Time Series Analysis of California's Unemployment

Sathvika Parimi

Department of Probability and Statistics, University of California, Santa Barbara

PSTAT 174: Time Series

Professor Raya Feldman

March 30, 2025

Time Series Analysis of California's Unemployment

Abstract

Unemployment counts are a key metric for analyzing the job market and the broader economy, making accurate predictions of unemployment a valuable tool. Using data published by the Bureau of Labor Statistics, I analyzed California's unemployment trends from 2010 to 2018. After transforming the data with a square root transformation and differencing at lag 1 to achieve stationarity, I evaluated a few potential ARIMA models. The ARIMA(4,2,0) model provided the best fit, passing residual diagnostic tests, and accurately forecast unemployment counts in California over a twelve month period.

Introduction

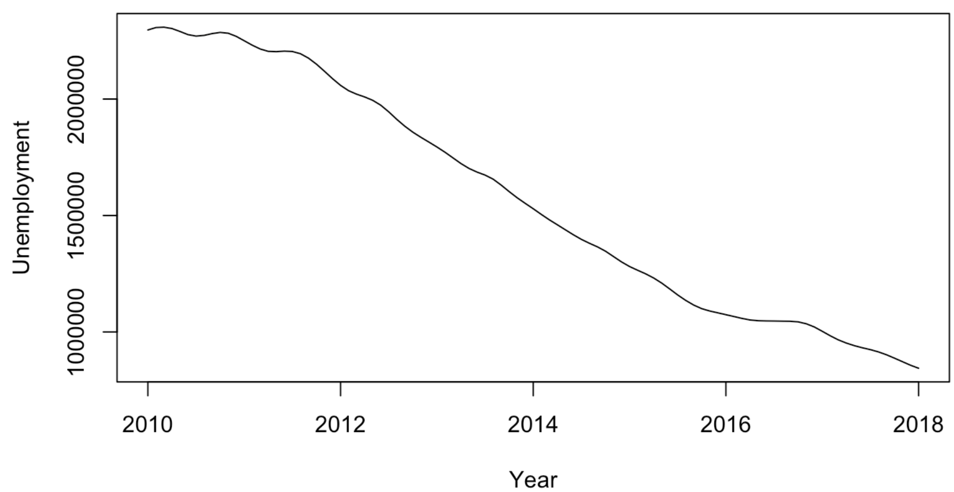
Unemployment is one of the most widely recognized indicators of economic health on both global and local scales. In particular, California's unemployment can provide information on job prospects for soon-to-be college graduates like myself who want to work in the state. Being able to forecast unemployment is a crucial tool for helping students decide whether they should seek employment or pursue further education. Using the Bureau of Labor Statistics's data on unemployment counts in America by state, I was able to subset the data for California only, specifically from January 2010 to 2018. I chose this time period because of the Great Recession that happened in 2007 and the COVID-19 pandemic of 2020, which severely affected unemployment. In R, I selected potential models using stationarity and differencing after transforming the data. I tested these ARIMA models using AICc and residual diagnostic tests to determine the best fit. I then used the best model to forecast data for the next 12 months and compared it to the real world unemployment statistics. Based on this, I can accurately predict unemployment in California post-graduation.

Modeling and Forecasting

Exploratory Analysis

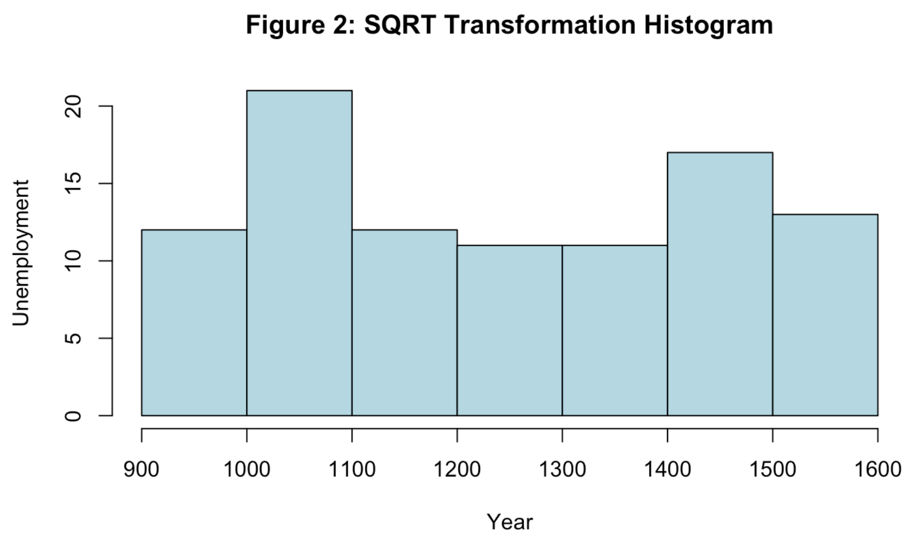
To preprocess the data, I subset it into training and testing groups, with the testing group including 12 months more than the training group. This will allow me to validate the accuracy of forecasting with the chosen ARIMA model. I then plotted the data as a time series and noticed an overall downward trend. From the time series, there does not appear to be any seasonality as shown in Figure 1.

Figure 1: Unemployment in California from 2010 to 2018



Transformation and Stationarity

To determine the best transformation, I used the Box-Cox function to calculate lambda. This function outputs a lambda value of 0.3838384, which suggests that a square root transformation should be used. Furthermore, I compared histograms of Box-Cox, log, and square root transformations and found square root to be the most symmetric.



Since we noticed a trend in the data, we detrend the data by differencing at lag 1. Even though there does not seem to be seasonality present, I compared the variance from differencing at lag 1 to differencing at lag 1 and 12. Since the former has a lower variance, we know that seasonality will not pose any issue, but the trend will. I plotted the ACF and PACF of the transformed data differenced at lag 1 and noticed that significant lags cut off before lag 1. This indicates that the data is most likely stationary.

Figure 3: ACF of Unemployment in California from 2010 to 2018

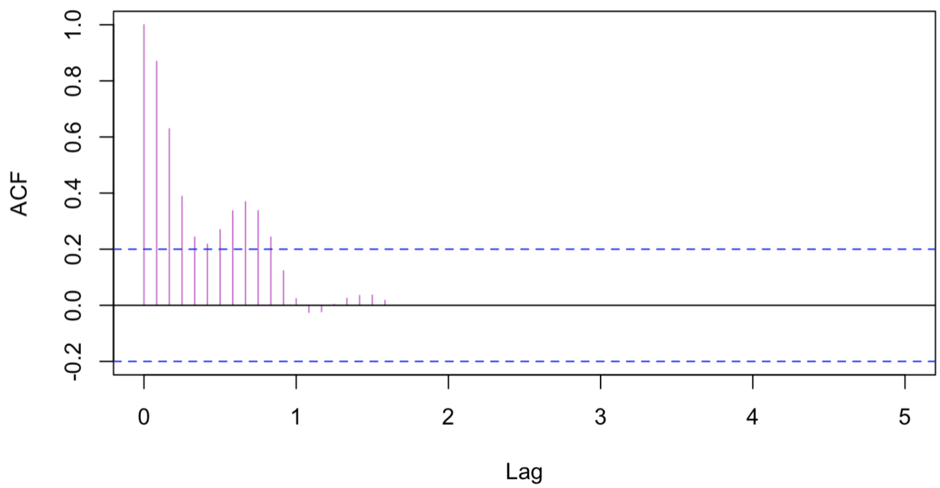
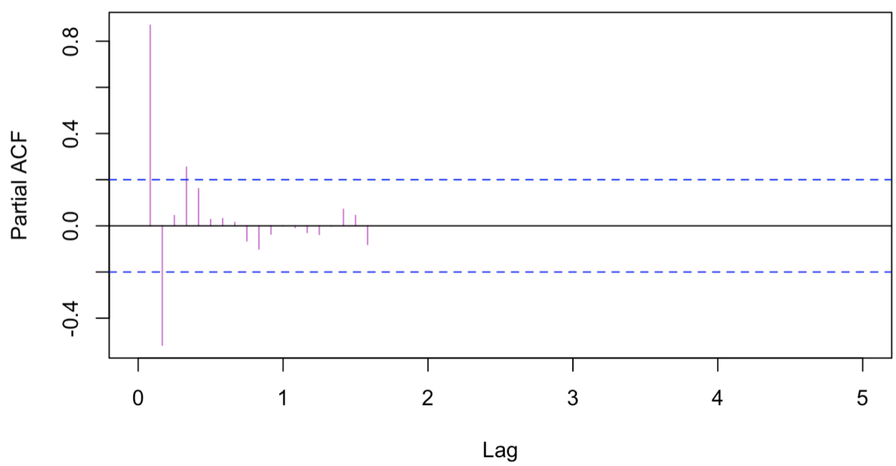


Figure 4: PACF of Unemployment in California from 2010 to 2018



Preliminary Selection

Based on the ACF and PACF plots, ARIMA looks like the most appropriate model. The value of p is most likely either 2 or 4 since there is a strong AR component, d is either 1 or 2 since we differenced at lag 1 to remove trend but might need to account for more differencing, and q is definitely 0 since there is no MA component (seasonality). Therefore, the potential models are ARIMA(2,1,0), ARIMA(4,1,0), ARIMA(2,2,0), and ARIMA(4,2,0).

Model Fit and Evaluation

To compare models, I used AICc scores and residual diagnostics. The ARIMA(4,2,0) model had the lowest AICc value, indicating that this model fits the data best. Furthermore, this was the only model not to fail the Shapiro-Wilk, Box-Pierce, Ljung-Box, McLeod-Li, or Yule-Walker tests. For each test, the p-value was greater than 0.05, indicating that the residuals are normally distributed, behave like white noise (no autocorrelation), and independently distributed. It also showed that the variance of the residuals is also white noise. The AICc and residual diagnostics both select for ARIMA(4,2,0), confirming this model is statistically sound and appropriate for this dataset. This model in algebraic form is:

(1) $\Delta_2 \sqrt{U_t} = (1 - 1.3482B + 1.1001B^2 - 0.3960B^3 + 0.2519B^4)Z_t$

Table 1: AICc and Residual Diagnostics for ARIMA Models

Model	AICc	Residual Diagnostics
ARIMA(2,1,0)	296.1931	Fails the Box-Pierce Test; Fails Ljung-Box Test
ARIMA(4,1,0)	214.1453	Fails the Box-Pierce Test; Fails Ljung-Box Test
ARIMA(2,2,0)	204.8874	Fails the Box-Pierce Test; Fails Ljung-Box Test
ARIMA(4,2,0)	203.4565	No Fails

Forecasting and Results

Using ARIMA(4,2,0), I forecasted unemployment for the next 12 months after the training period. When comparing to testing data, we see that the real data falls within the prediction interval for both transformed and untransformed data. This demonstrates strong model accuracy and reliability.

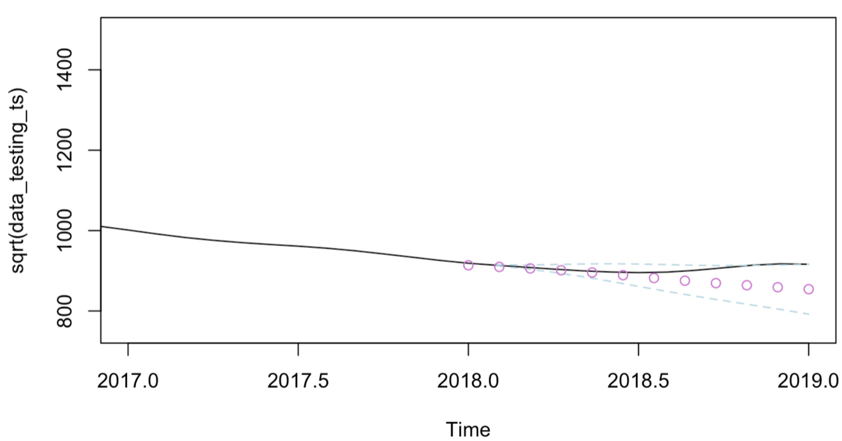


Figure 5: Transformed Unemployment with Forecasts (Zoomed)

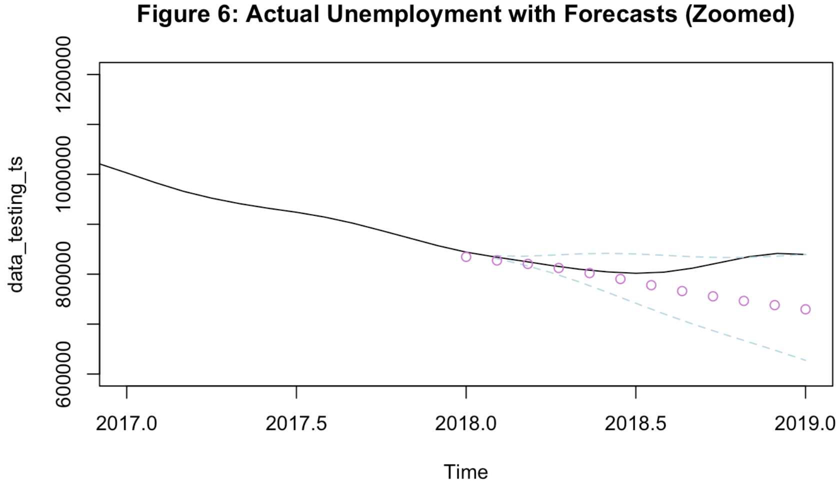


Figure 6: Actual Unemployment with Forecasts (Zoomed)

Conclusion

Out of all the tested models, ARIMA(4,2,0) proved to be the most effective for forecasting unemployment in California from January 2010 to 2018. This is supported by model selection criteria and diagnostic testing. The model successfully forecasted unemployment counts for the following year, which can be used to evaluate job market conditions. These forecasts can help new graduates decide whether to enter the workforce or continue towards higher education.

References

The Data Wrangler (2023, March 2). *Unemployment in America, per US state*. Kaggle. https://www.kaggle.com/datasets/justin2028/unemployment-in-america-per-us-state