International Journal of Scientific Research in Multidisciplinary Studies
Vol.6, Issue.11, pp.29-33, November (2020)

Research Paper
E-ISSN: 2454-9312
P-ISSN: 2454-6143

Forecasting Unemployment Rate in the Time of COVID-19 Pandemic Using Google Trends Data *(Case of Indonesia)*

M. Fajar[1*], O. R. Prasetyo[2], S. Nonalisa[3], Wahyudi[4]

[1, 2, 3, 4]Statistics Indonesia (Badan Pusat Statistik)

[*]Corresponding Author: mfajar@bps.go.id, Tel.: +62-89622459926

Available online at: www.isroset.org

Received: 08/Nov/2020, Accepted: 20/Nov/2020, Online: 30/Nov/2020

Abstract— The outbreak of COVID-19 is having a significant impact on the contraction of Indonesia's economy, which is accompanied by an increase in unemployment. This study aims to predict the unemployment rate during the COVID-19 pandemic by making use of Google Trends data query share for the keyword “phk” (work termination) and former series from official labor force survey conducted by Badan Pusat Statistik (Statistics Indonesia). The method used is ARIMAX. The results of this study show that the ARIMAX model has good forecasting capabilities. This is indicated by the MAPE value of 13.46%. The forecast results show that during the COVID-19 pandemic period (March to June 2020) the open unemployment rate is expected to increase, with a range of 5.46% to 5.70%. The results of forecasting the open unemployment rate using ARIMAX during the COVID-19 period produce forecast values are consistent and close to reality, as an implication of using the Google Trends index query as an exogenous variable can capture the current conditions of a phenomenon that is happening. This implies that the time series model which is built based on the causal relationship between variables reflects current phenomenon if the required data is available and real-time, not only past historical data.

Keywords*:* Unemployment, Google Trends, PHK, ARIMAX

I. INTRODUCTION

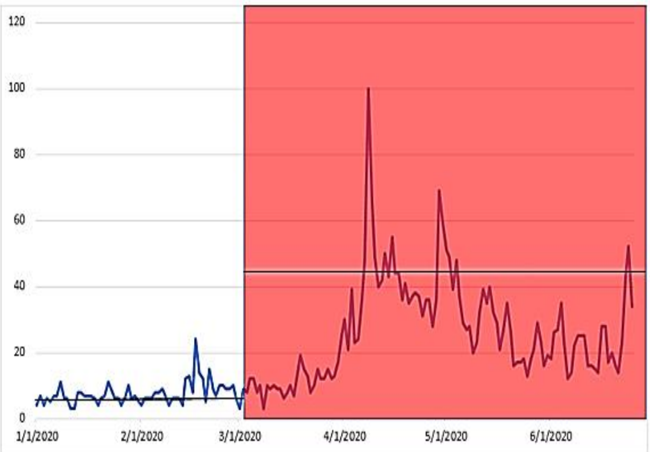
The cumulative number of COVID-19 confirmed cases in Indonesia from March 2 to June 27, 2020, have reached 52,812 cases, with 28,183 patients in treatment, 21,909 patients recovered, and 2,720 patients died. To slow down the spread of the virus, the government of Indonesia has implemented Large-Scale Social Restrictions (PSBB) since mid-March, 2020, which limits specific activities in an area suspected of being infected with COVID-19. Several interventions being implemented, such as the closure of schools, restaurants, factories, and public spaces, including modern shopping centers, restrictions on both domestic and international travel, enactment of working from home policy, as well as enforcement of learning at home regulations.

The implementation of PSBB, which is also followed by homecoming prohibition policy caused the contraction of the economy. According to Badan Pusat Statistik (Statistics Indonesia), the economic growth (year on year) in the first quarter of 2020 was 2.97%, much lower than the first quarter of 2019 [1]. Moreover, in April 2020 the decline in foreign tourist arrivals to Indonesia was 87.44% (year on year), and the room occupancy rate (TPK) of star-classified hotels was 12.67%, experienced a decrease of 41.23 points compared to the previous year or decreased by 19.57 points compared to TPK for March 2020 [2]. Also, there was a decrease (year on year) of passengers of railway, domestic flight, and international flight, by 83.55%, 85.18%, and 98.26% respectively [9]. In terms of export and import, the value of Indonesia's exports from January to May 2020 reached US$ 64.46 billion or decreased by 5.96% compared to the same period in 2019. Moreover, import value in May 2020 reached US$ 8.44 billion, decreased by 32.65% compared to April 2020 or experienced a decrease of 42.20% compared to May 2019 [3].

On the other hand, as many as 43% of micro, small, and medium enterprises (MSMEs) stopped operating during the COVID-19 pandemic in April 2020, 1,139 hotels closed, and 1,174 hotels laid off their employees. According to the Indonesian Chamber of Commerce and Industry, many companies filed for bankruptcy since the wake of COVID-19 pandemic. As a consequence, termination of employees in the affected business units occurred and has implications for an increase in unemployment.

Figure 1 presents the movement of the query index for "phk" (terms of abbreviation in the Indonesian Language) on the search engine by Google Trends for all categories. It can be seen that the "phk" query index during the COVID-19 pandemic tends to increase compared to the time before the pandemic COVID-19. It indicates that there has been an incline in layoffs by the business sector. The PSBB regulation that is implemented to control the spread of

COVID-19 affected population mobility, MSMEs, the tourism industry, and other businesses, which gave a significant impact on economic activity. The ancillary impact of economic activity reduction is implicated in the unavailability or lack of turnover income. However, at the same time, the business operating costs continued. For this reason, many employers have engaged in layoffs in reaction to economic uncertainty.



Notes: red shaded shows the COVID-19 pandemic period in Indonesia, March 2 to June 25, 2020

Figure 1. The movement of Google Trend “phk” (work termination) in Indonesia in the Period of January 1, 2020 to June 25, 2020

Furthermore, according to IMF prediction, Indonesia's economic growth is projected to contract at -0.3% in 2020. On this basis, the unemployment rate is also expected to increase as a result of economic contraction for the remaining three quarters of 2020.

Therefore, this study aims to predict the unemployment rate in Indonesia during the pandemic period by using Google Trends data as an exogenous variable in the model. Forecasting using Google Trends data is expected to reflect the condition in real-time, where official statistics by the government are not yet available. This paper covers four parts, namely: Part 1. The introduction contains a summary of the background of the research; Part 3. Related Work, contains a summary of several studies, Part 3. The method contains a summary of the data sources and methods used; Part IV. The result and Discussion contains a description of the estimation results, and Part V. The Conclusion and Recommendation contains conclusions on the discussion and recommendation.

II. RELATED WORK

Several studies of forecasting using Google Trends data have been carried out, such as forecasting cases of dengue fever in Surabaya [12], forecasting stock market movements [10], and forecasting private consumption [6]. Google Trends data is a real-time reflection of a phenomenon that is studied based on a query determined by the researchers.

The use of Google Trends data is intended as real-time data, where official statistics are not yet available for a phenomenon to be investigated. By entering search words related to a phenomenon in the specified period, Google Trends displays a query index visually and numerically. As Reference [10] uses Google Trends data as a proxy for investor attention, the signals derived from changes in search volume is conditional upon the sentiment inherent to the search terms. Research shows that Google search data can indeed be used as potential signals for stock market movement. However, the directional signal provided by a particular search volume index is conditional on the positivity or negativity of the initial search term.

Reference [12] uses Google Trend search queries related to dengue fever is determined based on 2016 related search results appeared after searching for “dengue fever”. The queries used for ARIMAX modeling are “dengue fever”, “dbd”, “fever” and “dengue”. In the English language “dengue fever” means dengue fever. Other terms in the Indonesian language of dengue fever is “dengue hemorrhagic fever”, shortened into “dbd”. In English "fever" means fever. These data are used as exogenous variables in the ARIMA model which aims to forecast the number of cases of dengue fever in Surabaya in the latest and updated manner. The results of this study found that the ARIMAX model has better performance than ARIMA.

On the other hand, the ARIMA model is still a reliable model to predict phenomena even without using Google Trends data, as a reference [11]. This study showed that the ARIMA (2, 2, 1) and ARIMA (2, 2, 1) -ARCH (1) models were suggested to give tuberculosis surveillance by providing estimates on tuberculosis incidence trends in Tamilnadu.

III. METHODOLOGY

Data Source

This study used the number of unemployed and workforce population data from the results of the labor force survey conducted by Badan Pusat Statistik (Statistics Indonesia) to calculate the open unemployment rate (%). Due to the unavailability of monthly data from the survey, we used the Chow-Lin method [5] to interpolate the annual and semi-annual data into monthly data. Firstly, annual data of the number of unemployed and workforce population for the period of 1986 to 2004 is interpolated by Chow-Lin method into semi-annual data. Secondly, we combined the 2004 semi-annual interpolated data results with semi-annual data (February-August) for the period of February 2005 to February 2020. Lastly, the semi-annual data from January 2004 to February 2020 is interpolated by the Chow-Lin method into monthly data. From these data, the monthly open unemployment rate data can be calculated by dividing the number of unemployed by the number of the labor force and multiplied by 100%. Then, data on the number of unemployed and the population of the labor force for January 2004 to February 2020 were obtained.

Also, we used Google Trends data with the term of "phk" in requesting data queries (z_t) as an exogenous variable, with the data period from January 2004 to June 2020. Specifically, in this study, the training data used is from January 2004 to January 2020, and the open unemployment rate forecast is carried out from February to June 2020.

ARIMA with Exogenous Variable (ARIMAX)

The forecasting method used in this study is ARIMAX. This method was chosen with the consideration that the univariate time series model includes exogenous variables so that forecasting is not only based on historical data values but also supported by exogenous variable information. In this study, ARIMA model $(p, d, q)(P, D, Q)^S$ for a *time series* x_t, which refers to the open unemployment rate, is formulated as follows:

$$\Phi_P B^S \phi_p(B)(1-B)^d(1-B^S)^D x_t = \mu + \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (1)$$

where:

μ : intercept
B : operator lag
p, q : non-seasonal autoregressive order and non-seasonal moving average order
P, Q : seasonal autoregressive order and seasonal moving average order
d : non-seasonal differencing order
D : seasonal differencing order
S : time span of repeating seasonal pattern (monthly (S = 12), quarterly (S = 4))
$\phi_p(B)$: non-seasonal autoregressive component
$\Phi_p B^S$: seasonal autoregressive component
$\theta_q(B)$: non-seasonal moving average component
$\Theta_Q(B^S)$: seasonal moving average component
$(1-B)^d$: non-seasonal differencing
$\left(1-B^S\right)^D$: seasonal differencing
ε_t : error term

ARIMA model parameter is estimated by using maximum likelihood methods. Moreover, to determine the ARIMA model, minimum Akaike Information Criterion (AIC) is used. AIC is formulated as follows:

$$AIC = -2 \times maximum \log likelihood\ of\ model\ (1) + 2 \times number\ of\ model\ parameters \quad (2)$$

In the exogenous variable element, z_t (query index of Google Trends with keywords "phk") is included to equation (1) and obtained formulation as follows:

$$\Phi_P B^S \phi_p(B)(1-B)^d(1-B^S)^D x_t = \mu + (1-B)^\delta z_t + \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (3)$$

with:

$(1-B)^\delta$: non-seasonal differencing for exogenous variable
δ : differencing order for exogenous variable

Equation (3) is ARIMAX [6], model (3) is estimated using maximum likelihood methods based on optimization algorithm of BFGS (Broyden-Fletcher-Goldfarb-Shanno).

To determine the order of integration in the ARIMAX model, this study uses the Philip-Peron test. The advantage of the Phillips-Perron test is in maintaining the possibility of serial correlation in error terms without adding lagged difference terms to the regression [4] and accommodating structural changes in time series [13].

Forecasting Accuracy

Forecasting accuracy of the test data (out sample) in this research uses MAPE (Mean Absolute Percentage Error) formulated as follows [7]:

$$\text{MAPE} = \frac{1}{v-T}\sum_{t=T+1}^{v}\left|\frac{F_t - A_t}{A_t} \times 100\%\right| \quad (11)$$

with F_t is the t-th forecasting result value, and A_t is the t-th actual value. MAPE characteristics: (1) If the MAPE < 10%, then the model perfomance is very good, (2) if the MAPE value is in the range of 10% - 20%, then the model forecasting performance is good, (3) if the MAPE in the range of 20% - 50%, then the model forecasting performance is adequate, and (4) if the MAPE > 50%, then the model forecasting performance is bad.

IV. RESULTS AND DISCUSSION

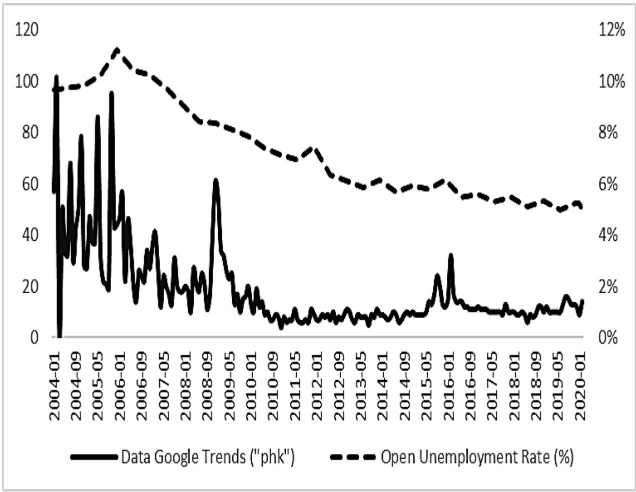


Figure 2. The Movement of Google Trend Query Index Data of "phk" and Open Unemployment Rate (Chow Lin Interpolation) in Indonesia in the Period of January 2004 to February 2020

Figure 2 shows the movement of Google Trends query index data of "phk" and open unemployment rate (%) from January 2004 to February 2020. The movement of the query index data of "phk" and open unemployment rate data experienced a downward trend during that period. However, from January 2004 to April 2009, the movement of the query index was very volatile. The linear correlation value between the two variables is 0.65, which means that the relationship between the two variables is strong and unidirectional, thus strengthening the position of the Google Trends query index as an exogenous variable in our study. As a first step to identify the ARIMAX model, the results of the Phillip-Peron test are presented in table 1. Based on the Philip-Peron test results, the open unemployment rate data is stationary at order 1 (first difference) and the query index is stationary at order 0 (data level).

Table 1. The Results of Phillips-Perron Test on the Open Unemployment Rate Variable and Google Trends Query Index (for the Keywords "phk")

	Open Unemployment Rate	**Google Trends Query Index**
Data level	(-1.524) [0.818]	**(-10.775)** **[0.000]**
First Difference	**(-4.946)** **[0.000]**	(-41.872) [0.000]

(…) presents adjusted t-statistic, and […] presents p-value

The determination of the best ARIMAX model is based on Akaike Information Criterion (AIC) using auto ARIMA in the package 'forecast'. The results reveal that the best model is ARIMAX (2,1,0) $(2,0,0)^{12}$, which means that the order of the non-seasonal components of AR is two and MA is 0 at the first difference, and the order of the seasonal components of AR is 2 and MA is 0 at the data level, with the seasonal period of 12.

$$\Delta x_t = 0.005 + 1.683\Delta x_{t-1} - 0.792\Delta x_{t-2} + 0.364x_{t-12} + 0.449x_{t-24} +$$

(0.047) (0.044) (0.044) (0.070) (0.076)

[0.917] [0.000] [0.000] [0.000] [0.000]

$$(1.776 \times 10^{-5})z_t + \varepsilon_t$$

(3.263×10^{-5})

[0.586] (4)

Adjusted R^2 = 0.999 AIC = -961.6939 Log-likelihood = 487.847, $\Delta x_t = x_t - x_{t-1}$ is the form of first difference x_t (…) presents *standard error*, […] presents *p-value*

The estimation results of the ARIMAX model are presented in equation (4). The adjusted R^2 value of the model is 99.99%, meaning that this model is very fit, or in other words, the model can explain the variation of the response variable by 99.99%, while the remaining 0.01% explained by other factors outside the model. It is also reflected in Figure 3 where the actual and forecast data movements coincide once. Moreover, the MAPE value result reaches 13.46%, which means that the forecasting performance of the ARIMAX model (2,1,0) $(2,0,0)^{12}$ is good.

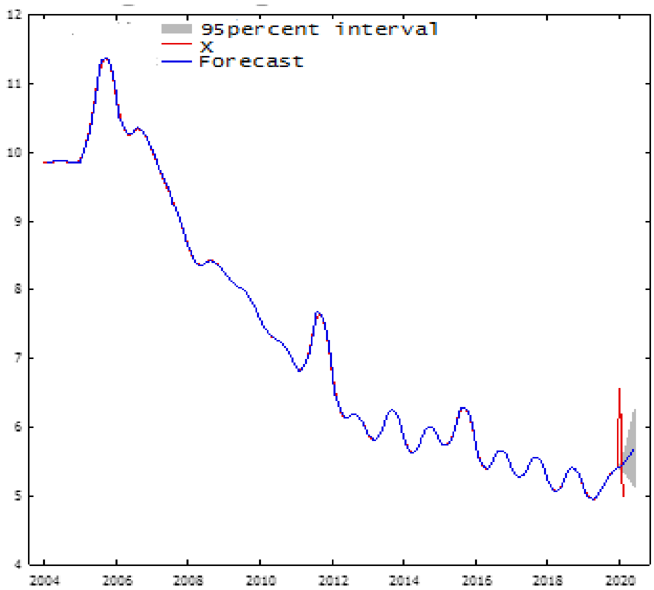


Figure 3. The Comparison of Actual and Forecast Data Results for Open Unemployment Rate for the Period of January to June 2020

Table 2 shows the forecasting results of the open unemployment rate for the period of February to June 2020 by making use of the ARIMAX model. The ARIMAX forecasting results show that there is an increase in the open unemployment rate from February to June 2020, where the open unemployment rate forecast for February, March, April, May, and June is 5.46%, 5.51%, 5.57%, 5.63%, and 5.70%, respectively. So, during the forecasting period, the open unemployment rate is expected to range between 5.46% to 5.70%. The results of forecasting the open unemployment rate using ARIMAX during the COVID-19 period produce forecast values are consistent and close to reality, as an implication of using the Google Trends index query as an exogenous variable can capture the current conditions of a phenomenon that is happening. This implies that the time series model which is built based on the causal relationship between variables reflects current phenomenon if the required data is available and real-time, not only past historical data.

The increase of open unemployment rate corresponds to the impact of COVID-19. As we mention in the previous part, the implementation of Large-Scale Social Restrictions (PSBB) affected the decline of economic activity. As the economic transactions decrease, many businesses suffer losses and stop operating. As a consequence, a number of companies made a company policy to lay off some employees, especially in the Greater Jakarta area. Thus, unemployment rate has increased during the COVID-19 pandemic period.

Table 2. The Open Unemployment Rate Forecasting Results Using ARIMAX Model

Period	Google Trends Query Index	Actual TPT	Forecasted TPT	Forecasted TPT (Lower)	Forecasted TPT (Upper)
February 2020	14	4.99%	5.46%	5.36%	5.56%
March 2020	17		5.51%	5.32%	5.71%
April 2020	69		5.57%	5.26%	5.88%
May 2020	47		5.63%	5.19%	6.07%
June 2020 [until June 25, 2020]	37		5.70%	5.13%	6.27%

V. CONCLUSION AND FUTURE SCOPE

In conclusion, the ARIMAX model has good forecasting capabilities in predicting the open unemployment rate in Indonesia. This is indicated by the MAPE value of 13.46%. The forecast results show that during the COVID-19 pandemic period (March to June 2020) the open unemployment rate is expected to increase, with a range of 5.46% to 5.70%. The results of forecasting the open unemployment rate using ARIMAX during the COVID-19 period produce forecast values are consistent and close to reality.

Because the open unemployment rate is forecast to increase during the COVID-19 pandemic, the Government must stimulate the micro, small, and medium enterprises (MSMEs) sector by providing credit, financial assistance, and facilitating investment regulation. So, people and economical units can survive during the pandemic period.
For the next research, the ARIMAX model can capture regime change by incorporating elements of the Markov model, so that not only the predicted value is obtained but also the probability and duration of the regime.

ACKNOWLEDGMENT
Give thanks to Allah, and for our institutions, Statistics-Indonesia (Badan Pusat Statistik).

REFERENCES

[1] Badan Pusat Statistik, "May 5 edition," *Berita Resmi Statistik (BRS)*, **Indonesia**, **2020**.
[2] Badan Pusat Statistik, "June 2 edition," *Berita Resmi Statistik (BRS)*, **Indonesia, 2020**.
[3] Badan Pusat Statistik, "June 15 edition," *Berita Resmi Statistik (BRS)*, **Indonesia, 2020**.
[4] D.N. Gujarati, "Basic Econome**c**trics, 4th Edition.," *The McGraw-Hill Companies Inc.*, **New York**, **2004**.
[5] G.C. Chow and A.L. Lin, "Best linear unbiased interpolation, distribution, and extrapolation of time series by related series," *The review of Economics and Statistics*, **372-375**, **1971**.
[6] H.J. Bierens, "ARMAX model specification testing, with an apllication to unemployment in the Netherlands," *Journal of Econometrics, 35 (1)*, **161-190**, **1987**.
[7] J.J.M. Moreno, A.P. Pol, A.S. Abad, and B.C. Blasco, "Using the R-MAPE index as a resistant measure of forecast accuracy," *Psicothema 25 (4)*, **500-506.**, **2013**.
[8] J. Woo and A.L. Owen, "Forecasting private consumption with Google Trends data," *Journal of Forecasting,38,* **81– 91**, **2019**.
[9] Lembaga Administrasi Negara, "Dampak COVID-19 terhadap kondisi sosial-ekonomi Indonesia," *Webinar of COVID-19 dan tantangan mewujudkan pembangunan berkelanjutan on June 27, 2020 addressed by Chief Statistician of Statistics Indonesia*, **2020**.
[10] M.Y. Huang, R.R. Rojas, and P.D. Convery, "Forecasting stock market movements using Google Trend searches," *Empirical Economics*, **2019**.
[11] S. Poyyamozhi and A. Kachi Mohideen, "Forecasting Analysis for Tuberculosis (TB) Incidence in Tamilnadu," *International Journal of Scientific Research in Mathematical and Statistical Sciences*, **2018**.
[12] W. Anggraeni and A. Laras, "Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia," *2016 International Conference on Information & Communication Technology and Systems (ICTS). IEEE, 114-11*, **2016**.
[13] W. Enders, "Applied Econometric Time Series, 2nd Edition," *John Wiley & Sons, Inc.*, **New York**, **2004**.

AUTHORS PROFILE

Muhammad Fajar is currently the Statistician of the BPS-Statistics Indonesia Office. In March 2017, he published paper with title; Spectral analysis and markov switching model of Indonesia business; in AIP Conference Proceedings, in November 2017 he published paper with title; PEMODELAN KURVA ENGEL SEDERHANA INDONESIA (Pendekatan Bayesian Quantile Regression); in National Statistics Seminar of Padjadjaran University. In February 2019 he published paper with title; The Estimation of Production Function and Technical Efficiency Shallot Farming; in jurnal matematika Mantik of UIN Sunan Ampel. In October 2019 he published paper with title; An Application of Hybrid Forecasting Singular Spectrum Analysis – Extreme Learning Machine Method in Foreign Tourists Forecasting; in jurnal matematika Mantik of UIN Sunan Ampel.

Octavia Rizki Prasetyo is currently working as a Junior Statistician at BPS-Statistics Indonesia. She graduated from Institute of Statistics in 2015. His research interests are poverty, export-import, and forecasting method comparisons. She published paper conference "A Pilot Project of Area Sampling Frame for Maize Statistics: Indonesia's Experience" at 2020 Asia-Pacific Statistics Week.

Septiarida Nonalisa is currently a Junior Statistician at BPS-Statistics Indonesia of Banten Province Office. Previously, she was a statistician at BPS-Statistics Indonesia of Kotabaru Municipality in South Borneo. She graduated from the Institute of Statistics in 2014. Her research interests are employment, demography, and population studies.

Wahyudi is currently a Junior Statistician of at BPS-Statistics Indonesia of Banten Province Office. He graduated from Institute of Statistics in 2014. Previously, He was a statistician at BPS-Statistics Indonesia of Landak Municipality in West Borneo. His research interests are employment, demography, modelling, and population studies.

