HASIL CEK_Jurnai AIM_SGY_UM

by Jurnai Aim Sgy Um

Submission date: 18-Oct-2021 09:22AM (UTC+0700)
Submission ID: 1676581853
File name: Jurnai AIM_SGY_UM.pdf (871.94K)
Word count: 3261
Character count: 17454

http://www.aimspress.com/journal/mine

Mathematics in Engineering, Volume (Issue): Page.
DOI: 10.3934/MinE.
Received:
Accepted:
Published:

Type of article

Forecasting Educated Unemployed People In Indonesia Using The Bootstrap Technique

Umi Mahmudah [1], Sugiyarto Surono [2]

[1] Department of Mathematics Education, IAIN Pekalongan ,Central Of Java Indonesia

[2] Department of Mathematics, Ahmad Dahlan University, Yogyakarta, Indonesia

* **Correspondence:** Email; sugiyarto@math.uad.ac.id

Abstract: Forecasting is an essential analytical tool used to make future predictions based on preliminary data. However, the use of small sample sizes during analysis provides inaccurate results, known as asymptotic forecasting. Therefore, this study aims to analyze the unemployment rate of educated people in Indonesia using the bias-corrected forecasting bootstrap technique. Data were collected from a total of 30 time series of educated unemployed from 2015 to 2019 using the bias-corrected bootstrap technique and determined using the interval prediction method. The bootstrap replication used is at intervals of 100, 250, 500, and 1000. The results obtained using the R program showed that the bootstrap technique provides consistent forecasting results, better accuracy, and unbiased estimation. Moreover, the results also show that for the next 10 periods, the number of educated unemployed people in Indonesia is projected to decline. The bootstrap coefficient also tends to decrease with an increase in the number of replications, at an average of 0.958. The interval prediction is also known to be smooth, along with a large number of bootstrap replications.

Keywords: AR model, Bias-corrected, Bootstrap, Forecasting

1. Introduction

Forecasting future data is an important process used to determine human resource requirements in order to make effective and efficient decisions while planning design. The Autoregressive (AR) model in time series is a tool used to make future predictions by analyzing previous values. It is sensitive to small sample sizes, thereby affecting the accuracy of forecasting results [1]. According to [2] during forecasting, accuracy is highly dependent on the length of the previous data series. However, sometimes it is difficult to get long data series, especially in the annual form.

This study combined a nonparametric bootstrap approach and an AR model to determine the relatively small sample size. The principle of the bootstrap approach is resampling without return based on the generated data, and it is an effective alternative used to produce better forecasting values. Moreover, the bootstrap approach is distribution-free [3] and is believed to provide relatively accurate prediction results using large and small samples [4]. Point forecasts under the bootstrap approach as well as bootstrap prediction interval are studied. Furthermore, this study follows the bootstrap procedure on the prediction interval model time series AR model [1], [5].

Numerous studies have been applied to the forecasting analysis using time series data [6]–[8]. Clements & Kim applied 3 methods to deal with small samples in forecasting analysis based on the bootstrap approach, namely the bias-corrected OLS, the Roy–Fuller, and the Andrews–Chen estimator. According to them, the Roy-Fuller estimator is generally superior to the other 2 [9]. Moreover, [10] used the bootstrap technique, which allowed for uncertainty in parameter estimation to reduce small sample bias. Meanwhile, a study carried out by [11] applied sieve bootstrap to autoregressive.

Furthermore, the bootstrap techniques application has also been found in forecasting future values in various fields. For instance, it is used to determine probabilistic forecasts from the past single-valued analysis offered by the Numerical Weather Prediction model in Eastern Canada [12]. This method is also used to predict pax in the air transportation industry [13]. Forecasting in the airline industry is also carried out, especially related to accurate demand [14]. Moreover, it is also used to get more accurate predictions on wind power integrated in a sustainable manner [15].

Pattern This study relies on using time-series data to determine the number of unemployment in Indonesia using the bootstrap technique. Data were collected from 30 time-series from 2005 to 2019, published by Statistics Indonesia twice a year and sourced from the national labor force survey (*Sakernas*)..

2. Materials and Methods

2.1. Autoregresive (AR)Model

The autoregressive model is defined by AR(p), which contains an arbitrary deterministic component, $D_{j,t}$, and includes intercepts, time trends, and dummy variables. Furthermore, the AR model is written as follows [1]:

$$Y_t = \sum_{i=1}^{p} \gamma_i Y_{t-1} + \sum_{j=1}^{m} \beta_j D_{j,t} + u_t \tag{1}$$

Where u_t indicates the error term. W is a size nxp matrix which is lagged dependent variable, while D is defined as a size nxm matrix with deterministic components. Then $Z = [W{:}\,D]$ is a size nxk matrix, where $k = p + m$. If $Y = (Y_1, Y_2, \dots, Y_n)'$, $u = (u_1, u_2, \dots, u_n)'$ then equation (1) is written as follows:

$$Y = Z\alpha + u \tag{2}$$

If the vector of the unknown coefficients is defined by $\alpha = (\gamma:\beta)'$, then the least-square estimator of α and σ is defined as follows:

$$\hat{\alpha} = (\hat{\gamma}, \hat{\beta}) = \frac{Z'Y}{Z'Z} \tag{3}$$

$$s^2 = \frac{e'e}{(n-k)}$$

Where $e = (e_1, e_2, \ldots, e_n)'$ shows residuals, assuming the optimal forecast is defined as follows:

$$Y_{n+h} = Y_n(h) + u_{n+h} \tag{4}$$

Where

$$Y_n(h) = \sum_{i=1}^{p} \gamma_i Y_n(h-i) + \sum_{j=1}^{m} \beta_j D_{j,n+h} \tag{5}$$

If $h \leq 0$ then $Y_n(h)$ is estimated as follows:

$$\hat{Y}_n(h) = \sum_{i=1}^{p} \hat{\gamma}_i Y_n(h-i) + \sum_{j=1}^{m} \hat{\beta}_j D_{j,n+h} \tag{6}$$

Equation (5), $\hat{Y}_n(h)$ is consistent and asymptotically normal [1]. Therefore, by using normal approximation, an asymptotic prediction interval is constructed. However, when the sample size used is a small category, the asymptotic predictions give poor results (deficiently). Therefore, alternative ways are needed to deal with small sample sizes and provide better predictive results [5], [16]. This is followed by using a bias-corrected method as an alternative to the bootstrap approach to provide good predictive results based on small sample sizes.

2.2. Bias- Corrected Estimators for AR models

Kim et al. (2010) [1] applied a nonparametric bootstrap approach using residual resampling to estimate the bias of $\hat{\alpha}$ on $O(n^{-1})$ based on equation (1). The sample bootstrap is defined with $\{Y_t^*\}_{t=1}^{n}$, which is degenerated using the starting point $\{Y_t\}_{t=1}^{p}$as:

$$Y_t^* = \sum_{i=1}^{p} \hat{\gamma}_i Y_{t-i}^* + \sum_{j=1}^{m} \hat{\beta}_j D_{j,t} + e_t^* \tag{7}$$

Where e_t^* is a random sample taken with returns based on $\{e_t\}_{t=1}^{n}$. Furthermore, a bootstrap estimator for α is generated and symbolized by $\hat{\alpha}^* = \frac{Z^*Y^*}{Z^{*\prime}Z^*}$. The estimator of the bias-corrected bootstrap is obtained from:

$$\hat{\alpha}_B^C = \hat{\alpha} - bias(\hat{\alpha}) \qquad (8)$$
$$\hat{\alpha}_B^C = [\hat{\gamma}_B^C : \hat{\beta}_B^C]$$

Using equation (8), the following are the steps of the bias-correction bootstrap estimation method:

Step 1:

Calculate the $\hat{\alpha}$ and s^2 estimators and the bias-correction estimator using equations (1) and (8). The residual from the calculation of $\hat{\alpha}$ di is defined by $\{e_t^C\}_{t=1}^n$.

Step 2:

Use starting point $\{Y_t\}_{t=1}^p$ to generate sample bootstrap $\{Y_t^*\}_{t=1}^n$ as $Y_t^* = \sum_{i=1}^p \hat{\gamma}_i^C Y_{t-i}^* + \sum_{j=1}^m \hat{\beta}_j^C D_{j,t} + e_t^*$. Where e_t^* sample is randomly determined using the bias estimation in step 1 defined as $\hat{\alpha}^{C*} = \hat{\alpha}^* - bias(\hat{\alpha})$.

Step 3:

Repeat step 2 in B times to generate a bootstrap distribution for the forecast, $\{Y_n^{C*}(h;j)\}_{j=1}^B$.

The nominal coverage rate of $100(1-\theta)\%$ for bias-corrected bootstrap on the prediction interval is given as follows:

$$[Y_n^*(h,\tau),\ Y_n^*(h,1-\tau)] \qquad (9)$$

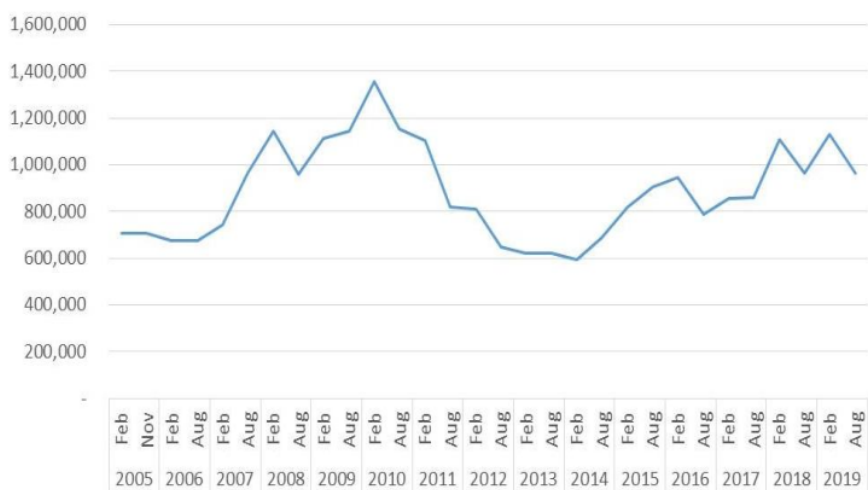
Where $Y_n^*(h,\tau)$ is the $100^{\text{th}}\tau$ percentile of the bootstrap distribution $\{Y_n^{C*}(h;i)\}_{i=1}^B$ and $\tau = 0.5\theta$.

3. Results And Discussion

This study was carried out using 30 educated unemployment data series from 2005-2019 published by BPS in 2020. Indonesia's open unemployment data is tabulated based on the National Labor Force Survey (SAKERNAS) and released twice a year. The time series consists of 2 categories, namely diplomas and universities, which represent the total number of educated unemployed in

Indonesia. Although data for the 2020 period was available at the time of this study, it was not included during the model estimation and forecasting process due to the emergence of the COVID-19 pandemic. [1] stated that the spread of the SARS virus is unpredictable because it had a negative impact due to extreme observations. Therefore, the last data series used was in August 2019, and it is hoped that the prediction results still reflect the real situation on the ground. Moreover, the results also represent normal conditions that do not pay attention to special events capable of effecting changes in data. In other words, when there is no Covid-19, the forecasting results tend to reflect on the future state of the number of educated unemployed in Indonesia.

The time-series data for the educated unemployment in Indonesia from 2005 to 2015 had an average number of 886,180. Meanwhile, the smallest and the highest numbers were 593,556 and 1,358,206, and they both occurred in February of 2014 and 2010. Figure 1 shows the time series data on educated unemployment from 2005-2019, collected only in February and August.



. **Figure 1: Plot of graduate unemployment series**

Figure 1 shows that the data series used are inconsistent due to the increase and decrease in the number of educated unemployed in Indonesia from 2005 to 2019. Deeper analysis found a significant increase from February to August 2011 at 284,419. Meanwhile, the largest decline occurred from August 2017 to February 2018, at a rate of 246,673. The lowest decline was experienced from 2005 to 2019, at a rate of 8,864.

Furthermore, to determine the point forecasts and prediction intervals, this study applied a bias-correction method to generate point forecasts and prediction intervals. Point forecasts are inseparable from the prediction interval because the predictive analysis only uses point estimates. The prediction results are also difficult to determine, therefore, without interval prediction values, point forecasts cannot be ascertained [2]. This makes it difficult to ascertain the value of the prediction interval in expressing the uncertainties that need to arise in the forecast analysis. Some important things that need to be highlighted in prediction intervals include uncertainty in the model and parameter estimates and individual randomness associated with the prediction of certain points.

Figure 2 shows the prediction points for 10 data using the bootstrap approach from February 2020 to August 2024. The predictive analysis under the bootstrap approach was carried out using the bootstrap replications number of 100, 250, 500, and 1000.

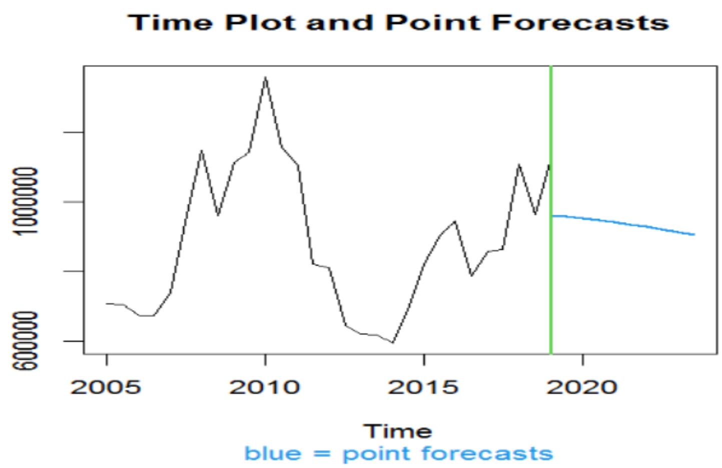


Figure 2: Point Forecasts of graduate unemployment series

Figure 2 shows that the blue line comprises a total of 10 predictive data. A point forecast is interpreted as a function that contains a points summary of the predictive distribution. Meanwhile, the green line is the deadline for data on the number of open unemployment used in August 2019. In general, Figure 2 shows that for the next 10 periods, the number of educated unemployed in Indonesia declined. The results also found that the bootstrap coefficient decreases with increase in the number of replications at an average coefficient of 0.958.

A good forecasting system, which produces residuals with an average of zero, is needed to make unbiased findings [2], [17]. The predictive results indicate that the average residual is 0.000 for all bootstrap (B) while using the replication counts. In other words, the bootstrap technique model carries out a decent job of projecting the educated unemployment rate in Indonesia. Furthermore, the results also indicate that point forecasts and bias-corrected-parameter estimates also provide consistent values, increasing the number of bootstrap replications due to a decrease in average value. Table 1 shows point forecasts based on bootstrap bias-corrected-parameter estimates for 10 future prediction data.

Table 1: Point Forecasts under bootstrap bias-corrected estimation

Point Forecasts	B=100	B=250	B=500	B=1000
h1	962326	960707	960408	960965
h2	958706	955715	955174	956184
h3	954459	950352	949625	950985
h4	949596	944635	943778	945388
h5	944128	938584	937651	939410
h6	938067	932215	931259	933068
h7	931422	925544	924618	926379
h8	924205	918586	917740	919359
h9	916425	911355	910640	912022
h10	908093	903866	903330	904383

Table 1 shows the predicted value in the future indicated by the point forecasts denoted by *h*. The results indicate that the average number of educated unemployed in Indonesia from February 2020 to August 2024 are 938743, 934156, 933422, and 934814 for B values of 100, 250, 500, and 1000. The figures are consistent and do not make a significant difference. Therefore, all point forecasts generated from the number of bootstrap replications indicate consistent results [2]. Figure 2 and Table 2 indicate that the point forecasts for the next 10 data periods decline due to the lack of fluctuations in the prediction data. This indicates that for the next 5 years, the number of educated unemployed will continuously decline. Therefore, the government's strategy to control or reduce the number of educated unemployed is appropriate.

Figure 3 shows that the plotting prediction intervals and point forecast that use bootstrap replications are B=100, 250, 500, and 1000. This study uses quantile estimates which are from the approximate distribution. Moreover, interval prediction is used to determine the range of coverage probabilities defined under the distribution. This study used 95% prediction intervals determined by the forecast distribution's 2.5% and 97.5% quantiles. The 95% prediction interval is commonly used in forecasting analysis besides the 80% [2], [18].

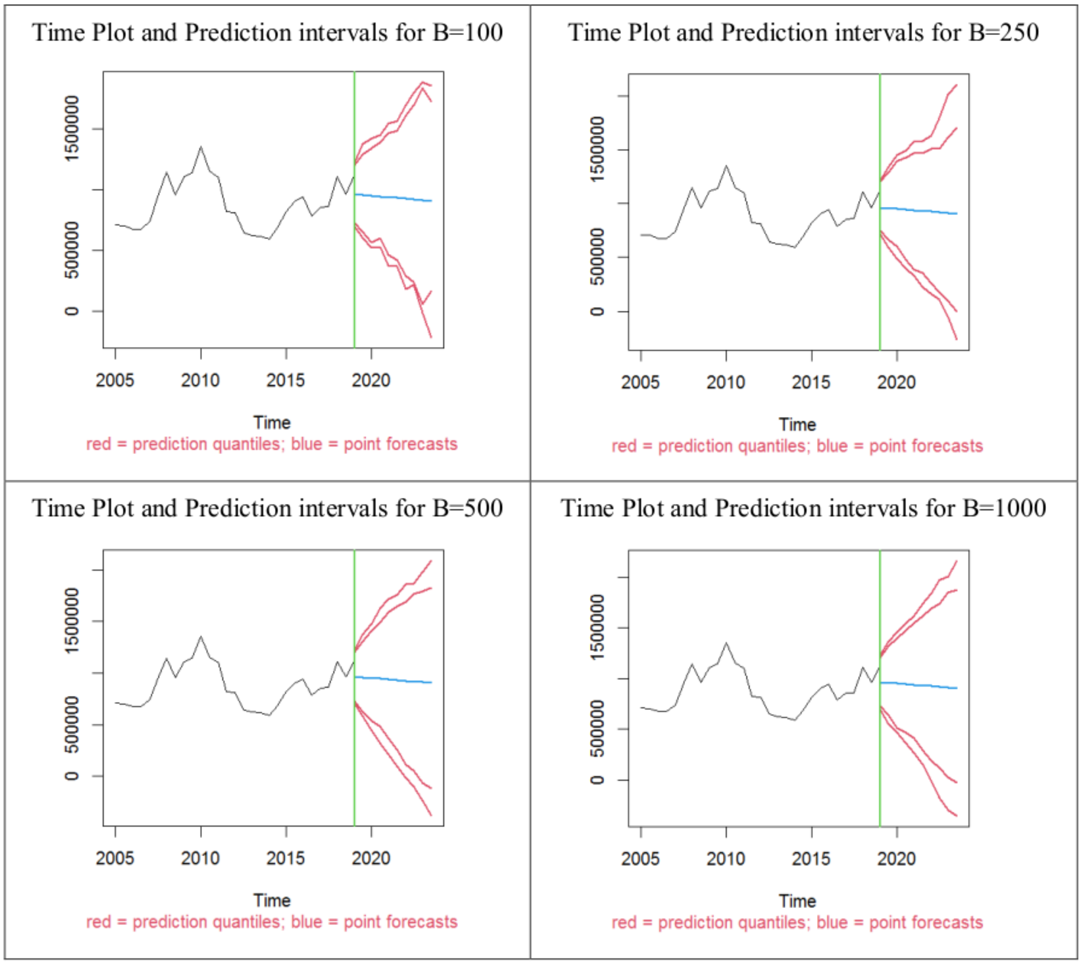


Figure 3. Time Plot and Prediction Intervals under Bootstrap Estimation

Figure 3 shows the green line of the data series in the forecasting analysis, which denotes the boundary. Meanwhile, the blue line shows point forecasts for predictive future data, using the next 10 periods. It is important to note that point forecast predictive analysis cannot provide the accuracy of the projected value without accompanying the prediction interval value [2], [18]. Moreover, when predictive analysis provides a greater degree of uncertainty, the prediction interval tends to widen.

The analysis results in Figure 3 indicate that Indonesia's actual educated unemployment rate needs to be within the prediction interval of 85% at a probability of 0.95. Furthermore, the prediction interval for 95% is used when forecasting over a longer period continuously.

The educated unemployed people in Indonesia are predicted to experience a continuous decline for each number of bootstrap replications used, such as 100, 250, 500, and 1000. Figure 3 shows that the prediction interval of the number of bootstrap replications at 100 and 250 are less smooth, while the number of replications for 500 and 1000 was smoother. Therefore, the bootstrap technique has the ability to provide forecasting results with a better level of accuracy than traditional methods [3].

4. Conclusions

This study successfully used the bootstrap technique to forecast models dealing the problems arise due to small data series. It is also free from data distribution assumptions because it belongs to the category of the nonparametric approach. The time-series data on the number of educated unemployed people in Indonesia from 2005 to 2019 was used to determine the approach used. The results indicate that the number is projected to decline in the next 10 periods. Moreover, the number of bootstrap replications also impacts the forecasting accuracy, where the higher the number of replications used, the smoother and more accurate the resulting prediction intervals. Furthermore, the bootstrap technique is suitable for future value forecasting models compared to traditional techniques.

Acknowledgments (All sources of funding of the study must be disclosed.)

The author thanks to head of center of study data science Department of Mathematics Ahmad Dahlan University.

Conflict of interest

The authors declare no conflict of interest.

References

[1] J. H. Kim, H. Song, and K. K. F. Wong, "Bias-corrected bootstrap prediction intervals for autoregressive model: New alternatives with applications to tourism forecasting," *J. Forecast.*, vol. 29, no. 7, 2010, doi: 10.1002/for.1150.

[2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts. Retrieved from https://otexts.com/fpp3/, 2018.

[3] G. Masarotto, "Bootstrap prediction intervals for autoregressions," *Int. J. Forecast.*, vol. 6, no. 2, pp. 229–239, 1990, doi: 10.1016/0169-2070(90)90008-Y.

[4] M. R. Chernick and R. A. LaBudde, *An introduction to bootstrap methods with applications to R*. John Wiley & Sons, 2014.

[5] J. H. Kim, "Bootstrap-after-bootstrap prediction intervals for autoregressive models," *J. Bus. Econ. Stat.*, vol. 19, no. 1, pp. 117–128, 2001, doi: 10.1198/07350010152472670.

[6] M. La Rocca, F. Giordano, and C. Perna, "Clustering nonlinear time series with neural network bootstrap forecast distributions," *Int. J. Approx. Reason.*, vol. 137, pp. 1–15, 2021, doi:

10.1016/j.ijar.2021.06.014.

[7] E. Paparoditis, “Sieve bootstrap for functional time series,” *Ann. Stat.*, vol. 46(6B), no. 3510–3538, 2018.

[8] A. Staszewska-Bystrova, “Bootstrap prediction bands for forecast paths from vector autoregressive models,” *J. Forecast.*, vol. 30, no. 8, pp. 721–735, 2011, doi: 10.1002/for.1205.

[9] M. P. Clements and J. H. Kim, “Bootstrap prediction intervals for autoregressive time series,” *Comput. Stat. Data Anal.*, vol. 51, no. 7, pp. 3580–3594, 2007, doi: 10.1016/j.csda.2006.09.012.

[10] M. P. Clements and N. Taylor, “Bootstrapping prediction intervals for autoregressive models,” *Int. J. Forecast.*, vol. 17, no. 2, pp. 247–267, 2001, doi: 10.1016/S0169-2070(00)00079-0.

[11] J.-P. Kreiss, E. Paparoditis, and D. N. Politis, “On the range of validity of the autoregressive sieve bootstrap,” *Ann. Stat.*, vol. 39, no. 4, pp. 2103–2130, 2011, doi: 10.1214/11-aos900.

[12] R. Errouissi, J. Cardenas-Barrera, J. Meng, E. Castillo-Guerra, X. Gong, and L. Chang, “Bootstrap prediction interval estimation for wind speed forecasting,” *2015 IEEE Energy Convers. Congr. Expo. ECCE 2015*, pp. 1919–1924, 2015, doi: 10.1109/ECCE.2015.7309931.

[13] M. R. Nieto and R. B. Carmona-Benítez, “ARIMA + GARCH + Bootstrap forecasting method applied to the airline industry,” *J. Air Transp. Manag.*, vol. 71, no. February 2017, pp. 1–8, 2018, doi: 10.1016/j.jairtraman.2018.05.007.

[14] N. A. Mobarakeh, M. K. Shahzad, A. Baboli, and R. Tonadre, “Improved Forecasts for uncertain and unpredictable Spare Parts Demand in Business Aircraft’s with Bootstrap Method,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15241–15246, 2017, doi: 10.1016/j.ifacol.2017.08.2379.

[15] F. Harrou, A. Saidi, and Y. Sun, “Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid,” *Energy Convers. Manag.*, vol. 201, no. May, p. 112077, 2019, doi: 10.1016/j.enconman.2019.112077.

[16] Y. S. Lee and S. Scholtes, “Empirical prediction intervals revisited,” *Int. J. Forecast.*, vol. 30, no. 2, pp. 217–234, 2014, doi: 10.1016/j.ijforecast.2013.07.018.

[17] A. M. de Livera, R. J. Hyndman, and R. D. Snyder, “Forecasting time series with complex seasonal patterns using exponential smoothing,” *J. Am. Stat. Assoc.*, vol. 106, no. 496, pp. 1513–1527, 2011, doi: 10.1198/jasa.2011.tm09771.

[18] M. Chamdani, U. Mahmudah, and S. Fatimah, “Prediction of Illiteracy Rates in Indonesia Using Time Series,” *Int. J. Educ.*, vol. 12, no. 1, pp. 34–41, 2019, doi: 10.17509/ije.v12i1.16589.

HASIL CEK_Jurnai AIM_SGY_UM

ORIGINALITY REPORT

9%	5%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Kebangsaan Malaysia Student Paper	3%
2	Umi Mahmudah. "Autoregressive Integrated Moving Average Model to Predict Graduate Unemployment in Indonesia", Practice and Theory in Systems of Education, 2017 Publication	1%
3	Imad A. Moosa, Jae H. Kim. "Direct and indirect forecasting of the money multiplier and velocity of circulation in the United Kingdom", International Economic Journal, 2007 Publication	<1%
4	Rachid Errouissi, Julian Cardenas-Barrera, Julian Meng, Eduardo Castillo-Guerra, Xun Gong, Liuchen Chang. "Bootstrap prediction interval estimation for wind speed forecasting", 2015 IEEE Energy Conversion Congress and Exposition (ECCE), 2015 Publication	<1%
5	www.jstage.jst.go.jp	

Internet Source

<1%

6 Jae H. Kim. "Forecasting autoregressive time series with bias-corrected parameter estimators", International Journal of Forecasting, 2003
Publication
<1%

7 arxiv.org
Internet Source
<1%

8 www.tandfonline.com
Internet Source
<1%

9 Jae H. Kim, Kevin Wong, George Athanasopoulos, Shen Liu. "Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals", International Journal of Forecasting, 2011
Publication
<1%

10 Michael P. Clements, Jae H. Kim. "Bootstrap prediction intervals for autoregressive time series", Computational Statistics & Data Analysis, 2007
Publication
<1%

11 Rafael Bernardo Carmona-Benítez, María Rosa Nieto. "SARIMA damp trend grey forecasting model for airline industry", Journal of Air Transport Management, 2020
Publication
<1%

12	hira.hope.ac.uk Internet Source	<1%
13	www.mitpressjournals.org Internet Source	<1%
14	"Nonparametric Statistics", Springer Science and Business Media LLC, 2020 Publication	<1%
15	Annisa Eka Haryati, Sugiyarto. "Clustering with Principal Component Analysis and Fuzzy Subtractive Clustering Using Membership Function Exponential and Hamming Distance", IOP Conference Series: Materials Science and Engineering, 2021 Publication	<1%
16	J D Bermúdez, J V Segura, E Vercher. "Bayesian forecasting with the Holt–Winters model", Journal of the Operational Research Society, 2017 Publication	<1%
17	journal.uad.ac.id Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		