifications of our choice of model will be shown.

1 Analysis of the Unemployment Data

We will start the analysis of the unemployment rate and select the model by checking the stationarity of the unemployment rate data. To consider the p of AR(12), we follow the convention that the maximum value of p we will probably choose is 12.

1.1 Statioanry

In the previous section, we had a brief look at the unemployment rate graph. To test whether there is a unit root formally, we use the Augmented Dickey-Fuller (ADF) test, setting ARD as the model with 11 lags of differenced terms. We select ARD since we can see from the plot that the unemployment rate fluctuates around some positive level and does not have a trend pattern. Thus, the test will follow:

$$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^{11} \gamma_i \Delta y_{t-i} + \varepsilon_t$$

By testing whether the $\rho = 0$, we can judge whether there exist unit roots. The p-value of the test is 0.0156, which rejects the null hypothesis of the unit root at 5% significance level. Besides viewing the graph and applying the ADF test, we also have some papers supporting the unemployment rate's stationarity. For example, the empirical finding of Omay et al[1]. supports the stationarity of the unemployment rate in 47 states. Also, Lanaspa[2] analyses the time series properties of the unemployment rates of the 50 US States and rejects the null hypothesis of a unit root. This guarantees our feasibility of using AIC and some other tests.

1.2 ACF of Unemployment Rate

To get a rough idea of the unemployment rate and the model we want to use, we begin by drawing the ACF. From the ACF function, we see that the autocorrelation decays gradually as the lag increases. There is no sudden cutoff at some lag, which suggests the data is likely to follow an autoregressive process rather than a moving average process. This result suggests we start with an AR model. To see whether the data follows an AR(p) process and what value of p we should choose, we need to work more on it.

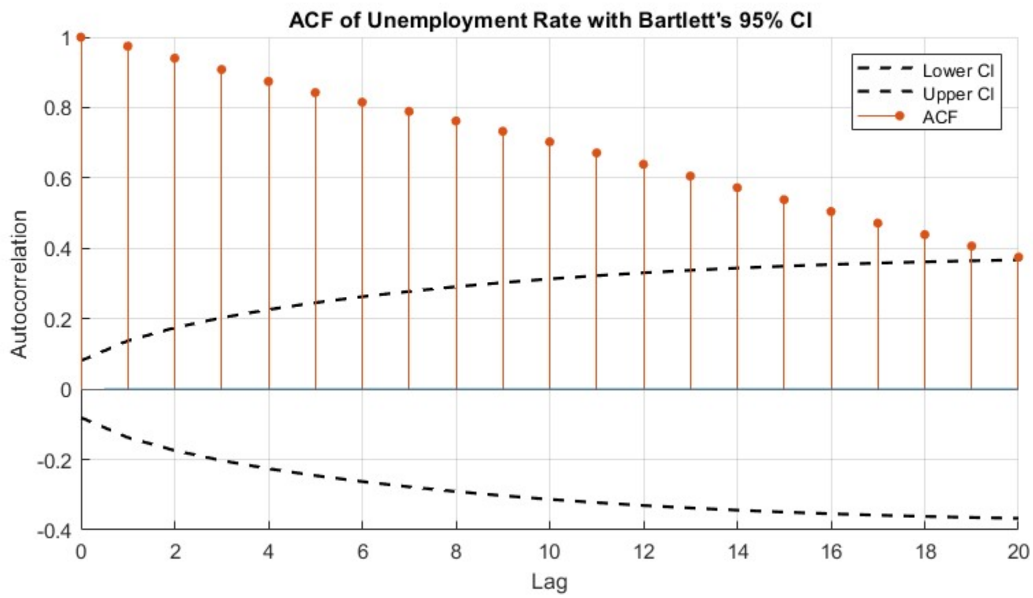


Figure 2: ACF graph

2 Model Selection and Baseline AR Model

From the plot of the unemployment rate, we see that although the unemployment rate varies across time, there is no clear upward trend or downward trend over time. Since it is the seasonally adjusted data, there is no seasonal pattern. Because of the mean of the unemployment rate is positive, we consider AR(p) model with an intercept.

2.1 Selecting AR(p) Model

We are selecting the optimal number of lags (p) for the AR(p) model. Since our goal is to have the best model in forecasting the upcoming unemployment rate, we would mainly use Akaike Information Criterion (AIC), which has AIC = Tln(SSR/T) + 2(1 + p) for the AR(p) model with intercept. Unlike BIC, which is used to select the true model or most likely true model, AIC, as the unbiased estimate of the mean squared forecast error(MSFE), is designed to find the models with low forecast risks. The model with the lowest AIC value is preferred for forecasting. Therefore, we estimate the AIC of a series of AR(p) models from p=1 to 12 with the same restricted sample set without the first 12 observations. Table 1 shows the results of AIC for 12 models. Notice that to compare different AIC values, we need the AIC value calculated based on the same sample.

Table 1: AIC Values for AR(p) Models

AR(p) model	AIC
1	796.52
2	776.53
3	776.33
4	**774.43**
5	775.16
6	774.94
7	776.66
8	778.35
9	778.28
10	780.23
11	781.27
12	782.8

The AIC result suggests we choose the AR(4) model with an intercept as our best model of those 12 for forecasting the unemployment rate. The model is $y_i = \alpha + \beta_1 y_{t-1} +$

$\beta_2 y_{t-2}+\beta_3 y_{t-3}+\beta_4 y_{t-4}+\epsilon_t$, where β_i captures the associations between the unemployment rate and its own lag when holding others constant. α is a constant term and ϵ is white noise. We may use OLS to estimate the value of these parameters.

2.2 Can AR(4) Model Make Residual White Noise

If a model is good enough to explain the data, the residual of the data, which is calculated by the real in-sample data minus the fitted value in the sample, should be white noise. That means the model captures all the predictable variation of the data and leaves with a white noise residual part. For the AR(4), we calculate the in-sample fitted value, plot the real data with the fitted value, and then calculate the residual. We find that, except for the COVID period, the residual follows the white noise characteristics. Moreover, we also plot the autocorrelation function of the residual. For all lags > 1, the autocorrelation remains indifferent to zero, suggesting that the residual is white noise. Figures 3, 4, and 5 show that the AR(4) model does a good job of explaining the data, besides being good at forecasting the future unemployment rate.

2.3 Forecasting the unemployment rate by AR(4)

Plug-in or iterated method in this case will be too complicated to use and may mess up the error term over 12 horizons. Therefore, to forecast the upcoming 12-period unemployment rate, we use the direct method. It means that we need to have 12 separate regressions to forecast y_{t+h} for h=1,2,...,12. It follows that: $y_t = \alpha + \beta_1 y_{t-h} + \beta_2 y_{t-h-1} + \beta_3 y_{t-h-2} + \beta_4 y_{t-h-3} + \epsilon_t$.

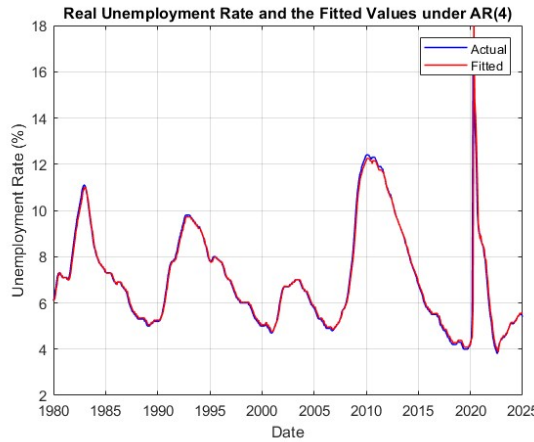


Figure 3: fitted vs real unemployment rate

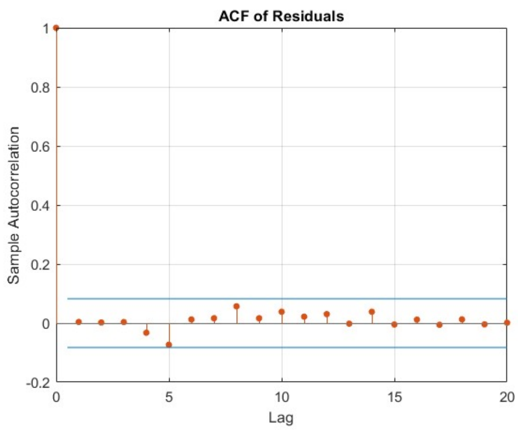


Figure 4: ACF of AR(4) Residual

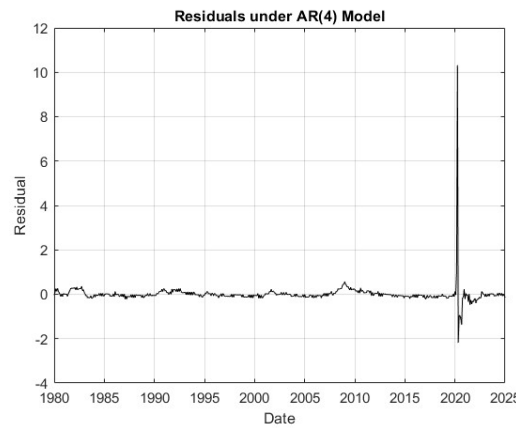


Figure 5: Residual of AR(4)

Once we have those regressions, we use the built-in command predict to get the point and interval forecast. Figure 6 shows the point and interval forecast of the upcoming 12 unemployment rate. To see all 12 regressions and coefficients, please check Appendix 2. The immediate point forecast for March is 5.4537, with a 90% forecast interval [4.6833, 6.2242] in percentage. One can check the specific values of the point and interval forecast for the next 12 months in Appendix 2.

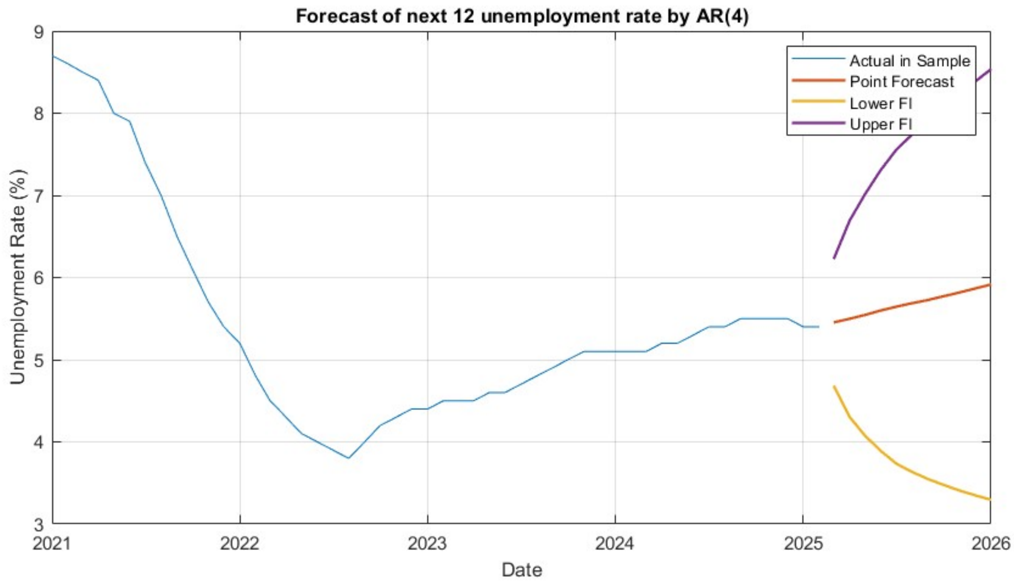


Figure 6: AR(4) forecast

3 Alternative ADL model

We have shown that AR(4) is good at fitting and forecasting. To try to improve the forecasting power, we will introduce the junk bond spread as a leading indicator to see whether it is helpful. There are papers showing that the spread of bonds is an excellent indicator for forecasting the unemployment rate. For example, Gertler and Lown[3] explore a model that captures the effect of financial factors, such as high-yield spread, on the business cycle. The change in the spread is a signal of the downward movement of the business cycle that influences the unemployment rate.

The junk bond spread is defined as the rate(Bond Yield) difference between the Moody's Seasoned BAA-rated bond and the Moody's Seasoned AAA-rated bond. The spread somehow reflects the investor's attitude towards the economy and their risk preference. In Figure 7, we can see that in some periods, an increase in the spread will be followed by an increase in the unemployment rate. To use junk bond spread as a predictor of the unemployment rate, we need to do the Granger causality test to show that the spread can predictively cause the unemployment.

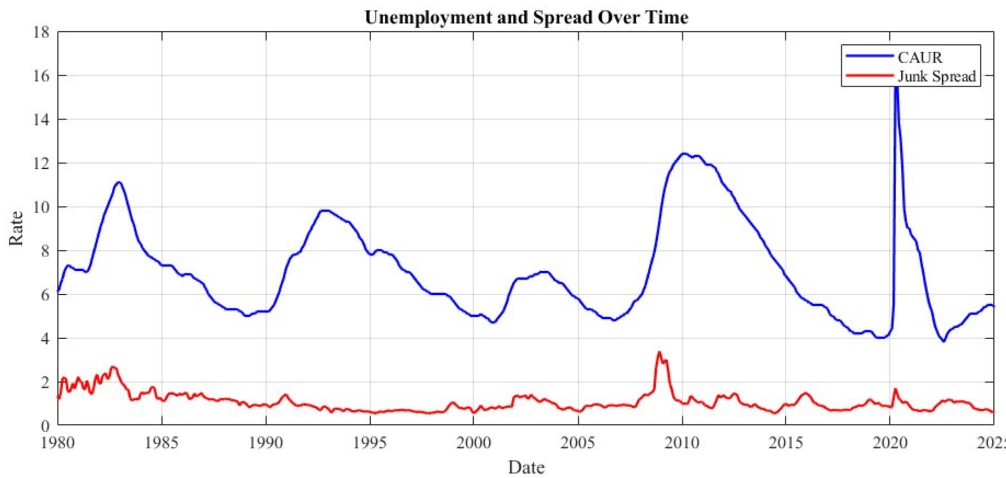


Figure 7: the unemployment rate and the spread

Meanwhile, by analyzing the correlogram of the spread, we observe that it likely follows an autoregressive (AR) process. This implies that the spread can be represented as a function of its own past values. Consequently, we can model the relationship between unemployment (y_t) and the spread (x_t) by incorporating the lagged values of x_t. That enables us to use the ADL model.

Besides the junk bond spread, there is another good indicator we mentioned in lecture, which is the treasury spread. There are 2 reasons that I do not use it. Firstly, it is somehow associated with the junk bond spread, and it is not as good as the junk bond spread as we showed in the lecture. More importantly, the data on treasure spread between 10 years and 3 months starts from 1982, which is much later than the start of the unemployment

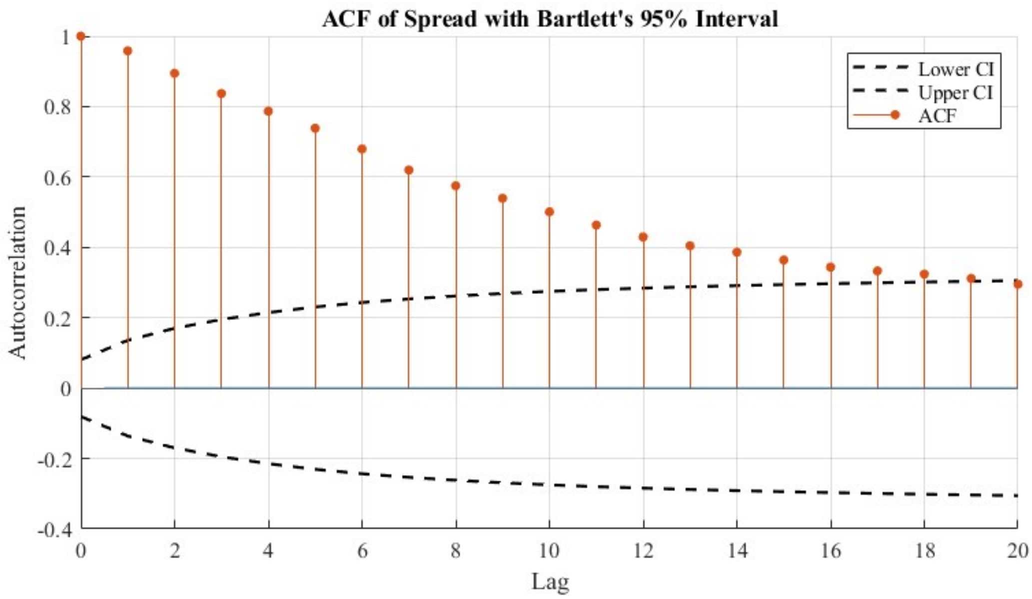


Figure 7.5: ACF of the Spread

rate data. To accommodate it, we may shrink the size of both the junk bond spread and the unemployment rate data. We do not have reasons or strong intuitions to discard early data, so we only consider the spread as a candidate.

3.1 Stationary and Granger Causality of Spread

To see whether spread is a good leading indicator of the unemployment rate, we want to show that the spread is a good time series that does not contain a unit root and has Granger causality on the unemployment rate. From the plot, we see that the spread does not have a trend pattern and fluctuates around a positive level. Thus, to test the unit root, we use the ADF test, specifying the model to be ARD with 11 lags as we defined previously. The result p-value is 0.0066. Thus, we reject the null hypothesis of the unit root and conclude that the spread is stationary.

Then, we want to formally show whether there is an association between the spread and the unemployment rate. We will use the Granger causality test, which tests whether the values of a variable have information that facilitates a forecast on another variable. Because the max model we consider is p=12 and q=12 by convention, we use the built-in gctest in MATLAB, selecting lags=12. The result p-value is $2.7228*10^{-6}$, which rejects the null hypothesis that all coefficients are zero. We conclude that the spread predicatively causes the unemployment rate. One may also use the Wald test under "hac" to test whether all coefficients are not statistically different from 0.

3.2 Selecting ADL(p) Model

Similar to AR(p) case, we use AIC to select the best ADL(p,q) model for forecasting. The ALD(p,q) is defined as:

$$y_t = \mu + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \beta_i x_{t-i} + \epsilon_t$$

where y_t is the unemployment rate and x_t is the spread in time t. The range of p is 1-12, and the range of q is 0-12. By using the built-in AIC syntax, we find that the smallest AIC value is 752.73 and the corresponding model is ADL(6,3). Since the max lag is 12, which is the same as we calculated the AIC of the AR(p) model, those AIC values can be compared with the AIC value of AR(p) models since their AIC values are calculated from the same sample. The AIC value shows that the ADL(6,3) model is better in forecasting the unemployment rate than AR(4).

3.3 Forecasting the unemployment rate by ADL(6,3)

We estimate the ADL(6,3) by the direct method to forecast the next 12 unemployment rates. Similarly, there are 12 separate regressions. For h=1,2,...,12, those 12 regressions each follow the model:

$$y_t = \mu + \alpha_1 y_{t-h} + \alpha_2 y_{t-h-1} + \alpha_3 y_{t-h-2} + \alpha_4 y_{t-h-3} + \alpha_5 y_{t-h-4} + \alpha_6 y_{t-h-5} + \beta_1 x_{t-h} + \beta_2 x_{t-h-1} + \beta_3 x_{t-h-2} + \epsilon_t$$

Similarly, we have intercept, associations between unemployment and its lag and the lag of spread, represented by $\mu.\alpha, \beta$. ADL(6,3) means the model have 6 lags in the unemployment rate and 3 lags in the spread. Once we have those regressions, we use the built-in command *predict* to get the point and interval forecast. Figure 8 shows the point and interval forecast of the upcoming 12 unemployment rate. To see all 12 regressions and coefficients, please check Appendix 3. The immediate point forecast for March is 5.3586, with a 90% forecast interval [4.6036, 6.1135] in percentage. The specific value of next 12 forecasting is also presented in Appendix 3.

4 Adjusting Data in COVID and Its Justification

In the previous section, we showed that including distributed lags, changing AR to an ADL model, will increase the AIC value. Thus, in this chapter, we will focus on different optimal ADL models.

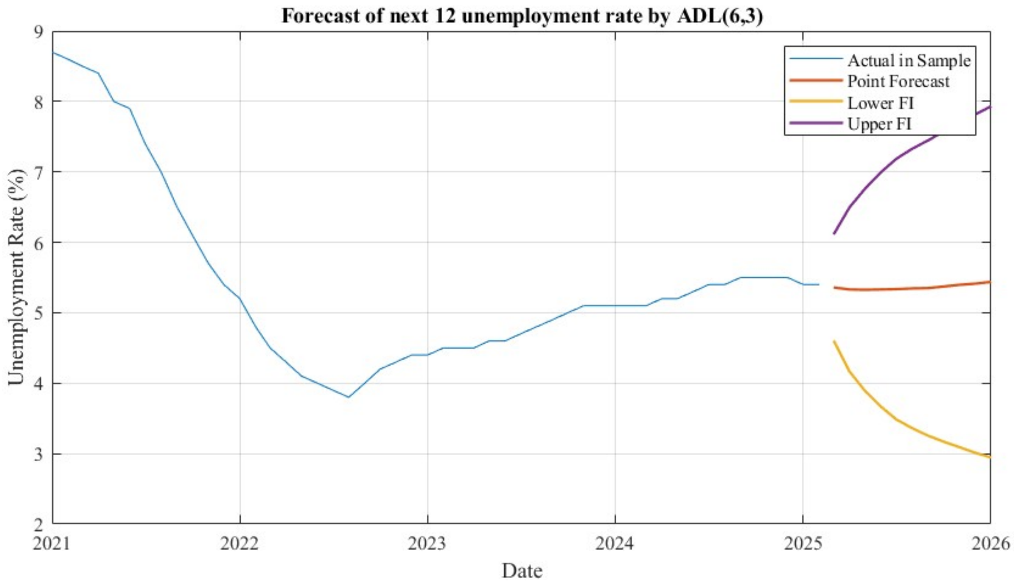


Figure 8: ADL(6,3) Forecast

4.1 Considering Extreme Value

There are 4 observations during the COVID period that can be considered as extreme values under the COVID shock. From April 2020 to July 2020, the unemployment reached an extremely high level, which is higher than at previous record. The unemployment rate in those months is 16.1, 15.8, 13.8, and 13.1, which is even higher than the previous record 12.4% in 2010. Those extreme value that happens during special shocks may not be useful. Excluding those variables can exclude some occasional shocks' effect on our estimated model, which is used to forecast the relatively stable unemployment rate in recent periods.

Besides those reasons, in the next part, we will show that by excluding those 4 observations and reselecting the optimal ADL model, the new optimal ADL(p,q) estimated from the sample without 4 observations will have smaller PLS than the previous ADL(6,3) estimated under the whole sample set. The smaller PLS gives us another justification for excluding several observations for forecasting.

4.2 New Optimal model and PLS comparison

Previously, we had 590 observations of the unemployment rate. By excluding 4 extreme values, we have 586 observations. In the new set with 586 observations. We calculate the AIC value of ADL(p,q) with p=1-12, q=0-12 again to select a new best forecasting model in a new set. One can check the full results of AIC in Appendix 4. We find that the smallest AIC value is 239.72 and the corresponding model is ADL(9,3).

We can not compare AIC values that come from different sample sets. To compare the forecasting performance of models in different sets, we would rely on the PLS criterion. To make the PLS values of difference models comparable, we need to make sure the number of pseudo-out-of-sample forecasts is the same. We would consider Jan 2022 to Feb 2025 as periods we want to "forecast" based on data before 2022. PLS is defined as

$$\mathrm{PLS} = \sqrt{\frac{1}{P}\sum_{t=R+1}^{T}(y_t - \hat{y}_t)^2}$$

P is the number of pseudo-out-of-sample forecasts defined as T-R. T represents the number of samples you have. R represents the threshold that you use to separate your sample into the in-sample and pseudo-out-of-sample parts. We need to estimate ADL(6,3) and ADL(9,3) in different in-sample sets and do a forecast for the same amount of period R. BY constructing syntax in MATLAB, we find that the PLS for ADL(9,3) is 0.1452.

Meanwhile, the PLS for ADL(6,3) is 0.1452, which is much higher than 0.0991. It implies that deleting the COVID data as well as changing the model are likely to have a better forecast with smaller error. Also, the varience of the forecast shrinks.

4.3 Best Forecasting the unemployment rate by ADL(9,3)

We would use the optimal forecasting model indicated by the AIC value in the dataset without the 4 observations during COVID. We estimate the ADL(9,3) by the direct method to forecast the next 12 unemployment rates, so there are 12 separate regressions. For h=1,2,...,12, those 12 regressions follow:

$$y_t = \mu + \sum_{i=1}^{9} \alpha_i y_{t-h-i+1} + \beta_1 x_{t-h} + \beta_2 x_{t-h-1} + \beta_3 x_{t-h-2} + \epsilon_t$$

Once we have those regressions, we use the built-in command predict to get the point and interval forecast. Figure 9 shows the point and interval forecast of the upcoming 12 unemployment rate. To see all 12 regressions and coefficients, check Appendix 4. The immediate point forecast for March is 5.3805, with a 90% forecast interval [4.8951, 5.8659] in percentage. The following table shows my final forecast of the next 12 unemployment rate. Since there are only 12 coefficients, overestimating is still not a big threat to our model.

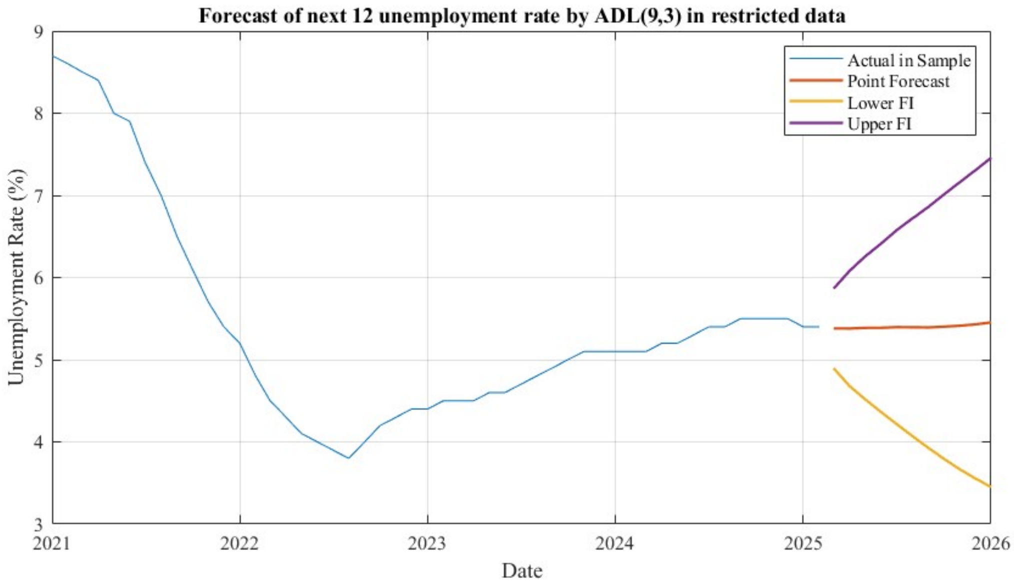
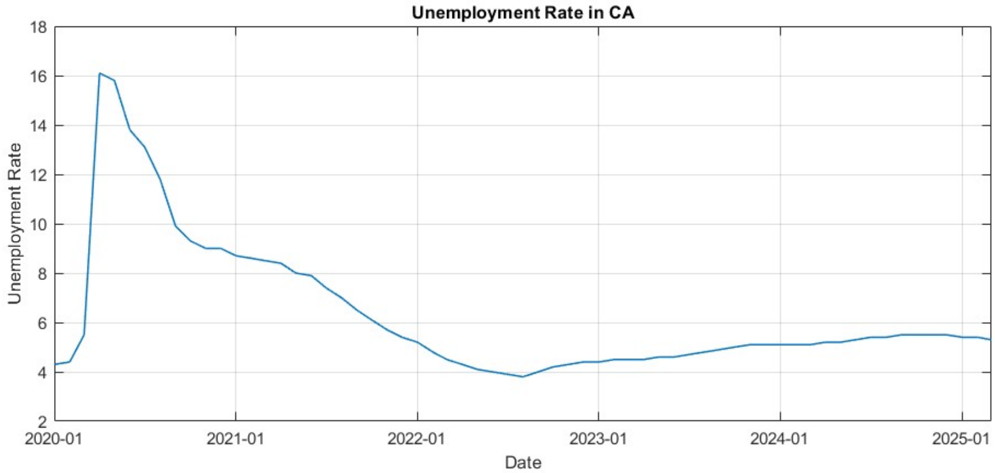


Figure 9: ADL(9,3) Forecast

Final point and Interval Forecast under ADL(9,3)

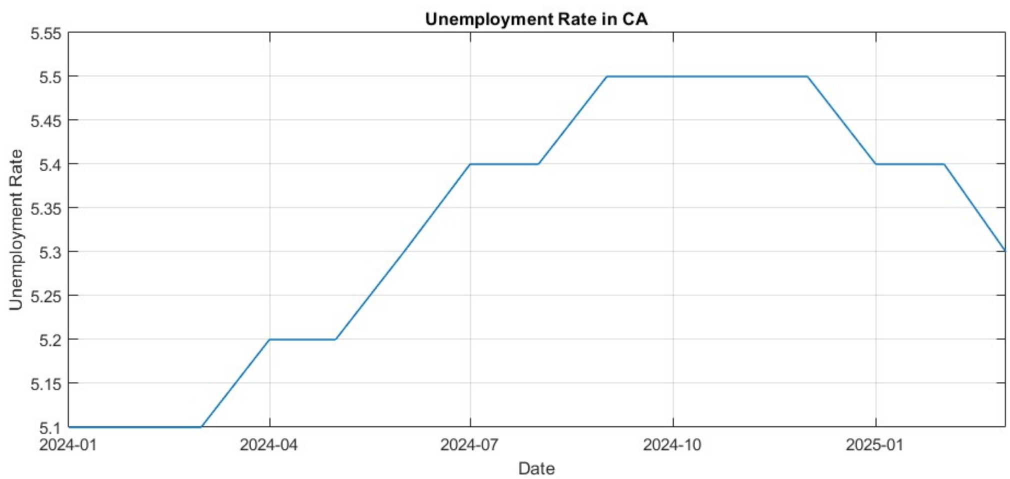
Month	Point Forecast	LB of Interval	UB of Interval
2025-3	5.3805	4.8951	5.8659
2025-4	5.3801	4.6784	6.0819
2025-5	5.3876	4.5190	6.2562
2025-6	5.3881	4.3641	6.4121
2025-7	5.3961	4.2185	6.5736
2025-8	5.3947	4.0710	6.7185
2025-9	5.3929	3.9272	6.8586
2025-10	5.4019	3.7946	7.0092
2025-11	5.4134	3.6675	7.1594
2025-12	5.4295	3.5563	7.3027
2026-1	5.4537	3.4510	7.4565
2026-2	5.4838	3.3628	7.6048

5 Forecast Evaluation



Real Unemployment Rate until March 2025

5.1 Real Result and My Forecast



Real Unemployment Rate until March 2025

The actual unemployment rate in March 2025 was 5.3%. My final point forecast, derived from data through February 2025, was 5.3805%, which is slightly above the realized value. The corresponding prediction interval, [4.8951%, 5.8659%] contains the new observation.

Overall, the projection is relatively accurate and precise because the realization lies within the interval, and the point estimate remains close to the true value. Figure 10 illustrates the position of the actual rate within the forecast region.

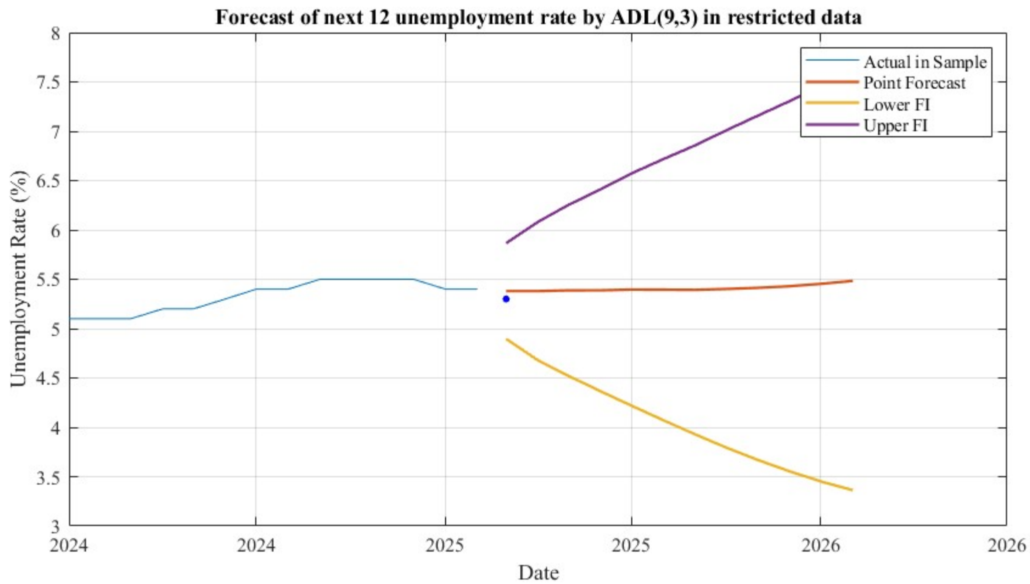


Figure 10: Forecast and Real Unemployment Rate

Decision makers will be wise to follow the forecast. That is because the model was chosen through a rigorous decision framework that utilizes a strong signal. Meanwhile, we exclude some extreme observations caused by shocks, thereby enhancing forecast performance, as confirmed by PLS diagnostics. The resulting forecast error is below 0.1%, a level sufficiently small for practical policymaking.

Although an alternative ADL(6, 3) model estimated, which is the best model under restricted data, delivers a point forecast of 5.3586%, which is slightly closer to the realized rate, I would not adopt it retrospectively. PLS diagnostics indicate that excluding extreme observations from the COVID-19 period will decrease the forecast error. Then, in the restricted dataset, the AIC selects a larger ADL(9, 3) specification as the best model, which approximately minimizes the risk of forecasting. This choice reflects more dependence of unemployment on its own historical values, consistent with typical business-cycle dynamics. In the economy, even remote past values can significantly influence current outcomes.

Moreover, although the unemployment plot reveals a six-month decline, suggesting that a model using less than six months of data might yield lower forecasts (closer to realized data), this brief downturn is unreliable because the sample is too small to establish a robust trend. By contrast, a three-year window shows unemployment has remained broadly stable—or even edged upward—offering perhaps stronger evidence than the recent dip. Consequently, I still decline to replace the larger model with a smaller one, both because statistical diagnostics favor the former and because unemployment exhibits pronounced persistence. Based on the restricted dataset, the more comprehensive ADL (9, 3) specification incorporates a longer history and therefore produces a higher forecast than its ADL (6, 3) counterpart. The larger model has only 13 coefficients, which is still too few to raise concerns about overfitting.

Beyond reasons already noted, I personally anticipate a higher unemployment rate because several new uncertainties have surfaced in recent months. The presidential transition and the early signals of a trade war, among other factors, have substantially heightened economic volatility. Microeconomic theory suggests that this instability is likely to prompt firms to adopt more cautious strategies, curtailing or even reducing production, which would keep unemployment steady or drive it higher. Therefore, many others and I will agree with a larger model giving a larger forecast value.

All in all, I would retain my model and expect it to provide a better forecast for April's unemployment rate. Due to the shock associated with the new president, the realized unemployment rate for April may turn out to be higher than the point forecasts from the ADL(9,3) model, but it is still likely to fall within the forecast interval.

Reference

U.S. Bureau of Labor Statistics, Unemployment Rate in California [CAUR], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CAUR, April 17, 2025.

Moody's, Moody's Seasoned Aaa Corporate Bond Yield [AAA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/AAA, April 17, 2025.

Moody's, Moody's Seasoned Baa Corporate Bond Yield [BAA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/BAA, April 17, 2025.

Federal Reserve Bank of St. Louis, 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity [T10Y3MM], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/T10Y3MM, April 17, 2025.

Omay, T., Ozcan, B., & Shahbaz, M. (2020). Testing the hysteresis effect in the US state-level unemployment series. Journal of Applied Economics, 23(1), 329–348.

Lanaspa, L. (2002). The unemployment structure of the US States.

M Gertler, CS Lown, The information in the high-yield bond spread for the business cycle: evidence and some implications, Oxford Review of Economic Policy, Volume 15, Issue 3, September 1999, Pages 132–150, https://doi.org/10.1093/oxrep/15.3.132

Appendix 2

Point and Interval Forecast under AR(4)

Month	Point Forecast	LB of Interval	UB of Interval
2025-3	5.4537	4.6833	6.2242
2025-4	5.4982	4.3030	6.6933
2025-5	5.5449	4.0742	7.0155
2025-6	5.6001	3.8876	7.3127
2025-7	5.6458	3.7344	7.5572
2025-8	5.6882	3.6322	7.7443
2025-9	5.7277	3.5460	7.9094
2025-10	5.7727	3.4756	8.0698
2025-11	5.8186	3.4075	8.2297
2025-12	5.8666	3.34925	8.3840
2026-1	5.9160	3.2953	8.5368
2026-2	5.9705	3.2519	8.6890

AR(4) h=1 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	0.20593	0.067595	3.0466	0.0024197
x1	1.1811	0.041349	28.565	8.9545e-113
x2	-0.28687	0.063783	-4.4976	8.302e-06
x3	0.15775	0.063783	2.4733	0.013672
x4	-0.081614	0.041339	-1.9742	0.048828

AR(4) h=2 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	0.43988	0.10495	4.1913	3.2059e-05
x1	1.1052	0.064143	17.231	4.9181e-54
x2	-0.16849	0.09894	-1.703	0.089112
x3	0.049124	0.098938	0.49651	0.61972
x4	-0.049157	0.064125	-0.76658	0.44364

AR(4) h=3 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	0.64536	0.12926	4.9928	7.8854e-07
x1	1.1296	0.07893	14.312	5.3655e-40
x2	-0.23667	0.12175	-1.9439	0.05239
x3	-0.013482	0.12175	-0.11074	0.91186
x4	0.027589	0.078908	0.34964	0.72674

AR(4) h=4 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	0.85943	0.15066	5.7045	1.864e-08
x1	1.0911	0.091912	11.871	3.0784e-29
x2	-0.30995	0.14177	-2.1863	0.029194
x3	0.082887	0.14177	0.58465	0.55901
x4	0.012134	0.091893	0.13204	0.895

AR(4) h=5 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.0722	0.16831	6.3704	3.8545e-10
x1	0.97413	0.10259	9.4957	5.7343e-20
x2	-0.21045	0.15824	-1.33	0.18405
x3	0.096057	0.15824	0.60705	0.54406
x4	-0.01429	0.10256	-0.13933	0.88924

AR(4) h=6 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.2745	0.18123	7.0328	5.758e-12
x1	0.93943	0.11035	8.5132	1.4732e-16
x2	-0.18049	0.17022	-1.0603	0.28944
x3	0.12428	0.17022	0.73013	0.4656
x4	-0.066942	0.11034	-0.60671	0.54428

AR(4) h=7 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.4755	0.19251	7.6646	7.6751e-14
x1	0.9302	0.11709	7.9441	1.035e-14
x2	-0.14985	0.18062	-0.82964	0.40708
x3	0.098839	0.18062	0.54723	0.58443
x4	-0.091876	0.11709	-0.78469	0.43296

AR(4) h=8 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.6846	0.20291	8.3023	7.3469e-16
x1	0.95285	0.12329	7.7287	4.8835e-14
x2	-0.18508	0.19017	-0.97322	0.33085
x3	0.12733	0.19017	0.66954	0.50342
x4	-0.13785	0.12328	-1.1182	0.26396

AR(4) h=9 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.9058	0.21323	8.9382	5.4086e-18
x1	0.94659	0.12941	7.3148	8.7362e-13
x2	-0.17261	0.19962	-0.86473	0.38755
x3	0.12694	0.19962	0.63593	0.52508
x4	-0.17543	0.12941	-1.3556	0.17575

AR(4) h=10 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	2.1324	0.22289	9.567	3.2525e-20
x1	0.95383	0.13511	7.0595	4.8616e-12
x2	-0.18046	0.20842	-0.86583	0.38694
x3	0.12931	0.20842	0.62043	0.53522
x4	-0.20968	0.13512	-1.5518	0.12126

AR(4) h=11 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	2.3634	0.23233	10.172	1.8863e-22
x1	0.95535	0.14066	6.7918	2.7864e-11
x2	-0.1837	0.21698	-0.84662	0.39756
x3	0.11205	0.21698	0.51643	0.60575
x4	-0.22373	0.14067	-1.5905	0.11227

AR(4) h=12 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	2.6031	0.24131	10.787	8.1418e-25
x1	0.95665	0.14591	6.5564	1.2365e-10
x2	-0.21322	0.22508	-0.9473	0.34389
x3	0.15107	0.22507	0.67121	0.50236
x4	-0.26874	0.14592	-1.8417	0.066031

Appendix 3

$p=$	0	1	2	3	4	5	6	7	8	9	10	11	12
1	796.52	780.58	766.39	766.33	768.06	770.06	770.92	771.82	773.81	775.75	777.74	779.17	781.17
2	776.53	765.60	756.46	754.73	756.43	758.40	759.09	760.31	762.31	764.13	766.10	767.59	769.52
3	776.33	763.94	755.71	754.99	756.47	758.43	759.27	760.58	762.57	764.37	766.36	767.80	769.74
4	774.43	763.45	754.66	754.32	756.14	757.99	758.86	760.06	762.03	763.89	765.89	767.44	769.36
5	775.16	763.14	754.45	753.79	755.68	757.67	758.90	760.07	762.06	763.96	765.94	767.54	769.49
6	774.94	761.55	753.41	752.73	754.69	756.65	758.34	759.84	761.82	763.75	765.75	767.42	769.35
7	776.66	763.55	755.41	754.72	756.68	758.63	760.30	761.84	763.82	765.75	767.75	769.41	771.35
8	778.35	765.54	757.38	756.68	758.63	760.59	762.27	763.82	765.80	767.71	769.71	771.38	773.31
9	778.28	766.43	757.79	757.15	759.11	761.09	762.77	764.42	766.41	768.15	770.10	771.75	773.66
10	780.23	768.43	759.77	759.15	761.11	763.09	764.77	766.42	768.41	770.15	772.10	773.74	775.65
11	781.27	769.87	760.78	760.31	762.24	764.23	765.88	767.59	769.58	771.38	773.36	775.16	776.98
12	782.80	771.00	762.31	761.97	763.90	765.89	767.55	769.24	771.24	773.04	775.02	776.97	778.50

Compact representation of the second dataset for different values of p

Point and Interval Forecast under ADL(6,3)

Month	Point Forecast	LB of Interval	UB of Interval
2025-3	5.3586	4.6036	6.1135
2025-4	5.3309	4.6128	6.4990
2025-5	5.3263	3.8904	6.7622
2025-6	5.3309	3.6658	6.6990
2025-7	5.3354	3.4833	7.1876
2025-8	5.3455	3.3606	7.3304
2025-9	5.3502	3.2527	7.4478
2025-10	5.3729	3.1696	7.5762
2025-11	5.3970	3.0897	7.7043
2025-12	5.4131	3.0129	7.8132
2026-1	5.4385	2.9437	7.9334
2026-2	5.4713	2.8858	8.0567

ADL(6,3) h=1 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	0.041594	0.075428	0.55143	0.58155
x1	1.1321	0.042178	26.841	1.9523e-103
x2	-0.23515	0.064133	-3.6666	0.00026867
x3	0.13859	0.064774	2.1395	0.032815
x4	-0.12206	0.063934	-1.9092	0.056739
x5	-0.018788	0.063104	-0.29773	0.76602
x6	0.072019	0.040909	1.7605	0.078858
x7	0.77641	0.17315	4.4841	8.8459e-06
x8	-0.88912	0.2749	-3.2343	0.0012896
x9	0.29192	0.17508	1.6674	0.095987

...

Omit 10 regressions from h=1 to 11 since they are all similar

...

ADL(6,3) h=12 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.7548	0.26293	6.6743	5.959e-11
x1	0.72771	0.14447	5.0371	6.3743e-07
x2	-0.12018	0.21965	-0.54714	0.5845
x3	0.07559	0.22184	0.34105	0.73319
x4	-0.088564	0.21195	-0.41785	0.67622
x5	0.1093	0.21602	0.50608	0.613
x6	-0.15351	0.14021	-1.0949	0.27402
x7	1.7692	0.59355	2.9808	0.0029995
x8	-1.3048	0.94198	-1.3851	0.16656
x9	0.84969	0.59991	1.4164	0.15722

Appendix 4

$p=,q=$	0	1	2	3	4	5	6	7	8	9	10	11	12
1	291.94	259.25	244.91	244.36	245.80	247.76	249.75	251.37	253.31	255.31	257.10	258.48	260.38
2	288.38	259.53	244.71	243.02	244.29	246.28	248.28	249.82	251.75	253.75	255.58	257.07	259.02
3	286.87	260.68	245.84	244.37	245.88	247.86	249.86	251.38	253.33	255.33	257.17	258.67	260.63
4	281.49	259.27	244.53	243.22	244.58	246.53	248.52	249.94	251.91	253.90	255.73	257.26	259.24
5	275.86	256.98	241.43	240.11	241.39	243.38	245.23	246.85	248.84	250.82	252.73	254.24	256.22
6	274.91	257.55	241.22	240.14	241.42	243.42	245.32	246.45	248.41	250.38	252.30	253.91	255.88
7	274.48	258.16	240.90	240.13	241.29	243.29	245.21	246.47	248.46	250.45	252.40	254.04	256.03
8	274.28	258.65	240.43	240.01	240.99	242.99	244.91	246.20	248.20	250.03	251.93	253.62	255.62
9	273.94	258.77	239.80	239.72	240.46	242.45	244.41	245.68	247.68	249.57	251.57	253.16	255.16
10	275.85	260.73	241.71	241.67	242.37	244.36	246.32	247.61	249.61	251.51	253.51	255.14	257.14
11	276.47	261.44	241.80	241.82	243.28	244.65	246.27	247.65	249.64	251.53	253.53	255.07	257.00
12	278.18	263.16	243.63	243.58	245.84	246.02	248.01	249.38	251.25	253.25	255.26	256.80	258.70

AIC value for Part 4

ADL(9,3) h=1 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	0.047988	0.051014	0.94068	0.34727
x1	1.0268	0.041951	24.476	1.0668e-90
x2	-0.035842	0.059487	-0.60251	0.54708
x3	0.04094	0.058519	0.6996	0.48447
x4	0.023447	0.058565	0.40035	0.68905
x5	-0.023894	0.058549	-0.51057	0.60985
x6	0.0038572	0.058553	0.065902	0.94748
x7	0.0022786	0.058504	0.038948	0.96895
x8	0.0043744	0.058501	0.074775	0.94042
x9	-0.060498	0.04075	-1.4846	0.1382
x10	0.63939	0.11049	5.7867	1.1913e-08
x11	-0.69059	0.17637	-3.9156	0.00010121
x12	0.16613	0.11394	1.4581	0.095987

...

Omit 10 regressions from h=1 to 11 since they are all similar...

...

ADL(9,3) h=12 Coefficients

	Estimate	SE	tStat	pValue
(Intercept)	1.5418	0.22765	6.7726	3.2418e-11
x1	1.2484	0.18336	6.8086	2.5744e-11
x2	-0.15263	0.26001	-0.58703	0.55742
x3	-0.070253	0.25575	-0.27469	0.78366
x4	-0.087093	0.25595	-0.34028	0.73378
x5	-0.087253	0.25589	-0.34098	0.73325
x6	-0.011124	0.25579	-0.043488	0.96533
x7	-0.029155	0.25567	-0.11403	0.90925
x8	-0.050136	0.25564	-0.19612	0.84459
x9	-0.12987	0.17825	-0.72863	0.46654
x10	1.8233	0.48326	3.773	0.000117872
x11	-1.308	0.777108	-1.6964	0.090378
x12	0.444114	0.4981	0.88564	0.3762