Forecasting New Hires in U.S. Labor Market

A GUIDELINE FOR U.S. COMPANIES TO MAKE INFORMED HUMAN CAPITAL DECISIONS

University of Southern California | Marshall
TEAM | QIAO LIU, ANQI PAN, RUOXIAN JIA, YOURAN GUI, WEI TANG

Table of Contents

Executive Summary

In this report, we explored ways to predict the number of new hires in the labor market through time-series forecasting models. We also investigated the impact of the 2008 financial crisis on the labor market. Our analyses are based on data retrieved from the FRED, Federal Reserve Bank of St. Louis. The number of new hires dropped after the 2008 financial crisis, but there is an upward increasing trend after the recssion. Based on our champion ARIMA model, the number of new hires in the U.S. is predicted to be 68,985,000 in the next year, an 6% increase from the previous year. Companies could make use of our predictions to make informed hiring decisions and headcount plannings. However, forecasts sometimes fail to capture sudden changes in the labor market, like the financial crisis in 2008. Therefore, one should be cautious when interpreting forcasting predictions and making decisions based on forecasts.

I. Introduction and Objective

The labor market is a major component of the economy, and it reflects the overall economic performance. Through understanding the labor market, businesses could make well-informed decisions and strategies based on the current economic situation. Number of new hires in the labor market is a good indicator of the labor market, and it should be considered when making recruitment budgeting and workforce planning decisions for the next fiscal year.

In this report, we focused on using time-series models to predict number of new hires and also tried to understand the impact of the 2008 financial crisis on the labor market and economy.

Objective 1: ***Predict the total number of new hires in the labor market and help companies to make informed human capital decisions***

Objective 2: ***Understand the impact of the 2008 financial crisis and help companies to understand potential risks associated with forecasts***

We retrieved aggregated monthly new hire data from FRED, Federal Reserve Bank of St. Louis for our analysis. We also collected relevant economic indicators that are associated with labor market conditions, including monthly GDP, CPI, DPI, and interest rate to build models and predictions. Our model would not only be helpful for companies to make informed decisions, but also for job seekers to better understand the job market and make relevant preparations.

II. Data Source and Explanation

Monthly aggregated new hire data in the US from December 2000 to September 2018 are available from FRED, Federal Reserve Bank of St. Louis. Beside the new hire counts, we also considered the economy impact. We also retrieved monthly economy indexes from FRED, which includes CPI, DPI and 10-year treasury constant maturity rate. We also collected quarterly GDP data, but in order to remain consisted with other monthly variables, we divided GDP by three and assigned the same GDP value to different months in the same quarter.

Figure1 represents the number of monthly new hire from 2000.12 to 2018.9. We noticed that the number of new hire follows additive seasonality. The orange line shows the average of new hire before 2018, while the yellow line shows the average after 2018. There is a difference of new hire counts before 2018 and after 2018. Figure2 represents the number of yearly new hire from 2001 to 2017. By looking at this graphic, we observed that the number of new hire kept decreasing in 2009 and 2010.

Figure 1. Number of New Hire Over Time(2000.12 - 2018.09)

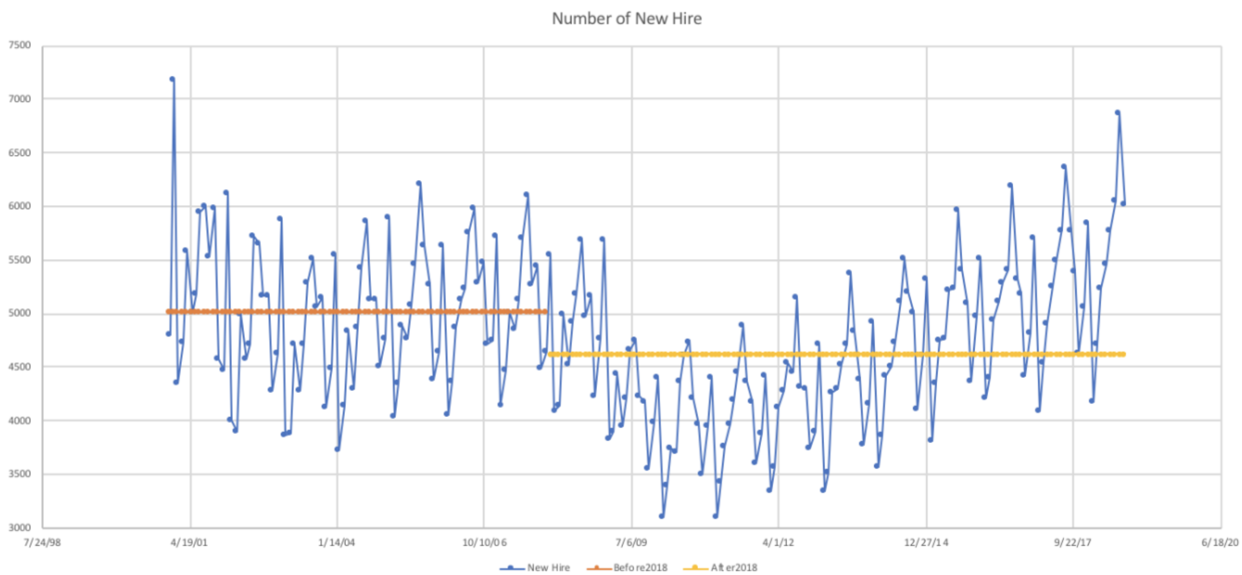
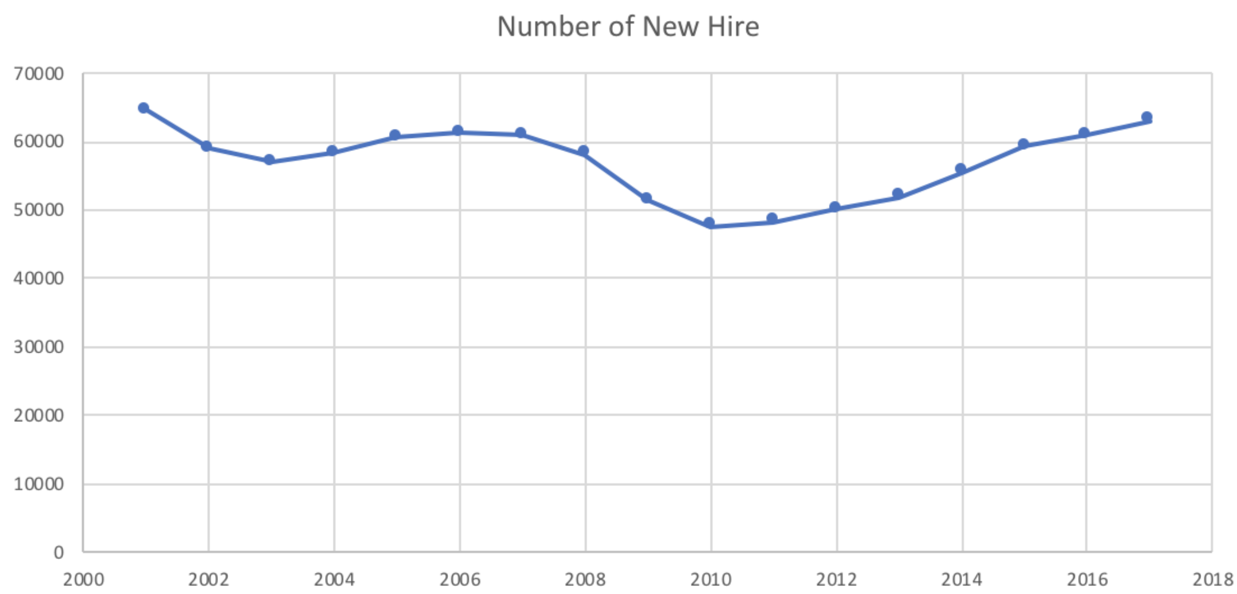


Figure 2. Number of New Hire by Year (2001-2017)



III. Methodology

a. Overview

Based on the dataset we have (data range: December 2000 – September 2018), we built three models. Each model has different training and testing windows. Model 2 and 3 serve to solve objective 1, while model 1 serves to solve object 2.

Table 1. Description of the Three Models

OBJECTIVE	MODEL	DATE RANGE
Objective 1	Model 2	*Training*: December 2000 to December 2016 *Testing 1*: Full year 2017 *Testing 2*: January 2017 - September 2018
	Model 3	*Training*: December 2000 – December 2017 *Testing 1*: January 2018 – September 2018
Objective 2	Model 1	*Training*: December 2000 - December 2007 *Testing 1*: Full year 2008

For Model 1, since there was a recession in 2008, we tried to exclude the impact of the recession in model 1. Under a scenario when there is no recession, we attempted to generate our predicted values. Then we compared our forecasts with the historical data from the testing period to recognize the impact of the economic crisis.

For Model 2, we generated two testing sets (next year, and the remaining period). From this model, we considered the recession and constructed a more robust model. This model will be able to adjust itself under some similar circumstances (e.g., a new recession hit the economy).

The difference between model 2 and model 3 is that we used more data to train model 3 to see whether it will generate a better result.

After dissecting different time windows for each model, we forecasted the total number of new hires from two perspectives: macroeconomic indicators and new hires historical data. We used various methods based on these two perspectives.

b. Macroeconomics Analysis

We assumed that several macroeconomic indicators are closely related to our prediction on the total number of new hires. We used four different indicators (interest rate, CPI, DPI, GDP) in addition to new hires data.

Multi Linear Regression:

For the linear regression model using macroeconomic indicators, we realized that many indicators had very high p-values and were not significant.

Table 2. Macroeconomics Indicators Coefficients

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3184.90	7305.99	0.44	0.66	-11354.50	17724.28
Interest rate	264.87	163.39	1.62	0.11	-60.28	590.02
CPI	12.20	78.07	0.16	0.88	-143.16	167.56
DPI	-1.51	0.89	-1.70	0.09	-3.28	0.26
GDP	2.90	1.94	1.49	0.14	-0.96	6.77

c. Time-series Analysis

Due to the fact that macroeconomic indicators are not reliable in terms of predicting new hire rates, we tried to take a closer look at the historical data. We used a new dataset that only included new hire values and time periods and used different methods including Regression, Seasonal Naive, Smoothing and Arima. Arima had the lowest testing MAPE in all three models and was the best model in this case.

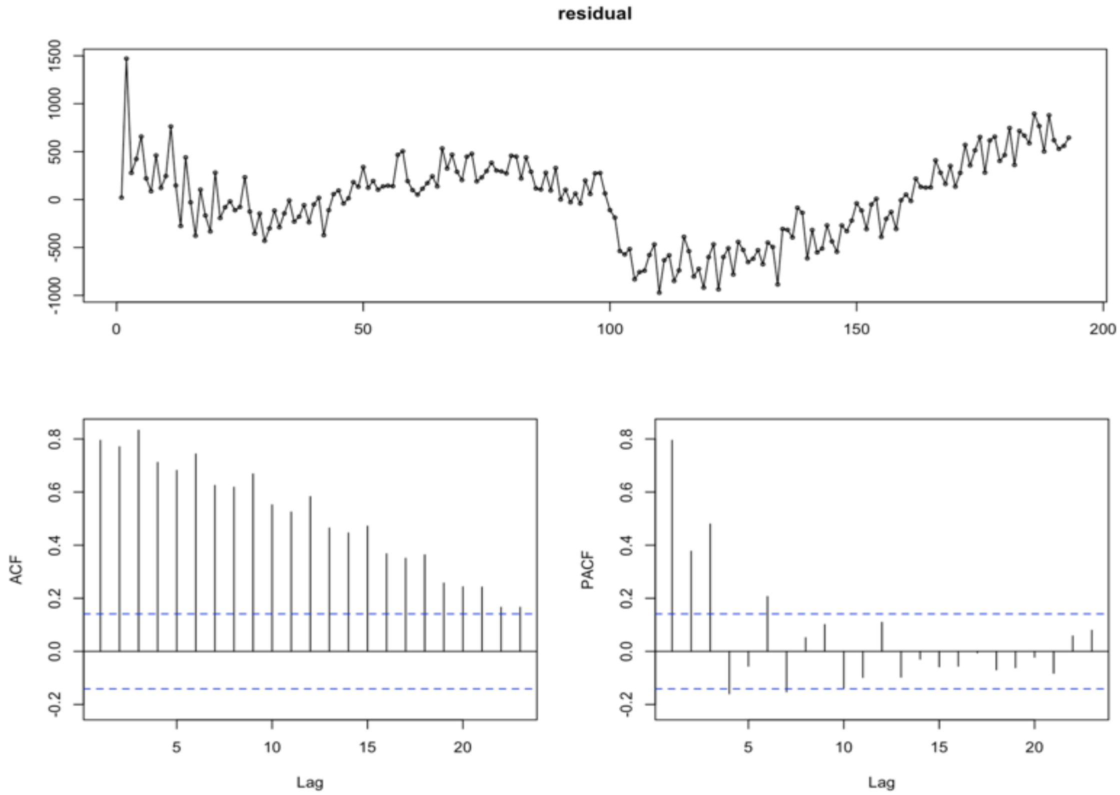
1. Regression

To understand the historical data better, we added the following dummy variables:

- **Trend**: we assigned “1” to the initial point of the time period and filled the series automatically.
- **Monthly Variables**: We created 11 dummy variables to capture the monthly trend.

Then we ran a regression analysis with residuals based on different models. From the residual plot, we could tell the residuals followed a clear pattern and were not random. Thus, we considered to add lag variables. We then used ACF and PACF to determine how many lags to include in our regression analysis. As shown below, we considered adding 3 lags in the regression analysis.

Figure 3. Autocorrelation and Partial Autocorrelation Function



Next, we created three more lag variables and ran a new regression analysis. The residuals seemed random and our prediction seemed more powerful at this point.

Figure 4. Residuals

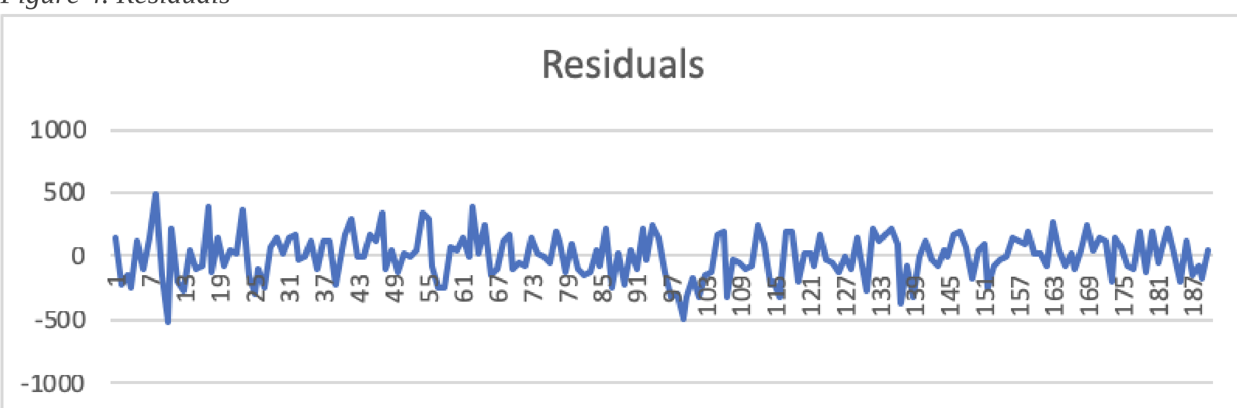
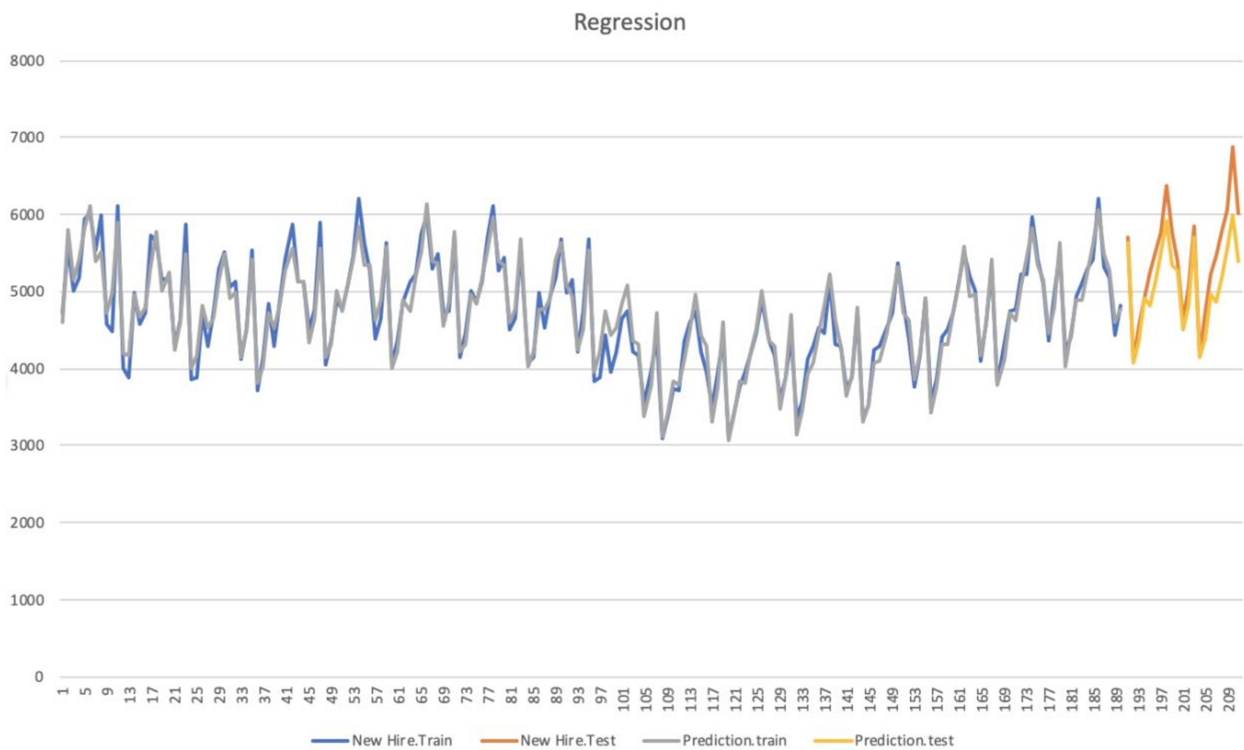


Figure 5. Predictions of Regression



In addition, we took the coefficients from the regression analysis to forecast the values from both testing and training periods. In the end, we calculated the MAPE, RMSE values for the training period and testing periods.

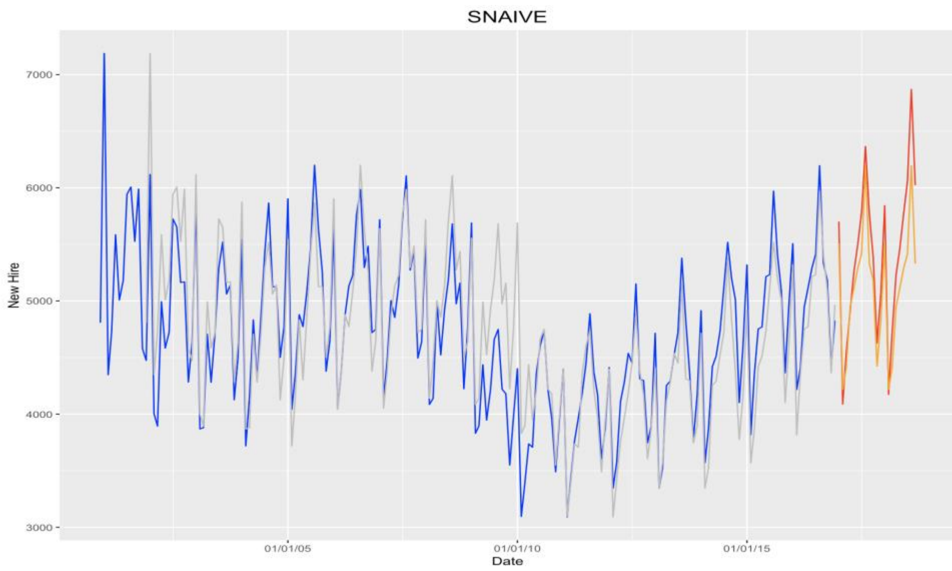
2. Seasonal NAÏVE

The seasonal naive method accounts for seasonality by setting each prediction to be equal to the last observed value of the same season.

In this case, we applied snaive() function in r to build the snaive model, and calculated the MAPE, RMSE values for the training period and testing periods by using accuracy() function.

Then, We conducted residual diagnostics to check the residual with checkresidual() function in r to identify the quality of this model

Figure 6. Predictions of Seasonal Naive



From Figure 7, we could see the Residuals of SNAIVE model do not resemble white noise which means there are patterns that seasonal naive model was unable to capture. Also there seems to be outliers and it would be interesting to see what happened during months of extreme outlying points.

Figure 7. Residual Diagnostics of Seasonal Naive

3. Smoothing

Smoothing is a method to covert non-stationary time series data to stationary data by adjusting the patterns existing in the data. The Smoothing model we used was ETS(A, Ad, A) - additive noise(first A), additive damped (second A) and additive seasonality (3rd A). We obtained these values through the ETS() function in r by setting lambda='auto' to select the best model for transformed data. The residuals look like random but still have some outliers.

Figure 8. Predictions of Smoothing

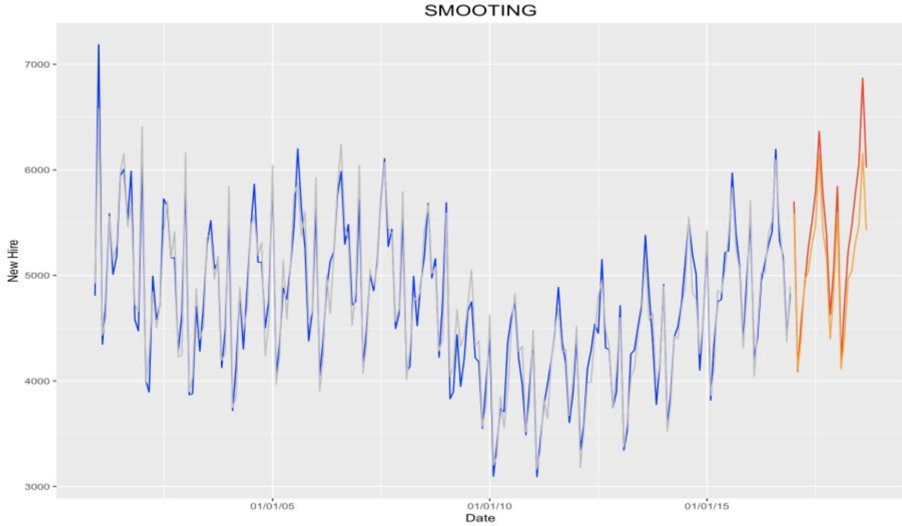
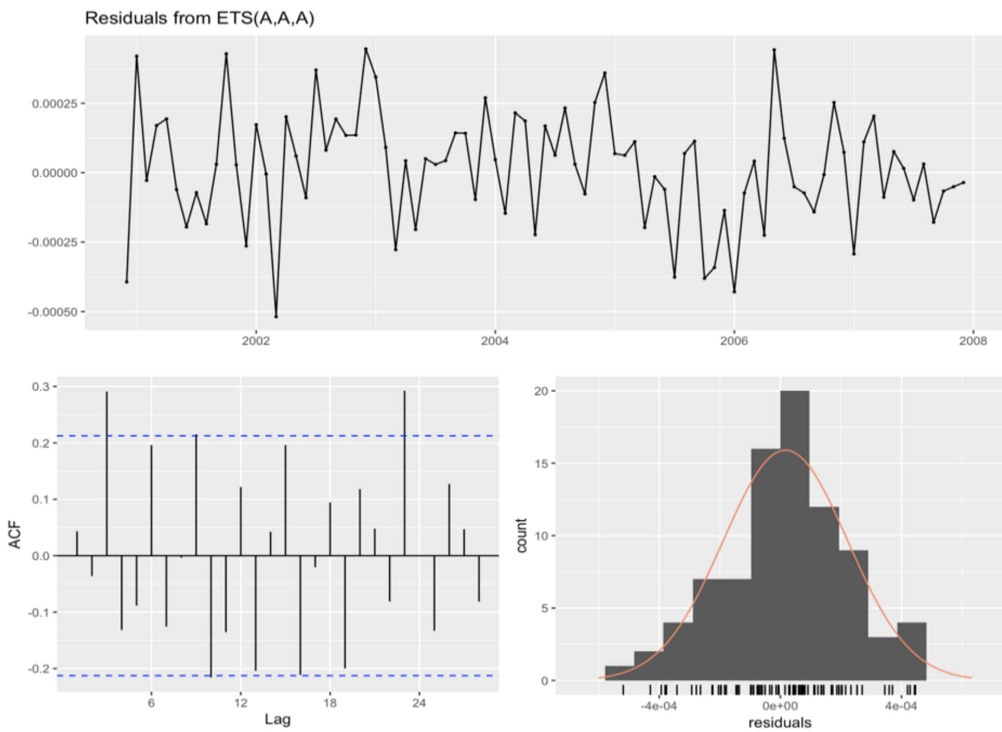


Figure 8. Residual Diagnostics of Smoothing



4. ARIMA

ARIMA model is a generalization of an autoregressive moving average (ARMA) model. ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step can be applied one or more times to eliminate the non-stationarity. The ARIMA model we used in there was ARIMA(2,1,2)(1,1,2)[12]. We obtained these values through the auto.arima() function in r

Figure 9. Predictions of ARIMA

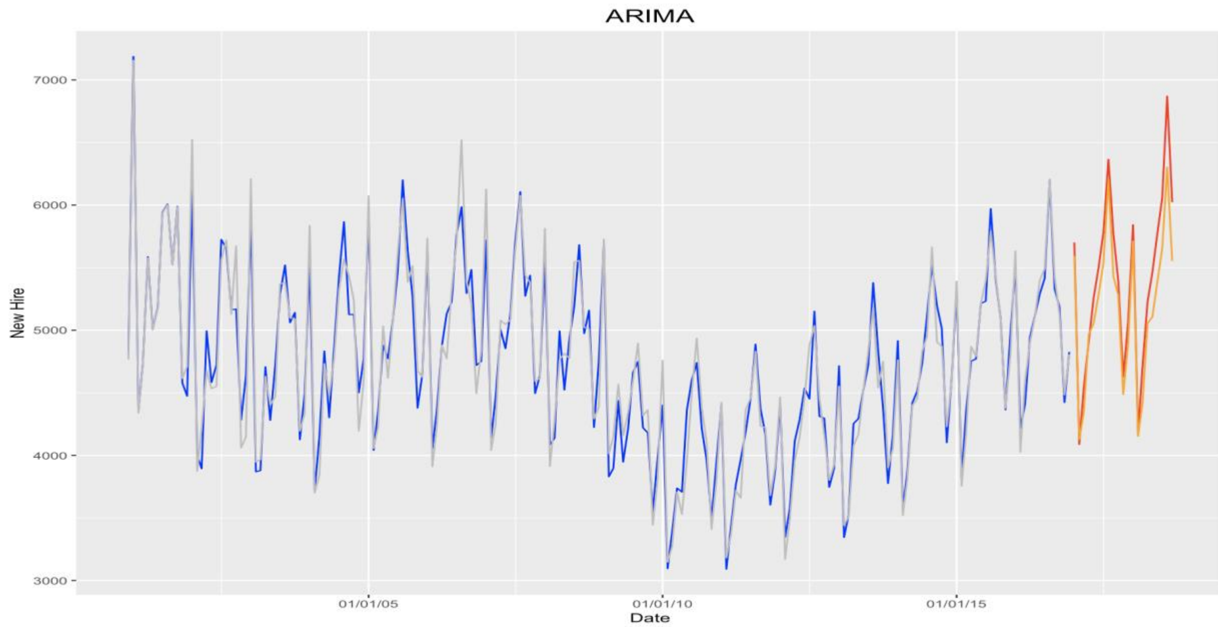
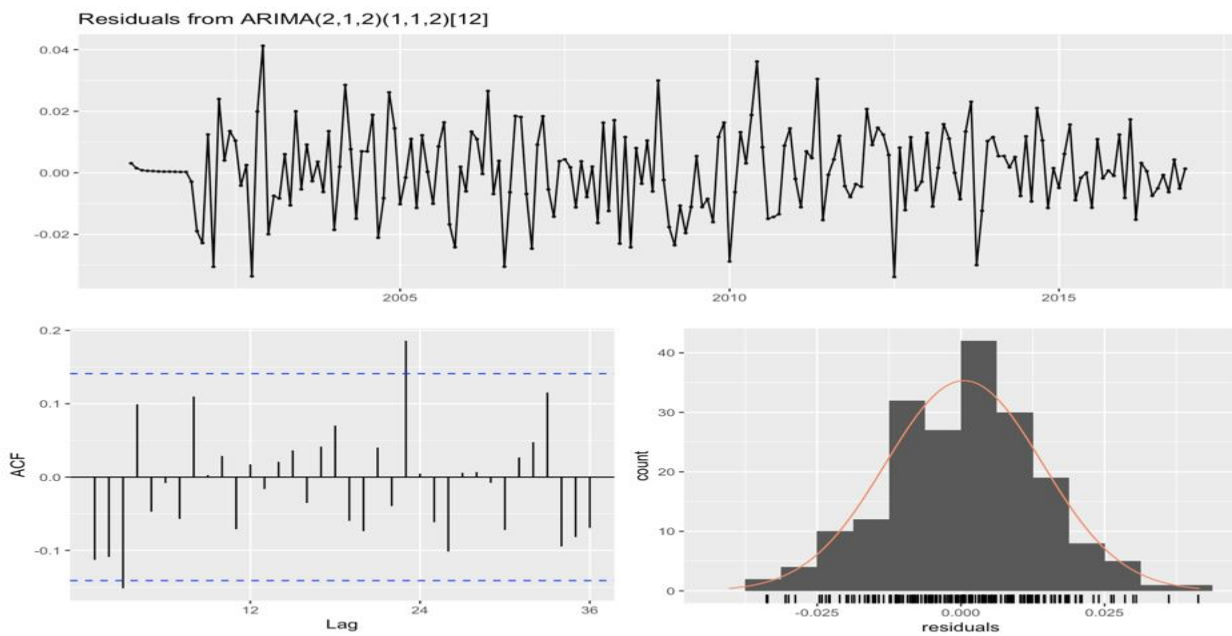


Figure 10. Residuals of ARIMA



Residuals look like white noise, which looks like the stationary uncorrelated sequence of numbers.

d. Combined Macroeconomics and Time-series Data

After using different methods from two separate perspectives, we tried to run a regression analysis using both macroeconomic indicators and historical data. We used lm() function in R to run the multiple regression and calculated MAPE, RMSE, etc.

We also used vif() function in R to test out the multicollinearity in our model. We observed that many variables have multicollinearity, which might also affect our results. The data exploration on macroeconomic indicators can be found in the excel worksheet and R code.

Figure 11. The multicollinearity of Combined Macroeconomics and Historical Data

```
> car::vif(lmodel)
       Trend           M1           M2 
  256.489512     3.368245     7.294170 
          M3           M4           M5 
    3.205946     5.002442     6.116801 
          M6           M7           M8 
    3.890310     3.759153     5.235893 
          M9          M10          M11 
    6.471240     3.093558     3.120314 
        Lag1         Lag2         Lag3 
   12.864980    11.237333     9.860457 
         CPI          DPI          GDP 
   96.754288   297.617325   290.084258 
Interestrate 
    7.114322 
```

IV. Forecast and Evaluation

Objective 1*: Predict the total number of new hires in the labor market*

We selected ARIMA as the final model based on its smallest validation RMSE and MAPE. Model comparison is shown in Table 3.

Table 3. Model Comparison of New Hires Prediction

Model	Training		Validation	
	RMSE	MAPE	RMSE	MAPE
Seasonal Naïve	336.16	5.54	353.45	5.32
Smoothing	180.18	3.08	335.53	5.02
Regression	172.98	2.99	392.34	5.58
ARIMA	174.41	2.92	269.06	4.07

Then, we reconstructed ARIMA model (shown in Figure 9) based on whole dataset to predict the total number of new hires in the US labor market for the next year in the future. The total number of new hires is predicted to continuly increase with a certain seasonality pattern.

Table 4. Predicted Total Number of New Hires in the Labor Market (Level in Thousands)

Month	Forecast	Lower 95%	Upper 95%
Oct-18	5764	5355	6204
Nov-18	4950	4556	5377
Dec-18	5385	4913	5902
Jan-19	6256	5606	6980
Feb-19	4528	4024	5093
Mar-19	4972	4384	5638
Apr-19	5559	4848	6371
May-19	5747	4978	6631
Jun-19	6065	5216	7049
Jul-19	6342	5410	7431
Aug-19	7119	6037	8392
Sep-19	6298	5303	7475

Objective 2: Understand the impact of the 2008 financial crisis

In order to understand the impact of the recession on employment, we built models to predict the 2008 new hires and compared it to the actual new hires. Similarly, we developed four models and chose the best one with smallest validation RMSE (shown in Table 5). ARIMA appears to be our final model.

Table 5. Model Comparison of Recession Impact Evaluation

Model	Validation	
	RMSE	MAPE
Seasonal Naïve	288.62	5.16
Smoothing	265.31	4.51
Regression	301.46	5.1
ARIMA	244.17	4.43

Table 6. Actual and Predicted New Hires in Year 2008

Month	Actual	Forecast	Residuals	Lower 95%	Upper 95%
Jan-08	5555	5656	-101	4387	7638
Feb-08	4085	4064	21	3258	5245
Mar-08	4138	4379	-241	3443	5804
Apr-08	4993	4912	81	3762	6753
May-08	4523	4840	-317	3679	6725
Jun-08	4930	5092	-162	3804	7256
Jul-08	5188	5634	-446	4100	8355
Aug-08	5682	6008	-326	4282	9204
Sep-08	4974	5247	-273	3799	7843
Oct-08	5159	5370	-211	3839	8185
Nov-08	4224	4491	-267	3289	6594
Dec-08	4758	4624	134	3345	6921
Grand Total	*58209*	*60317*	*-2108*	*44987*	*86524*

The great recession of 2008 had a negative impact on employment. Shown in Table 4, in 2008, the total number of new hires decreased by 2,108,000, which is 4% of the total new hires of that year. Nevertheless, the number is still within the predicted lower and upper bound.

V. Conclusion and Recommendation

Overall, there is an upward trend of new hires since recession. We observed an annual recruitment cycle: there exist a peak around August and a small peak around January, followed by the lowest point around February. Based on our research, the main reasons behind the recruitment cycle are graduation season and company budget.

For companies, HR managers should always be aware with the labor market demands and recruitment cycle. Hence, they are able to build recruitment strategies in advance. Next year, the total number of new hires in U.S. is predicted to be 68,985,000 employees, increased by 6% than the previous year. The labor market seems to be vigorous and competitive. A competitive wage rate might not be the only determinant factor for job seekers. In this case, a good HR manager should manage to build a friendly working environment in addition to a reasonable wage rate, because job seekers tend to be selective.

Sometimes, the forecast fails to capture a sudden change in the labor market, most likely caused by the macro-economic factor. This is also why there is a huge difference between the actual and predicted number of new hires during the years of recession around 2008-2009. Therefore, one should be careful with predictions when making decisions based on forecasts.

VI. Reference

1. U.S. Bureau of Labor Statistics, Hires: Total Nonfarm [JTUHIL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/JTUHIL, December 3, 2018.

2. U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIAUCSL, December 4, 2018.

3. U.S. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP, December 4, 2018.

4. U.S. Bureau of Economic Analysis, Disposable Personal Income [DPI], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DPI, December 5, 2018.

5. Board of Governors of the Federal Reserve System (US), 10-Year Treasury Constant Maturity Rate [GS10], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GS10, December 5, 2018.