Article

Forecasting the Unemployment Rate in the United States

Skyline Business Journal
20(1) 25-33, 2024
https://doi.org/10.37383/SBJ200102
www.skylineuniversity/Journal

Jesús Constant
Business School, Tecnologico de Monterrey, Mexico

Jaime Ruiz
Business School, Tecnologico de Monterrey, Mexico

Martha Martínez
Business School, Tecnologico de Monterrey, Mexico

Orlando López
Business School, Tecnologico de Monterrey, Mexico

Jesús C. Téllez*
Business School, Tecnologico de Monterrey, Mexico
Cuauhtemoc.tellez@tec.mx
*Corresponding Author

Abstract

The U.S. unemployment rate serves as a critical economic indicator, playing a pivotal role in assessing economic well-being and societal welfare and attracting the attention of both national and international corporations. The present study takes into account the underlying causes, effects, and potential remedies associated with the unemployment rate. The study also aims to construct a robust forecasting methodology for the unemployment rate indicator, with the help of diverse set of variables such as the S&P 500 index, inflation rates, currency fluctuations, and other associated factors. Several statistical models, including logistic regression OLS (Ordinary Least Square), neural networks, k-nearest neighbours, and supervised machine learning algorithms, have been examined. In the present study, the logistic regression model achieved a relatively high accuracy of 67%, which is compatible with the inherent volatility of the unemployment rate.

Keywords

Unemployment rate, machine learning, confusion matrix, forecasting

1 . Introduction

The percentage of the labor force in a region that is jobless yet actively seeking for work is referred to as the unemployment rate, and it is an essential figure in economic analysis. This fundamental economic indicator forms a complex framework with numerous other economic variables, indicating not just the well-being of a nation or region, but also the way they interact. Businesses keep a careful eye on the unemployment rate due to its many facets.

Market Labor Conditions: Companies can learn about the current state of the labor market by interpreting the unemployment rate. This affects the workforce planning, talent acquisition tactics, and hiring decisions.

Consumer Spending Levels: A rising unemployment rate often leads to reduced spending. Corporations adjust their sales forecasts and marketing strategies accordingly.

Revenue Impact: Unemployment affects consumers' purchasing power, directly impacting corporate revenues. Companies must adapt to the changing demand patterns.

Business Sales Cycles: Unemployment fluctuations influence business cycles. Corporations anticipate shifts in demand and adjust their production and inventory levels accordingly.

Financial Stress: High unemployment places financial stress on individuals, affecting their ability to make purchases. Corporations consider this when assessing market dynamics.

Companies can manage costs by using the unemployment rate as a reference. For example, during downturns, companies which sell non-essential products could see a decline in demand. Consumer budget tightening brought on by rising unemployment rates results in lower expenditures on luxuries.

Economists monitor the rate of unemployment with great care.

Predicting Economic Cycles: Unemployment rate trends provide insight into economic cycles. Economists analyze these patterns to understand labor market health and anticipate economic shifts.

Labor Health Assessment: Monitoring unemployment helps economists assess the overall health of the labor force. It also provides policy recommendations and economic research.

Governments and central banks, among other authorities, base their choices on the unemployment rate.

Interest rates and fiscal policies: When making decisions regarding interest rates and fiscal policies, policymakers consult the data on unemployment. For instance, expansionary policies may be used to promote economic growth during times of high unemployment.

In result, there is a lot of potential for the unemployment rate prediction model. Forecasting unemployment patterns accurately can assist firms in preparing for swings in sales, direct economic policy, and guide government budgeting. In a changing economic environment, adaptation and resilience can be guaranteed by efficiently handling these changes.

2 . Literature Review

2.1 Method: Long Short-Term Memory

Yurtesever's (2023) long short-term memory is a kind of recurrent neural network, which is an artificial intelligence that receives information continuously and then uses that information to improve its own answers. This type of technology was developed by Hochreiter and Schmidhube in 1997, and he created this method to address the problem of diminishing gradients in traditional neural networks. LSTM employs memory gates and three gates: input, output, and forget gates, which enable the technique to specifically maintain or discard information, permitting it to effectively maintain long-term dependencies. The framework of the LSTM algorithm includes memory cells that store important data over time and are able to control the flow of information and keep it updated. The main advantage of LSTM is to manage carefully information to keep or discard to capture long-term dependencies and prevent information loss and also, making them highly effective for sequential data processing.

2.2 Forecasting the United States Unemployment Rate (Using Recurrent Neural Networks with the application of Google Trends Data)

The Kundu and Singhhania (2020) method was applied to improve the future forecast of the unemployment rate in the United States with the help of online search data. It is important to have a good prediction of this rate, particularly in the short term. To tackle this problem and have a good prediction model, LSTM was the best option using the Google Trends query share for certain keywords. This method works with the correlation between many keywords and different aspects of the United States economy. To save information, they used data from January 15, 2019. With that information adding another method was necessary, in this case was the VAR model combining the official unemployment claims series with the search trends for the keyword 'job offers' taken from Google Trends and an LSTM model with only the Google trends time series data for the complete set of identified keywords. The LSTM model significantly improves all the capacities of the VAR method.

2.3 Time series analysis and forecasting of the unemployment

According to Törnqvist (2021), it is important to understand that it is not necessary to find a predicted model that obtains the highest accuracy, because sometimes the model can give you a really good result, but if you analyze, you can see that the results are not reasonable. For example, in the case of Swedish data, an Arima (3,1,1) model may have been beaten by the Holt-Winters nonseasonal smoothing model, even though the Holt-Winters nonseasonal smoothing model has a better percent accuracy, and the Arima model could predict this rate with greater effectiveness. However, in the case of Italy, we see the opposite; the Holt-Winters nonseasonal smoothing model was the most effective in predicting the unemployment rate. Therefore, in conclusion, we need to understand the context and analyze the model to determine the best option.

2.4 Lessons for Forecasting Unemployment in United States of America, Application of Flow Rates, Mind the Trend

Meyer and Tasci (2015) examined the effectiveness of various approaches for forecasting the U.S. unemployment rate, highlighting two key lessons. First, leveraging unemployment flow rates, such as those in the FLOW-UC model, significantly improves near-term unemployment rate forecasts, particularly around business cycle turning points. Second, it is crucial to consider the long-term trend in the unemployment rate, especially during periods when it deviates from the natural rate.

Forecasting models are capable of predicting the unemployment rate and more accurately reflect the dynamics of the labor market by incorporating data on flow rates and long-term patterns, especially during economic recovery. The findings of this study highlight the significance of these two variables in raising the precision of unemployment rate estimates and facilitating better economic judgment.

2.5 U.S. unemployment rate prediction (Time-series model)

Using a variety of time-series modeling approaches, Zou's 2024 research aimed to forecast the unemployment rate in the United States (Zou, 2024). The study emphasized the importance of precisely predicting the unemployment rate to make well-informed decisions and implement effective policy measures. Frequently utilized techniques include the seasonal effects-incorporating ARIMA model, the linear and nonlinear pattern-capturing ETS model, and the nonlinear relationship-modeling Neural Network Autoregression (NNAR) model.

To enhance forecasting effectiveness, hybrid models, such as ARIMA-GARCH, have also been utilized to combine the advantages of various methodologies. The study also highlights the impact of important economic indicators on the unemployment rate, highlighting the complexity of the factors influencing its prediction. These indicators include GDP, employment statistics, and consumer spending. This study highlights the value of sophisticated time-series modeling methods and economic indicators in predicting the U.S. unemployment rate, providing valuable insights for decision-makers.

2.6 Forecasting the U.S. Unemployment Rate (Another Look)

The usefulness of using leading economic indicators and weekly unemployment claims data as inputs to forecast the U.S. unemployment rate was examined by Xiao et al. (2022). They discovered that both these data sources are statistically significant inputs in forecasting models, indicating that they contain important information for predicting changes in unemployment.

Furthermore, the authors compared the performance of a seasonal ARIMA (SARIMA) model against a straightforward no-change benchmark, and discovered that the SARIMA model outperformed the benchmark across all forecasting horizons, indicating that more complex time-series models can result in more accurate forecasts of the unemployment rate. Furthermore, the study revealed that SARIMA and transfer function models were particularly effective during the 2008-2019 Global Financial Crisis period, outperforming a null, no-change forecast and suggesting their ability to capture the complexities of the labor market during economic downturns.

2.7 Forecasting US Unemployment Rate (Job Openings Index)

In order to predict the US unemployment rate, Huang (2015) concentrated on using the Bureau of Labor Statistics' Job Openings Index as a leading indicator (Huang, 2015). Huang used multivariate vector autoregressive models and integrated autoregressive moving average with external variables models, using the Job Openings Index and beginning unemployment insurance claims as external variables. The results show that the Job Openings Index can greatly improve US unemployment rate forecasting accuracy, beating models that rely on benchmark models or initial claims for unemployment insurance.

Huang came to the conclusion that forecasting models can produce more accurate estimates of the US unemployment rate by using the Job Openings Index. The Job Openings Index is a useful leading indicator for predicting the US unemployment rate, as Huang's research showed. Huang demonstrated that it is feasible to increase the precision of estimates of the unemployment rate by integrating this index into forecasting models, offering important information to economists and policymakers.

2.8 Forecasting Unemployment Rates in USA (Box-Jenkins Methodology)

The Box-Jenkins approach was used by Dritsakis and Klazoglou (2018) to predict unemployment rates in the United States.

From January 1955 to July 2017, they gathered monthly data on the unemployment rate, examined its stationarity, performed ACF and PACF studies to find a suitable ARIMA model, estimated the model's parameters, used residual analysis to confirm the model's adequacy, and ultimately used the model for forecasting. Notwithstanding obstacles, including the intricacy of economic events and problems with data quality, the methodology offers a methodical way to comprehend and forecast unemployment rates in the United States.

Dritsakis and Klazoglou discovered that the Box-Jenkins technique performed well in predicting unemployment rates in the United States. The chosen ARIMA model yields trustworthy projections for future unemployment rates with accurately reflecting the dynamics of the unemployment rate time series. The accuracy of the projections was influenced by the methodological approach, thorough assessment of data features, and model suitability of the methodology, indicating the usefulness of the Box-Jenkins methodology in economic forecasting.

2.8 Predicting changes in the U.S. Unemployment Rate at Covid-19 (Time Series)

In order to produce a novel LSTM-GRU hybrid model, the authors of the paper Gao et al. (2023) combined LSTM and GRU architectures in a hybrid deep learning technique. The purpose of this model is to analyze monthly data on the unemployment rate from January 1983 to May 2022 to identify both short- and long-term patterns in the data. The hybrid model's LSTM component handled long-term dependencies well, while the GRU component was used to identify short-term patterns in the data. Researchers have aimed to improve the overall prediction accuracy by combining the capabilities of both the models through the integration of these two architectures.

We used several important performance measures, such as the Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE), to assess the forecasting accuracy of the LSTM-GRU hybrid model. They evaluated the hybrid model's overall and COVID-19 pandemic-specific performance against those of the standalone LSTM and GRU models. The goal of the evaluation was to determine whether the hybrid technique might perform better than the individual GRU and LSTM models, particularly when it came to detecting the complex and nonlinear patterns found in data on unemployment rates, particularly during the pandemic.

2.9 Short-term forecasting of the US unemployment rate

Maas used MIDAS regression models in the Benedikt Mass 2019 study in order to predict the US unemployment rate (Mass, 2019). In addition to standard variables like weekly initial jobless claims, he employed a novel predictor, a weekly Google search index for "unemployment". The goal of this strategy is to improve the forecast timeliness and accuracy. Diffusion index variations, which fused the Google search index with other indicators, such as monthly latent components indicating the status of the economy, were also employed by Maas.

Maas's research was practical in that it was intended for real-time forecasting. He compared short- and long-term forecasts to evaluate the forecast performance over various time frames. By integrating diverse data sources and innovative methods, Maas's research offers a comprehensive framework for short-term forecasting of the US unemployment rate with potential implications for economic policymaking and market analysis.

3 . Data and Methodology

The first step was to select variables that could explain the movement of the unemployment rate, considering that all variables should be within the same time interval and periodicity to build up the database. Thus, this study considers 13 variables from January 2001 to October 2023 monthly frequency. The following are the variables included in the database downloaded from FRED economic data:

1.Urate: Unemployment rate published in the FRED economic data.
2.GDP: Measure the production of goods and services in the country.
3.S&P500: Market capitalization weighted index of the 500 leading publicly traded companies in the U.S.
4.Inflation: Indicator of price increases in the USA States
5.ACT: Level of labor activity that represents income spent on discretionary purchases and the current level of activity in labor markets.
6.HR WAGE: Unweighted Median Hourly Wage Growth.
7.Interest rate: Federal funds' effective rate.
8.USD/MXN: Value of the US currency against Mexican currency.
9.CPI: Consumer Price Index, considering basic food basket prices.
10.RUSSELL 2000: Contemplates 2000 small and medium-sized companies listed in the RUSSELL 3000 index.
11.HOUSING: Purchase-Only House Price Index for the U.S.
12.INDS PROD: Or Industrial Production as an indicator of the growth of the US economy owing to infrastructure investment.

The data were transformed to log-returns, as shown in Expression (1):

$$rets = log\left(\frac{P1}{P0}\right) \qquad (1)$$

where and are the current and previous values, respectively. Only interest rates were transformed to first differences, as in Expression (2):

$$\Delta P = P1 - P0 \qquad (2)$$

Then, ret-logs and first difference variables were transformed to a categorical variable to represent up and down movements, where 1 and 0 are related to bull and bear movements, respectively.

Forecasting was performed by applying Logistic Regression, Support Vector Machine (SVM), a Neural Network, and KNN models. All estimations were performed using Orange Release version 3.36.2. To compare the accuracy of the predictive models, ROC Analysis was performed, and the Confusion Matrix was estimated.

4 . Analysis of Results

Based on the logistic regression model, there is an intercept of 0.1992, which indicates a direct impact of ACT, HR WAGE, and INT. The RATE and CPI variables, with HR WAGE as the main variable with a coefficient of 0.9381, the second highest is INT. RATE with a coefficient value of 0.0740. Notably, HR WAGE can considerably affect the Unemployment Rate. In the indirect impact, the main factor is INFL. where -0.2311 is the coefficient of the variable. In this case, when inflation in the US economy increases, the unemployment rate decreases, and vice versa. The test and score reflect that the model with the highest accuracy is the logistic model with 67% accuracy. Other models were close to this value, with 63.3%, 64.9%, and 53.5% being the SVM, Neural Network and KNN models, respectively.

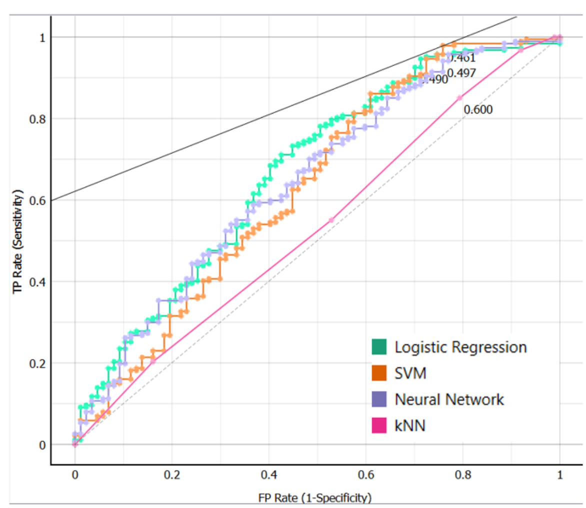
Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.670	0.719	0.700		0.719	0.276
SVM	0.633	0.741	0.686	0.764	0.741	0.343
Neural Network	0.649	0.715	0.699	0.698	0.715	0.253
KNN	0.535	0.646	0.609	0.600	0.646	0.071

Authors' calculations.

Predicted

Actual	0	1	Σ
0	171	16	187
1	61	26	87
Σ	232	42	274

Authors' calculations.



5 . Conclusion

Businesses, financiers, politicians, and economists closely follow the unemployment rate. An accurate prediction model for this rate would mean that all the interested groups mentioned would be able to make more accurate decisions, thus developing into a better-managed society.

The efforts gathered in this study achieved a 67% accurate prediction model, which is relatively high considering the complexity of the movement at this rate. This study is open to the use of other prediction models because it only considers GDP, S&P500, inflation, level of activity, hourly wage, interest rate, USD/MXN, CPI, Russell 2000, housing prices, and industrial production variables.

It is speculated that a model using more variables could achieve a higher accuracy level, considering that the results of the models depend on the source and variables used, as can be seen in other articles where methodologies could be similar but show other results.

References

Dritsakis, N., & Klazoglou P., (2018). Forecasting Unemployment Rates in USA Using Box-Jenkins Methodology, *International Journal of Financial and Economic Issues*, 8(1), 9- 20. https://www.econjournals.com/index-.php/ijefi/article/download/5550/pdf/*14966*

Gao, J., & Li, Z., & Li, W., & Shi, B., (2023). Using a time series to predict changes in the U.S. unemployment rate at covid-19, *Highlights in Business, Economics and Management*. 10, 33-43. DOI: https://doi.org/10.54097/h-bem.v10i.7924

Hochreiter, S., and Schmidhuber, J. (1997) LSTM can solve difficult long-time lag problems. In *Advances in Neural Information Processing Systems* 9. MIT Press, Cambridge MA. Presented at NIPS *96.*

Huang, X., (2015). Forecasting the US Unemployment Rate with Job Openings Index, Open Access Master's Theses. Paper 699. University of Rhode Island. Available from: https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1716&context=theses

Kundu, S. and Singhania, R. (2020). Forecasting the United States Unemployment Rate by Using Recurrent Neural Networks with Google Trends Data. *International Journal of Trade, Economics, and Finance*. 11(6) 135-140. DOI: 10.18178/ijtef.2020.11.6.679

Mass, B., (2019). Short-term forecasting of the US unemployment rate, *Journal of Forecasting,* 39(3) 394- 411. https://doi.org/10.1002/for.2630

Meyer, B., & Tasci, M., (2015). Lessons for Forecasting Unemployment in the United States: Use Flow Rates, Mind the Trend Working Paper 2015-1, February. Federal Reserve Bank of Atlanta Working Paper Available from: https://www.atlantafed.org/-/media/documents/research/publications/wp/2015/01.pdf

Törnqvist V. (2021). Time series analysis and forecasting of the unemployment Orebro University School of Business, Available from: https://www.diva-portal.org/smash/get/diva2:1632335/FULLTEXT01.pdf

Xiao, H., & Chen, R., & Guerard, J., (2022). Forecasting the U.S. Unemployment Rate: Another Look. Rutgers University, Piscataway, NJ. Available from: https://statweb.rutgers.edu/hxiao/unemp.pdf

Yurtsever, M., (2023). Unemployment rate forecasting: LSTM-GRU hybrid approach, J Labour Market Res 57, 18 . https://doi.org/10.1186/s12651-023-00345-8

Zou, X., (2024). U.S. unemployment rate prediction using time series model, *Theoretical and Natural Science.* 30(1) 255- 262. DOI: 10.54254/2753-8818/30/20241129.