BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 0879–0888
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp0879-0888

POSITIVE CONFIRMED PREDICTION OF COVID-19 IN EAST JAVA USING COUNT TIME SERIES BASED DOUBLE POISSON INAR(p) PROCESS

A'yunin Sofro[1*], Framitha Septian Subiantoro[2], Khusnia Nurul Khikmah[3]

[1,2,3]Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya
Jln. Ketintang, Surabaya, Jawa Timur, 60231, Indonesia

*Corresponding author's e-mail: *ayuninsofro@unesa.ac.id*

ABSTRACT

Article History:

Received: 29th July 2024
Revised: 27th January 2025
Accepted: 27th February 2025
Published: 1st April 2025

Keywords:

COVID-19;
East Java;
INAR;
INARDP.

In December 2019, there was a virus outbreak caused by a virus disease with a relatively high spread in Indonesia, one of which was in East Java Province. It is proven by the number of new cases on January 15, 2021, in East Java, reaching 12818 cases. This is why researchers predict the number of positive cases of COVID-19 in East Java so that the Government can anticipate an increase in the number of COVID-19 patients. This study uses data on the addition of positive COVID-19 cases in East Java from May 16, 2020, to January 24, 2021. Because the count time series data shows overdispersion, predictions are made by modeling the COVID-19 data using the INAR(p). development model, namely Double Poisson INAR(p). Several tests were carried out with data from the Double Poisson distribution, and then the ACF and PACF plots were analyzed to find the order of INARDP. After obtaining the order, the model can be constructed and estimated using MLE. Then, the prediction of adding COVID-19 cases in East Java on January 25, 2021, obtained 949 cases with an estimated error of 13.73 percent. So, the model show that the accuracy of the forecasted value with actual value is 86.17 percent.

How to cite this article:

A. Sofro, F. S. Subiantoro and K. N. Khikmah., "POSITIVE CONFIRMED PREDICTION OF COVID-19 IN EAST JAVA USING COUNT TIME SERIES BASED DOUBLE POISSON INAR(p) PROCESS," *BAREKENG: J. Math. & App.*, vol. 19, iss. 2, pp. 0879-0888, June, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

At the end of December 2019, there was an outbreak in China. In January 2020, various provinces in China were attacked by the spread of rampant epidemics, then spread to other countries, including Indonesia. This outbreak is known as COVID-19 [1]. This virus can spread to other people through fluids from the nose or mouth of an infected person, and the liquid can stick to surfaces [2]. In Indonesia, the first reported cases occurred on February 6th, 2020, there were 28276 confirmed cases with 565 deaths globally up until this study [1]. The spread of COVID-19 in East Java, Indonesia, is relatively high. This is evidenced by many positive cases of Corona in East Java, reaching 16862 cases in July 2020. These reasons underlie the research that needs to be carried out. In this research, we will predict the number of additional positive confirmed cases of COVID-19 in East Java on the following day using positive confirmation data from the previous day. Hopefully, this research can help the government anticipate the explosion of hospital patients handling cases due to COVID-19.

The prediction will be made by building a model from positive confirmation data for COVID-19 in East Java. So, this research focuses on the feasibility of the model used by looking at the high enough accuracy errors from the prediction results. This study uses data on the number of additional positive confirmed cases of COVID-19 per day in East Java from May 16, 2020, to January 24, 2021, obtained from the official website of the East Java Provincial Government. Because the data is the sum of an event, the data is called the count data [3]. In addition, it is also time-series data because it is obtained based on time sequences [4]. It can be concluded that the data used is time-series data with count data types, and an observation-driven model (ODM) approach will be used for time-series data with count data types, namely the INAR(p) model. INAR(p) or Integer-valued Autoregressive with p order is one example of the ODM approach [5] with a Poisson distribution. One of the leading models as the basis for INAR(p) is the Autoregressive Moving Average (ARMA(p, q)), with the need to adapt the model to time-series data of the count data type.

Count data often use a regression model based on the Poisson distribution. However, because there is only one parameter in the Poisson distribution, it is a significant drawback for overdispersed or under-dispersed data. Based on research conducted by Wardrop [6], it is known that there is one alternative that can be used in dealing with data that shows overdispersed or under-dispersed, namely the double Poisson distribution [7] with the assumption that the probability mass function approach can be used without adding them one by one, it is discussed in [8], [9]. However, in [3], it is written that the distribution used in the INAR model is the Poisson distribution. In contrast, the data used in this study is assumed to have a double Poisson distribution because it is overdispersed. The prediction will be follow ed by the Double Poisson INAR(p) or INARDP(p) approach as a model of time series data with the type of count data and is overdispersion.

Research involving positive cases of COVID-19 was also carried out by [10] regarding Spatial Modelling Confirmed COVID-19 in East Java using GWR-Negative Binomial used to analyze the influence of the population with positive confirmation cases in the province of East Java. Then, to overcome the data that shows overdispersion, under-dispersion, and equidispersion, a model variation of the Poisson INAR(1) was carried out by [11]. However, in this study, the only estimated parameters and other weaknesses are the order used, which is only 1, so it is necessary to generalize the p-order used. This study models the data that will be used to predict the addition of positive cases of COVID-19 in East Java using one of the Poisson INAR(p) variations with the order of 7, namely INARDP(7).

2. RESEARCH METHODS

2.1 Integer-valued Autoregressive (INAR(p)) Model

In knowing INAR(p), it is necessary to know the $\circ$ operator, usually called the binomial thinning operator. Suppose Y is a random variable non-negative integer. For each $\alpha \in [0,1]$, the operator $\circ$ is defined as:

$$\alpha \circ Y = \sum_{\{i=1\}}^{\{y\}} Y\_i \tag{1}$$

where Y_i is an independent identically distributed random variable (i.i.d), independent of Y.

$$P_r(Y_i = 1) = 1 - P_r(Y_i = 0) = \alpha \tag{2}$$

From the definition of the operator $\circ$ describe in [12], [13], it is known that $0 \circ Y = 0, 1 \circ Y = Y, E\left(\alpha \circ Y = \alpha\, E(Y)\right)$ and for any $\beta \in [0,1), \beta \circ (\alpha \circ Y) = (\beta\alpha) \circ Y$. Next, the INAR(1) $\{Y_t : t = 0, \pm 1, \pm 2, \dots\}$ process will be defined where the process will be defined where $\alpha \in [0,1]$ and ε_t are a sequence of non-negative random variables with uncorrelated integer values that have a mean μ and the variance σ^2 as $Y_t = \alpha_i \circ Y_{t-1} + \varepsilon_t$ there is equation is a first-order Integer-valued Autoregressive model commonly written as INAR(1) [14]. However, this model is limited to stationary data [15].

In [16], the INAR(p) parameter estimation has been discussed. Suppose there is an INAR model with p-order, with $\theta = \left(\alpha_1, \alpha_2, \dots, \alpha_p; \mu\right)'$ is the vector of the INAR parameter which will be estimated by maximizing the following conditional log-likelihoods.

$$l\theta = \sum_{t=p+1}^{n} \ln P(Y_t = y \mid Y_{t-1}) = y_{t-1}, \dots, y_{t-p} \tag{3}$$

Where

$$\begin{aligned} &P\left(Y_t = y \mid Y_{t-1} = y_{t-1}, \dots, y_{t-p} \right) \\ &\quad = \sum_{i_1=0}^{\min(y_{t-1},y)} \binom{y_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{\{y_{t-1}-i_1\}} \times \dots \\ &\quad \times \sum_{i_1=0}^{\min\left(y_{t-p},y-(i_1+\dots+i_p)\right)} \binom{y_{t-p}}{i_p} \alpha_p^{i_p} \left(1-\alpha_p\right)^{y_{t-p}-i_p} \times \frac{\exp(-\mu)\mu^{y-(i_1+\dots+i_p)}}{\{y-\left(i_1+\dots+i_p\right)\}!} \end{aligned} \tag{4}$$

where $i_p = \min\left\{y_{t-p}, y - \left(i_1 + \dots + i_p\right)\right\}$. It is a pdf of INAR with the order p. Estimating parameters of the INAR model with p order and using MLE has also been discussed in [17].

2.2 Double Poisson INAR(p) Process

One of the double exponential families, namely the Double Poisson (DP) distribution, which has been explained in sub-chapter one that this model is indexed by two parameters, namely $\mu > 0$ and $\theta > 0$. From [7] it is found that the form of this approach is close to an infinite series $\sum_{z=0}^{\infty} \tilde{p}\,(z; \mu, \theta)$ and indicated that $\tilde{p}(z; \mu, \theta)$ would be good enough when the value of μ is large $\mu = 10$, and not good enough when the value of μ is smaller. Applying the double Poisson distribution without normalized constants provides good suitability of mean and variance, as was done in [18]. The expected value and the variance are given by approximation $\$(Z) \approx \mu$ and $Var(Z) \approx \frac{\mu}{\theta}$. Thus, the DP distribution allows for overdispersion ($\theta < 1$) and under-dispersion ($\theta > 1$). If $\theta = 1$, then the double Poisson distribution will combine to become a Poisson distribution. $\varphi_{Z(s)} = \frac{Z(\mu s, \theta)}{Z(\mu, \theta)}$. Then, let $\{\epsilon_t\}_{t \in Z}$ be a sequence of discrete random variables (i.i.d) following a Poisson double distribution with PMF is given. This process is commonly called the INARDP(p) process. Thus, the probability is obtained.

$$\begin{aligned} &P_r\left((Z_t = k) \mid Z_{t-1} = l)\right) \\ &\quad = Z(\mu, \theta)\sqrt{\theta} e^{-\theta\mu} \sum_{i=0}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \times \frac{e^{k-i}(k-i)^{k-i}}{(k-i)!} \cdot \left(\frac{e\mu}{k-i}\right)^{\theta(k-i)} \end{aligned} \tag{5}$$

The mean, variance, and Fisher spread index of $\{X_t\}_{t \in Z}$ are given by $\mu_Z = \frac{\mu}{1-\alpha}, \sigma_Z^2 = \frac{\mu(1+\alpha\theta)}{\theta(1-\alpha^2)}$, and $FI_Z = 1 + \frac{\alpha\theta}{\theta + \alpha\theta}$ is estimated using MLE by maximizing the conditional log-likelihood.

Table 1. Dispersion Index of The INARDP(p)

$\boldsymbol{\theta}$	$\boldsymbol{\alpha} \rightarrow$	$\boldsymbol{\alpha = 0.3}$	$\boldsymbol{\alpha = 0.5}$	$\boldsymbol{\alpha = 0.7}$
$\theta = 0.7$		1.3297	1.2857	1.2521
$\theta = 0.5$		1.7692	41.6667	1.5882
$\theta = 0.3$		2.7949	2.5556	2.3725
$\theta = 1$		1	1	1
$\theta = 1.3$		0.8225	0.8462	0.8642
$\theta = 1.5$		0.7436	0.7778	0.8039
$\theta = 1.7$		0.6833	0.7255	0.7578

This process shows equivalence for $\theta = 1$, under-dispersion for $\theta > 1$ and overdispersion for $\theta < 1$. The μ parameter does not change the process dispersion index. Table 1 contains the dispersion indices of the INARDP(p) model for various parameter values. The conditional expectation and conditional variance are given by $E(Z_t \mid Z_{t-1}) = \alpha Z_{t-1} + \mu$ and $Var(Z_t \mid Z_{t-1}) = \alpha(1-\alpha)Z_{t-1} + \frac{\mu}{\theta}$. In practice, the actual value of the model parameter of a process is not known but must be estimated from the realization process $Z_1, \dots, Z_T$. Then, in this article, the estimation method is used, namely MLE (Maximum Likelihood Estimation). However, before parameter estimation can be done after getting the order from the INARDP model, proceed with the model's residual test to find out whether there is a correlation between the resulting residuals [19]. After getting the best model, predictions can be made.

2.3 Data

This study is designed to predict the addition of positive confirmed cases of COVID-19 in East Java Province on January 25, 2021. This study uses secondary data taken from the official website of the East Java Provincial Government (http://infocovid19.jatimprov.go.id/). The data is on adding positive confirmation cases daily from May 16th, 2020, to January 24th, 2021. Then, the analysis obtained with the help of RStudio.

2.4 Stages of Analysis

This study begins with a stationarity test, the ADF test. The data is assumed to be Poisson distributed (where the Poisson distribution ignores overdispersion). The Poisson distribution used in this study has only one parameter, the average incidence rate. The Poisson distribution is a useful tool for studying count data because it has a number of benefits that make it easy to understand and use, especially for simulating unusual events that occur randomly over a predetermined period of time or space. Then, the analysis of the ACF and PACF plots obtained with the help of RStudio will be conducted to select the significant lag (lag through the significance line or the blue dotted line in the PACF plot) as the order of the INAR model. After the order is obtained, parameter estimation to build the INAR model. However, when the second overdispersion test was continued, the results showed that the data used showed overdispersion.

Meanwhile, INAR ignores overdispersion data. The data will be processed using the double Poisson process to handle data that shows overdispersion, and the INARDP model will be used. After going through the double Poisson process, it is continued with the analysis of ACF and PACF data with a new distribution, namely double Poisson, to determine the significant lag that will be used as the order of the INARDP model. After the model order is obtained, parameter estimates are carried out to build the INARDP model. The model is built. After the model is completed, it will be used to predict the number of additional positive cases of COVID-19 in East Java on January 25, 2021. Because the data used is an overdispersion time series count data (where the variance value is greater than the mean), the INARDP model was chosen for use in this study because of its ability to handle data suggesting overdispersion. The response variable of this study is the prediction result in the next period. In contrast, the predictor variable is the data on the addition of positive cases of COVID-19 per day in East Java. The following is a flow chart from the explanation above.

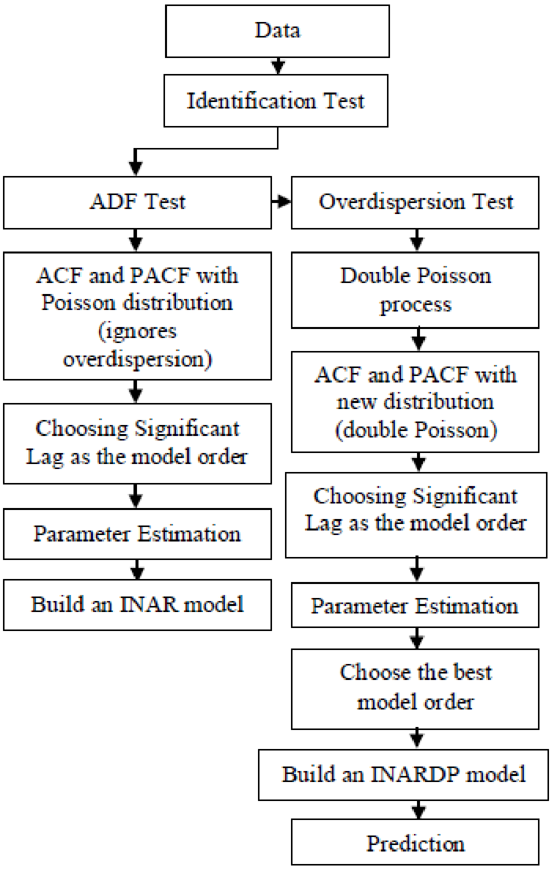

Figure 1. Research Methods Flowchart

3. RESULTS AND DISCUSSION

The INAR model will be used with data directly assumed to be Poisson distributed. The INAR model requires stationary data conditions. So that the first step to be carried out is the stationarity test using the Augmented Dickey-Fuller (ADF) test. In conducting the ADF test, the following hypothesis is used are H_0 data is not stationary and H_1 data is stationary. After performing the ADF test, it was found that the Dickey-Fuller value was -3.4569, and the p-value was 0.0474. The p-value obtained is smaller than the significant level of the test used, namely 0.05, so that H_0 is rejected, and it can be concluded that the data is stationary. Since the data is known to be stationary, the INAR(p) model can be constructed. Selection of the order for INAR via plot analysis of ACF and PACF data. The PACF and ACF plots show that the significant lags are the second, fourth, fifth, sixth, and 12th. The significant lag used in the INAR model's order is the most extensive significant lag, namely the 12th lag with non-zero coefficients on the second, fourth, fifth, sixth, and 12th lags. So, parameter estimation is carried out to build the INAR(12) model; with the help of the R application, the INAR(12) model is obtained as.

$$Y_t = 37.30293 + 0.29668Y_{t-2} + 0.01178Y_{t-4} + 0.00933Y_{t-5} + 0.13710Y_{t-6} + 0.11639Y_{t-12} \quad (6)$$

Another data identification test was conducted namely the overdispersion test. The results of data identification show that the mean data value is 404.8071 and the variance value is 57539.79, so it can be assumed that the data used is overdispersion because the variance value is greater than the mean. To ensure this assumption, an overdispersion test will be carried out. Suppose a model has a Poisson regression distribution where the mean (μ) equals the variance. This condition is called equidispersion. The overdispersion test assesses the hypothesis of the equidispersion condition against the alternative hypothesis, namely the variance more remarkable than the mean, which indicates overdispersion. The following is the hypothesis used in the overdispersion test H_0 is $Var(x) = \mu$ and H_1 is $Var(x) > \mu$ with the help of RStudio, the test results were obtained with a z-value of 5.788 with a $p - value = 3.561 \times 10^{-9}$ which is smaller than the significant level of 0.05. So, it can be concluded that H_0 is rejected. So, the data used indicates overdispersion. Because the INAR model uses the Poisson distribution, where the Poisson distribution is unable to handle the overdispersion of data, based on [6], a double Poisson distribution will be used to deal with data that shows overdispersion. So, the model to be used is Double Poisson INAR or INARDP(p).

This study will focus on the INARDP model, with the basic model used as INAR with Double Poisson distributed data. So, the next step is to ensure that the basic model used is the AR model by using the plot results of the ACF and PACF from the data that has gone through the double Poisson process. The double Poisson process is carried out with the assistance of R software. The following is the plot of ACF and PACF data that has gone through the double Poisson distribution process by Rstudio.

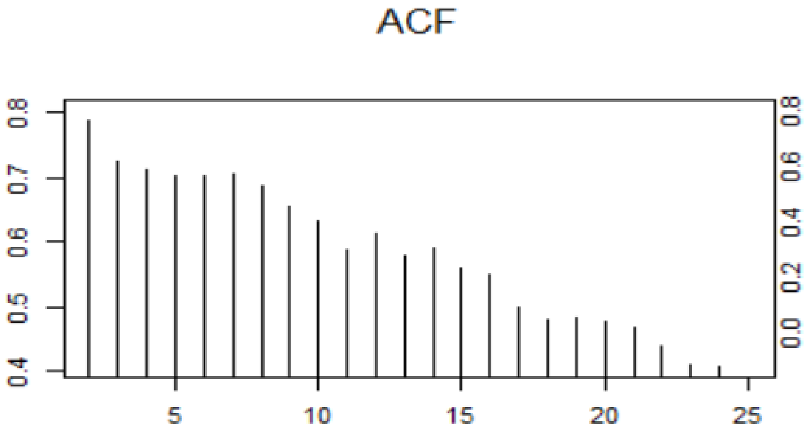

Figure 2. Plot ACF Data

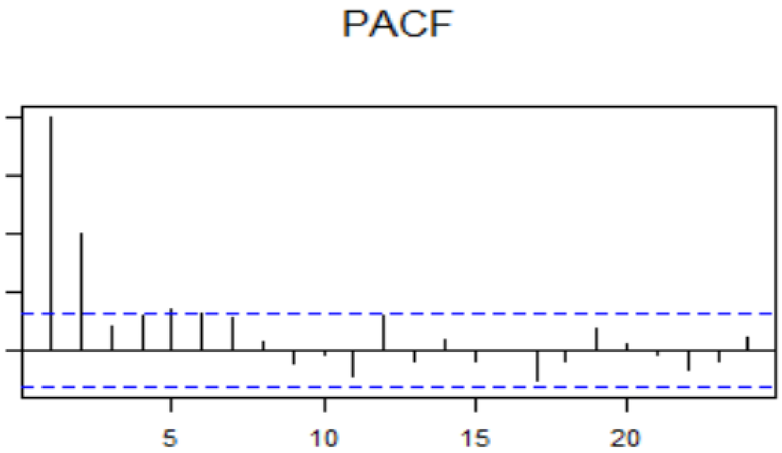

Figure 3. Plot PACF Data

From **Figure 2** and **Figure 3** above, it is known that **Figure 2** shows a pattern with a constant decreasing line commonly known as a dying down or dies down pattern. Meanwhile, **Figure 3** shows an irregular pattern commonly known as a cut-off pattern. Because the ACF and PACF plots show a dying down and cut-off pattern, it can be ascertained that the basic model used is the AR model. The PACF plot (cut-off) shows the lag of significance (which crosses the significant line or the blue dotted line) on the first, second, fourth, fifth, sixth, seventh, and 12th lags.

Because the model to be used is INARDP, it should be noted that the INAR model requires a condition where the data used is stationary. By using the same ADF test results in the first step, the INARDP(p) model can be built because the data is stationary. In selecting the order from INARDP using analysis from the results of the ACF and PACF plots in **Figure 2** and **Figure 3**, the order is selected from the largest significant lag. **Figure 3** shows that the lags are significant, namely the first, second, fourth, fifth, sixth, seventh, and 12th. So, what is used as the order should be 12 with a non-zero coefficient on the first, second, fourth, fifth, sixth, seventh, and 12th lags. Then, the parameter estimation for INARDP(12) should be carried out, and the residual will be tested. However, after testing the residuals, it is found that the value of $\chi^2_{0.05;1}$ is 43.427 and the $p - value = 4.4 \times 10^{-11}$ which is smaller than the significant level. So that H_0 is rejected. Moreover, it can be concluded that the residual INARDP(12) is correlated with residuals. So, the INARDP(12) model cannot be trusted for predictions, and the order chosen is 7. With the help of R, the estimation results of the INARDP(7) model are listed in Table 2 below.

Table 2. Estimated Parameters Model

Coefficients	Estimate	Std. Error	Cl (Lower)	Cl (Upper)
Intercept	21.95566	3.1649	15.75256	28.1588
α_1	0.33935	0.0116	0.31668	0.3620
α_2	0.34486	0.0125	0.32033	0.3694
α_4	0.01836	0.0136	-0.00828	0.0450
α_5	0.00201	0.0133	-0.02414	0.0282
α_6	0.09025	0.0134	0.06395	0.1165
α_7	0.15922	0.0127	0.13425	0.1842

Using the R application's help, the AIC value of the INARDP(7) model is 9148.678. Because the value of the INARDP parameter has been obtained, the INARDP(7) model can be built and the parameter values in Table 2 to obtain Z_t.

$$\begin{gathered} Z_t = 0.33935Z_{t-1} + 0.34486Z_{t-2} + 0.01836Z_{t-4} + 0.00201Z_{t-5} + 0.09025Z_{t-6} \\ + 0.15922Z_{t-7} + 21.95566 + \epsilon_t \sim DP(\theta, \mu) \end{gathered} \tag{7}$$

Equation (7) show that these parameters represent the autoregressive coefficients. They capture the dependence of the current count on the counts from the previous 7 time periods. For example, α_1 represents the influence of the count in the previous time period $(t-1)$ on the current count (t). Equation above is a model from $INARDP(7)$ built from data on the addition of COVID-19 cases in East Java. After that, a diagnostic test by using R studio of the INARDP model will be carried out and obtained.

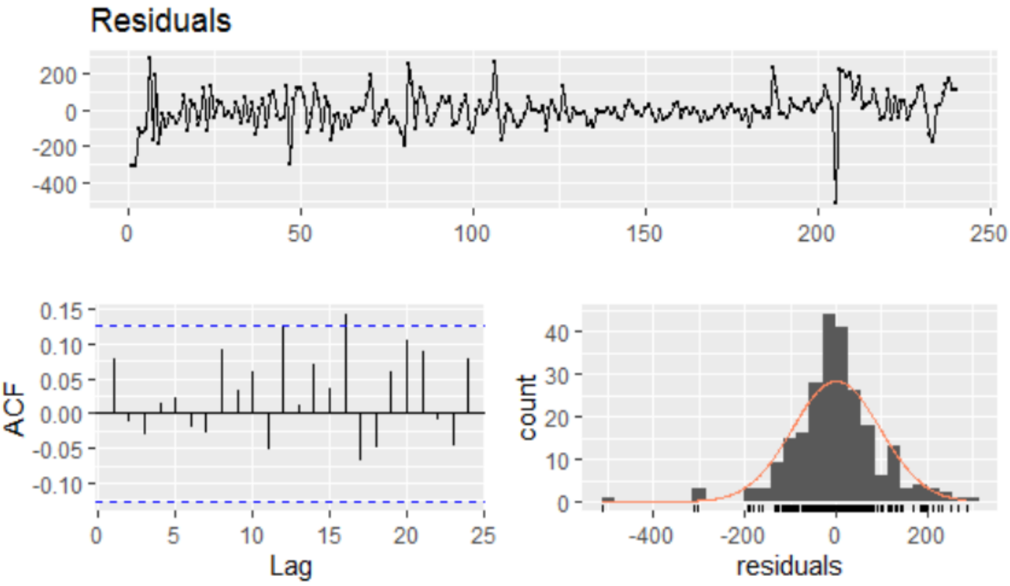

Figure 4. Residual Plot of INARDP(7) Model

The residual homogeneity test is obtained from the scatter plot analysis in Figure 4. The top image is a scatter plot from the INARDP(7) residual results, where the plot does not show a specific pattern so that it can be seen that the residuals are identical. Next is to test the correlation between residuals using the Ljung-Box (LB) test, and the following hypotheses are used H_0 is the residuals are uncorrelated and H_1 there is a correlation between residuals.

This LB test wants a condition where H_0 or the initial hypothesis fails to be rejected. After the LB test is carried out, the value of $\chi^2_{0.05;1}$ is 1.4997, and the p-value $= 0.2207$. It is known that the p-value is more significant than the significance level, so H_0 fails to be rejected. Moreover, it can be concluded that the INARDP(7) residuals are not correlated. The results of the two residual tests that have been carried out show that the residuals from INARDP(7) are identical and not correlated. So that the INARDP(7) model can be trusted, this will be followed by a prediction for the number of additional positive cases in the next 15 days with the INARDP(7) model. For comparison, here are the AIC results from INAR and INARDP.

Table 3. AIC and BIC Value

Model	AIC	BIC
INAR(12)	9150.869	9175.234
INARDP(7)	9148.678	9173.042

The statistical measures used in this research are aimed at comparing the quality of different statistical models. It indicates the best model among candidate models by balancing model fit and model complexity. Where lower AIC and BIC values indicate a better model fit. The AIC and BIC value of INARDP(7) is smaller than the AIC and BIC of INAR(12) because INAR ignores data overdispersion conditions. This proves that the INARDP(7) model is better used to predict the number of additional positive cases of COVID-19 in East Java than the INAR(12) model. Then, the INARDP(7) model will be used to predict. Predictions were made to predict the number of additional positive cases of COVID-19 in East Java Province the following day. Since the INARDP(7) model has been built. The data used to build the INARDP(7) model will be divided into train and test data. The distribution of train data and test data has yet to be determined. In this study, the first 240 data sets will be used as train data, and the remaining 14 will be used as test data. The train data will be used to build the model, and the model will be used to predict for the next 15 days. The prediction results will be compared with the original data and presented in Table 4. The difference between the prediction and the original data will be used to calculate the MAPE and determine the estimated prediction error. Since the number of additional positive cases of COVID-19 must be an integer, the prediction results are rounded up to 949 [20].

Table 4. Prediction Results of INARDP(7)

Test Data	Prediction	Difference
792	903.0476	111.0476
844	921.6320	77.632
815	877.3599	62.3599
981	845.3670	135.633
1198	862.3670	335.0591
1160	901.3463	258.6537
974	1034.3520	60.352
848	1076.7295	228.7295
972	974.4336	2.4336
955	877.5119	77.4881
1134	886.2412	247.7588
1056	965.7587	90.2413
919	1049.0745	130.0745
901	1063.5212	162.5212
888	949.0818	61.0818

The prediction column is the prediction result of the built INARDP(7) model, while the difference column is the difference between the test data and the predicted data. Suppose the original data is Z_i, then the prediction data is $\{\hat{Z}_i\}$, so the MAPE obtained is 13.73 percent for the INARDP(7) model.

Research conducted around this outbreak has also been carried out by [21] with data-driven analysis techniques using an observation-driven model approach where the data held is used to predict the number of positive case data for COVID-19 the following day. Besides, [22] predicted COVID-19 using the ARIMA method and an exponential development model, [23] analyzing and predicting the number of cases in three countries. For prediction using the INAR model, it is also carried out by [19] with the Markov Chain approach, where INAR-Poisson demands data to follow the characteristics of the Poisson distribution. This needs to be more flexible in the model used. Then, to overcome the data that shows overdispersion, underdispersion, and equidispersion using a variation of the model from Poisson INAR(1) have the weakness of the article is the use of order 1. Because generalization is needed, order p. The order p used in this study is 7, so if order 1 is enforced for this article, predictions will be inaccurate. This study models the data that will be used to predict the addition of positive cases of COVID-19 in East Java using one of the Poisson INAR variations of the order of 7, namely INARDP(7). It is carried out in sufficient detail using general theories used in statistics learning. The discussion of these fundamental theories is carried out in the hope that this article is easy to digest and understand.

4. CONCLUSIONS

Data for adding COVID-19 cases in East Java is time-series data, so the model used is the INAR type. However, INAR usually uses the Poisson distribution, and the overdispersion of this data must be addressed. Then, the INAR development model, namely the INARDP model, will be used. The assumption of stationarity is also influential in several models. It is proven by INARDP(p) that it cannot handle data that is not stationary because the basic model of INARDP is INAR. It has the same characteristics, which cannot handle non-stationary data. INARDP model with order 7 was chosen as a predictor because order 12 did not fulfill the conditions as a predictor. Besides, INARDP(7) has the smallest AIC value than INAR(12). So, from the INARDP(7) model, the prediction results obtained from the number of additional positive cases of COVID-19 in East Java on January 25, 2021, amounted to 949 people with an estimated MAPE prediction error of 13.73 percent. This value indicates that based on the model used to forecast the data has a forecasting accuracy of 86.27 percent. This value is obtained based on the error of the forecasting value compared to the actual value.

REFERENCES

[1] N. Coronavirus, “Available online: https://www. who. int/emergencies/diseases/novel-coronavirus-2019,” *Accessed on*, vol. 10, 2020.

[2] P. S. Blagov, “ADAPTIVE AND DARK PERSONALITY IN THE COVID-19 PANDEMIC: PREDICTING HEALTH-BEHAVIOR ENDORSEMENT AND THE APPEAL OF PUBLIC-HEALTH MESSAGES,” *Soc Psychol Personal Sci*, vol. 12, no. 5, pp. 697–707, 2021.

[3] C. H. Weiß, “A POISSON INAR (1) MODEL WITH SERIALLY DEPENDENT INNOVATIONS,” *Metrika*, vol. 78, no. 7, pp. 829–851, 2015.

[4] D. C. Montgomery, C. L. Jennings, and M. Kulahci, INTRODUCTION TO TIME SERIES ANALYSIS AND FORECASTING. John Wiley & Sons, 2015.

[5] W. W. S. Wei, “TIME SERIES ANALYSIS: UNIVARIATE AND MULTIVARIATE,” *Methods. Boston, MA: Pearson Addison Wesley*, 2006.

[6] R. Wardrop, “REGRESSION MODELS FOR COUNT DATA BASED ON THE DOUBLE POISSON DISTRIBUTION,” 2016.

[7] B. Efron, “DOUBLE EXPONENTIAL FAMILIES AND THEIR USE IN GENERALIZED LINEAR REGRESSION,” *J Am Stat Assoc*, vol. 81, no. 395, pp. 709–721, 1986.

[8] A. C. Cameron and P. Johansson, “COUNT DATA REGRESSION USING SERIES EXPANSIONS: WITH APPLICATIONS,” *Journal of Applied Econometrics*, vol. 12, no. 3, pp. 203–223, 1997.

[9] Y. Zou, S. R. Geedipally, and D. Lord, “EVALUATING THE DOUBLE POISSON GENERALIZED LINEAR MODEL,” *Accid Anal Prev*, vol. 59, pp. 497–505, 2013.

[10] R. Fitriani and I. G. N. M. Jaya, “SPATIAL MODELING OF CONFIRMED COVID-19 PANDEMIC IN EAST JAVA PROVINCE BY GEOGRAPHICALLY WEIGHTED NEGATIVE BINOMIAL REGRESSION,” *Communications in Mathematical Biology and Neuroscience*, vol. 2020, pp. 1–17, 2020.

[11] M. Bourguignon, J. Rodrigues, and M. Santos-Neto, “EXTENDED POISSON INAR (1) PROCESSES WITH EQUIDISPERSION, UNDERDISPERSION AND OVERDISPERSION,” *J Appl Stat*, vol. 46, no. 1, pp. 101–118, 2019.

[12] F. W. Steutel and K. van Harn, “DISCRETE ANALOGUES OF SELF-DECOMPOSABILITY AND STABILITY,” *The Annals of Probability*, pp. 893–899, 1979.

[13] K. Van Harn, “CLASSIFYING INFINITELY DIVISIBLE DISTRIBUTIONS BY FUNCTIONAL EQUATIONS,” 1978.

[14] G. S. Osho, “A GENERAL FRAMEWORK FOR TIME SERIES FORECASTING MODEL USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE-ARIMA AND TRANSFER FUNCTIONS,” *Int J Stat Probab*, vol. 8, no. 6, pp. 1–23, 2019.

[15] K. F. Pinontoan, “PEMODELAN SINGLE EXPONENTIAL SMOOTHING (SES) DAN INTEGER AUTOREGRESSIVE (INAR) PADA PERAMALAN PERMINTAAN INTERMITTENT,” *FRONTIERS: JURNAL SAINS DAN TEKNOLOGI*, vol. 1, no. 2, 2018.

[16] R. Bu and B. McCabe, “MODEL SELECTION, ESTIMATION AND FORECASTING IN INAR (P) MODELS: A LIKELIHOOD-BASED MARKOV CHAIN APPROACH,” *Int J Forecast*, vol. 24, no. 1, pp. 151–162, 2008.

[17] X. Pedeli, A. C. Davison, and K. Fokianos, “LIKELIHOOD ESTIMATION FOR THE INAR (P) MODEL BY SADDLEPOINT APPROXIMATION,” *J Am Stat Assoc*, vol. 110, no. 511, pp. 1229–1238, 2015.

[18] F. Zhu, “MODELING OVERDISPERSED OR UNDERDISPERSED COUNT DATA WITH GENERALIZED POISSON INTEGER-VALUED GARCH MODELS,” *J Math Anal Appl*, vol. 389, no. 1, pp. 58–71, 2012.

[19] N. Silva, I. Pereira, and M. E. Silva, “FORECASTING IN INAR (1) MODEL,” *REVSTAT-Statistical Journal*, vol. 7, no. 1, pp. 119–134, 2009.

[20] F. C. Drost, R. Van den Akker, and B. J. M. Werker, “EFFICIENT ESTIMATION OF AUTO-REGRESSION PARAMETERS AND INNOVATION DISTRIBUTIONS FOR SEMIPARAMETRIC INTEGER-VALUED AR (p) MODELS,” *J R Stat Soc Series B Stat Methodol*, vol. 71, no. 2, pp. 467–485, 2009.

[21] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, “ESTIMATION OF THE REPRODUCTIVE NUMBER OF NOVEL CORONAVIRUS (COVID-19) AND THE PROBABLE OUTBREAK SIZE ON THE DIAMOND PRINCESS CRUISE SHIP: A DATA-DRIVEN ANALYSIS,” *International journal of infectious diseases*, vol. 93, pp. 201–204, 2020.

[22] M. S. A. Abotaleb, “PREDICTING COVID-19 CASES USING SOME STATISTICAL MODELS: AN APPLICATION TO THE CASES REPORTED IN CHINA ITALY AND USA,” *Academic Journal of Applied Mathematical Sciences*, vol. 6, no. 4, pp. 32–40, 2020.

[23] D. Fanelli and F. Piazza, “ANALYSIS AND FORECAST OF COVID-19 SPREADING IN CHINA, ITALY AND FRANCE,” *Chaos Solitons Fractals*, vol. 134, p. 109761, 2020.