MATEMATIKA: MJIAM, 2020, Volume 36, Number 1, 15–30

Modelling Asthma Cases using Count Analysis Approach: Poisson INGARCH and Negative Binomial INGARCH

[1]'Aaishah Radziah Jamaludin, [2]Fadhilah Yusof* and [3]Suhartono

[1,2]Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia

[3]Department of Statistics, Faculty of Mathematics and Natural Sciences
Institut Teknologi Sepuluh Nopember, Surabaya, Jawa Timur 60111, Indonesia

*Corresponding author: fadhilahy@utm.my

Article history
Received: 14 February 2019
Received in revised form: 23 September 2019
Accepted: 23 January 2020
Published online: 1 April 2020

Abstract Pollution in Johor Bahru is an issue that needs adequate attention because it has contributed to a number of asthma cases in the area. Therefore, the goal of this study is to investigate the behaviour of asthma disease in Johor Bahru by count analysis approaches namely; Poisson Integer Generalized Autoregressive Conditional Heteroscedasticity (Poisson-INGARCH) and Negative Binomial INGARCH (NB-INGARCH) with identity and log link function. The estimation of the parameter was done by quasi-maximum likelihood estimation. Model assessment was evaluated from the Pearson residuals, cumulative periodogram, the probability integral transform (PIT) histogram, log-likelihood value, Akaike's Information Criterion (AIC) and Bayesian information criterion (BIC). Our result shows that NB-INGARCH with identity and log link function is adequate in representing the asthma data with uncorrelated Pearson residuals, higher in log likelihood, the PIT exhibits normality yet the lowest AIC and BIC. However, in terms of forecasting accuracy, NB-INGARCH with identity link function performed better with the smaller RMSE (8.54) for the sample data. Therefore, NB-INGARCH with identity link function can be applied as the prediction model for asthma disease in Johor Bahru. Ideally, this outcome can assist the Department of Health in executing counteractive action and early planning to curb asthma diseases in Johor Bahru.

Keywords Asthma cases; Pollution; Count data; Poisson INGARCH; NB-INGARCH

Mathematics Subject Classification 62P10, 62-07, 97K80

1 Introduction

Asthma is a typical incessant disease that influences numerous individuals of any age in all parts of the world. It is a burden to individuals, frequently causing reduced personal satisfaction, because of its physical impacts, as well as its mental and social impacts. The common

symptoms of asthma diseases are wheezing, coughing, shortness of breath and chest tightness. Inadequately controlled asthma can prompt numerous visits to the emergency room and if it worsens can lead to hospital admission. Three major feature of asthma are inflammation, airway obstruction and airway irritability. Referring to the Centers for Disease Control of America, in 2017, 18.4 million adults, or 7.6% of the adult residents, have asthma. Almost one of every 12 American kids currently have asthma. Beginning in 2015, an expected 6.2 million kids under age 18 have been identified to have this disease The normal triggers for asthma are infections, for example, sinusitis, colds and influenza, allergens, for example, dusts, shape, aggravations, for example, solid scents from fragrances or cleaning arrangements, and air contamination, tobacco or smoke, climate; changes in temperature or humidity, cool air, and so forth.

This study was conducted in Johor Bahru since it is an important metropolitan city and one of the fastest-growing cities in Malaysia after Kuala Lumpur. The rapid development of Johor Bahru in terms of transportation, construction, manufacturing, electrical and electronics processing may contribute to the air pollution episodes. As a result, the residents of Johor Bahru are exposed to asthma diseases and other respiratory problems. Therefore, this study is one of the initiatives in order to determine the behaviour of asthma diseases in this city from a statistical point of view. The weekly asthma data was obtained from Johor Bahru Department of Health for the period of 2012 to 2015. The data will be analysed by using the count time series approach. Modelling asthma diseases by using time series approach has been widely applied by many researchers all around the world. The most common time series models that have been used to analyse asthma data are Poisson regression models [1,2]Poisson generalized linear model [3], Poisson Generalized additive model [4] and others. So far as our knowledge, modelling asthma diseases by using Poisson INGARCH and NB INGARCH is still limited and there is considerable room for that. These two methods were chosen due to their capabilities in handling problem such as over dispersion, outbreak and serial correlation that exhibits in the dataset. Modelling and forecasting of asthma disease are important especially in the health sector to execute the preparation in terms of facilities, medicine supplies and medical staff. It is hoped that this piece of knowledge is able to contribute to the Health department of Johor Bahru and the public as well

The outline of the article is as follows. Section 2 discusses the literature review on the previous asthma analysis and count analysis method. Section 3 describes the methodology of Poisson INGARCH and NB-INGARCH. Section 4 gives some empirical results and in section 5 we conclude with some recommendations for further studies.

2 Literature Review

The analysis of asthma diseases that have been conducted by using time series methods for the past few years is reviewed in this paper. The risk of hospitalization for asthma in children after exposure to air pollutants in Southeast Brazil was estimated by [1] using Poisson regression generalized additive models. They found that exposure to particulate matter (PM10) and Sulphur dioxide (SO_2) were related with significant relative risks of hospital admission due to asthma on the same day and within 3 days after exposure. Meanwhile, the risk of ozone concentration and hospitalizations for asthma was estimated by [2] using generalized additive Poisson's regression model and superposed epoch. They found that the relative risks of ozone and asthma are 0.9975 which is considered high. The seasonal influence of air pollutants and

aeroallergens on the risk of asthma hospital admissions for adults and children in Adelaide, South Australia was assessed by [5] using generalized log-linear quasi-Poisson and negative binomial regressions. They found that the major effects on asthma admissions related to particulate matter with 2.5 mm in size ($PM_{2.5}$), Nitrogen Dioxide (NO_2), particulate matter with 10 mm in size (PM_{10}) and the major consequence for ozone was found in the warm season for children. The influence of pollutants on asthmatic children in Chongqing, China was assessed by using time-stratified case-crossover and conditional logistic regression [6]. They found that short-term exposure to PM10, PM2.5, sodium dioxide, nitrogen and carbon monoxide could trigger the hospital visits for asthma in children. The short-term relation between ambient concentrations of several aeroallergens and hospitalizations due to asthma was examined by using quasi-Poisson regression with distributed lag models [7]. They found that airborne grass, birch and hornbeam pollen are connected with severe asthma exacerbations in the Brussels region. These mixtures seem to perform in interaction with air pollution and to more precisely affect young and intermediate age groups. Therefore based on the previous studies, it was proven that air pollutants have significant effects on asthma cases.

Meanwhile, in term of climates influence, [8] estimated rate ratios (RR) between daily maximum temperature (T_{max}) and asthma visits, monitoring for time and meteorology by using Poisson generalized linear models. They found that higher RRs between T_{max} and asthma between males compared to females, non-white children compared to white children and children with private insurance compared to children with Medicaid. The association between diurnal temperature range (DTR) and childhood asthma by was examined by [3] using Poisson generalized linear model combined with a distributed lag non-linear model. They found that The DTR effect on childhood asthma increased above a DTR of $10C^0$. Male children and children aged 5–9 years looked to be more vulnerable to the DTR effect than others. The short-term effects of daily mean temperature on asthma hospital admissions by using Poisson generalized additive model (GAM) incorporated with a distributed lag non-linear model was assessed by [4]. They found that the relative risk of asthma hospital admissions connected with cold temperature was 1.20 at lag 0 to 14. However, warmer temperatures were not related to asthma. Our previous work [9], applied Poisson Generalized Linear Model and Negative Binomial Model to model the asthma disease in Johor Bahru for the period of 2012 to 2013. They found that humidity and temperature have a significant relationship with asthma diseases. However, both models have a drawback in terms of normality.

Dealing with integer-valued time series model, it is common to assume the model as a simple linear autoregressive where the observation relies on its past, similar in the popular class of autoregressive moving average (ARMA) models and its extensions. Unfortunately, for integer-valued time series or sometimes called as count analysis, this would not guarantee the observations to be non-negative integers. As a solution selecting a suitable distribution for count data and an appropriate link function can be helpful. Two common methods of count data that has been widely used are Integer autoregressive (INAR(p)) and INGARCH (p,q) [10]. Based on our review the application of count data analysis namely Poisson INGARCH and NB-INGARCH is still limited in the asthma data. Poisson INGARCH has been introduced by [11–13]. This model can be assumed as an integer-valued complement to the conventional GARCH model. As an alternative to the conditional Poisson distribution, [14] and [15] considered the negative binomial distribution. The conditional variance of the negative binomial distribution is greater than the conditional variance of the Poisson case, which reflects the conditional over

dispersion that is assumed to be present on real series [16].

The studies of Poisson INGARCH in many fields has been witnessed such as [17] who introduced a generalized Poisson INGARCH model in analysing the series of annual counts of major earthquakes for the years of 1900 to 2006. The outcome demonstrates that the proposed model does not just has great execution in displaying over-dispersed information yet, in addition, has the capacity to show the under-dispersed occurrence The issue of evaluating prediction of monthly number of measles at Sheffield for the period of 1978 to 1987 was considered by [18] which is based on either the Poisson distribution or the negative binomial distribution. The result shows that the negative binomial prediction is superior compared to the Poisson prediction. Then in the following year, [15] studied the inference and diagnostics for count time series regression models in negative binomial processes that comprise a feedback mechanism in transactions data. They found that the negative binomial distribution is outperformed than the Poisson distribution Hence, in this study we applied both; Poisson distribution and negative binomial distribution to see which distribution fits the asthma data best. The different methods for modelling intervention effects in time series of counts of campylobacterosis cases was studied by [19] concentrating on the integer-valued GARCH model. They found some robustness and usefulness to incorporate the intervention effects in INGARCH model. Therefore, the intervention effects should not be neglected if the outbreak exists in the data. Motivated by this finding, we include the intervention effects in our model. The asymptotic distribution of the estimator is when the parameter belongs to the interior of the parameter space was studied by [16] and when it lies at the boundary of INAR and INGARCH model for the daily number of trades of six stocks listed in the NYSE Euronext group. The result shows that Poisson with Quasi-Maximum Likelihood Estimation provides a general approach for estimating the conditional mean parameters of time series of counts. The INGARCH model with poisson and negative binomial distribution was applied by [20] to forecast the future trends of target keywords of Apple in order to know the future technology of Apple. From the results of the Apple case study and they found that which technological keywords are more important or critical in the entire structure of Apple's technologies. Based on our review, there is none of the studies that applies Poisson INGARCH and NB-INGARCH in asthma data. Therefore, with this limitation, we aim to explore this approach in our data.

3 Study Area

The study was carried out in Johor Bahru, the third metropolitan city in Malaysia. This city with an area of 220 km^2 and 497,067 residents is a vital industrial sector and is transacted by the most important and busiest highway in Malaysia. It was located at the strategic commercial centre at the growth triangle of Indonesia–Malaysia–Singapore. Tertiary-based industry dominates the economy with many international tourists from the regions visiting the city. It is the centre of commerce and retail, financial services, hospitality, arts and culture, urban tourism, electrical and electronics, food processing and plastic manufacturing. The Industrial development brought about drastic changes in the social and economic lives of Johor Bahru's residents. The benefits of the development are felt even today, but so are the adverse effects. The rapid industrialization brought about industrial pollution. The effects of industrial pollution are vast, causing water contamination, a release of toxins into the soil, the air and environmental disasters. Air pollution that caused by the smoke released by various industries

has been the culprit for many illnesses such as asthma, respiratory problems, conjunctivitis and others. Therefore it is essential to study the behaviour of asthma disease in Johor Bahru. Figure 1 shows the map of Johor Bahru

Figure 1: Map of Johor Bahru

4 Methodology

4.1 Poisson INGARCH and NB-INGARCH

The asthma data is in the form of count or discrete which the observations can take only the non-negative integer values such as 0, 1, 2, 3 and so forth. These integers ascend from counting rather than ranking which is regularly associated with Poisson or Negative binomial distribution. Therefore in this study our aim is to explore on count data approach in order to analyse the asthma data. In this study, Poisson INGARCH and Negative Binomial INGARCH will be applied to model the asthma data. We consider the Poisson INGARCH (p, q) or Negative Binomial INGARCH (p, q) as introduced by [21] as:

$$Y_t | \mathcal{F}_{t-1} \sim Poisson(\lambda_t)$$

$$\text{or } Y_t | \mathcal{F}_{t-1} \sim NegBin(\lambda_t, \phi)$$

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^{p} \beta_k g(Y_{t-i_k}) + \sum_{l=1}^{q} \alpha_l g(\lambda_{t-j_l}) + \eta^{\mathrm{T}} \mathbf{X}_t + \sum_{m=1}^{s} \omega_m \delta_m^{t-\tau_m} \mathrm{I}(t \geq \tau_m)$$

where $g : \mathbb{R}^+ \to \mathbb{R}$ is a link function and $g : \mathbb{N}_0 \to \mathbb{R}$ is a transformation function. The parameter vector $\eta = (\eta_1, \ldots\ldots.., \eta_r)^{\mathrm{T}}$ corresponds to the effects of covariates. $\{\mathbf{X}_t : t \in \mathbb{N}\}$ is a time-varying r-dimensional covariate vector, say $\mathbf{X}_t = (\mathrm{X}_{t,1}, \ldots\ldots \mathrm{X}_{t,r})^{\mathrm{T}}$. To allow for regression on arbitrary past observations of the response, define a set $P = \{i_1, i_2, \ldots\ldots, i_p\}$ and integers $0 < i_1 < i_2 \ldots. < i_p < \infty$ with $p \in \mathbb{N}_0$. This enables us to regress on the lagged observations $Y_{t-i_1}, Y_{t-i_2}, \ldots\ldots., Y_{t-i_p}$. Analogously, define a set $Q = \{j_1, j_2, \ldots\ldots, j_q\}$, $q \in \mathbb{N}_0$ and integers $0 < j_1 < j_2 \ldots. < j_q < \infty$, for regression on lagged conditional means $\lambda_{t-j_1}, \lambda_{t-j_1}, \ldots\ldots., \lambda_{t-j_q}$. Meanwhile, β_0, β_1, and α_1 are the model's parameter and ϕ is the dispersion parameter. The model with logarithmic link function should be added with 1 to each observation for avoiding zero values. Example of NB-INGARCH(1,1) model with the logarithmic link function:

$$Y_t | \, \mathcal{F}_{t-1} \sim NegBin\,(\lambda_t, \phi)$$

$$\log(\lambda_t) = \beta_0 + \beta_1 \log(Y_{t-1} + 1) + \alpha_1 \log(\lambda_{t-1}), \quad t = 1, \ldots, n$$

In many applications sudden changes or outbreak occur. Thus, the term $\sum_{m=1}^{s} \omega_m \delta_m^{t-\tau_m} \mathbb{I}\,(t \geq \tau_m)$ refers to the intervention effect will be added in the model, where ω_m, $m = 1, \ldots, s$ are the intervention sizes, δ is the decay rate and $\mathbb{I}$ is the outbreak occur at the time of occurrence, τ.

4.2 Quasi Maximum Likelihood Estimation (QMLE)

The estimation of the model's parameter was determined by quasi-conditional maximum likelihood estimation (QMLE) as explained by [16]. Denoted by $\theta = (\beta_0, \beta_1, \ldots., \beta_p, \alpha_1, \ldots.., \alpha_q, \eta_1, \ldots., \eta_r)^T$ is the vector of regression parameters and the parameter space for the INGARCH model regardless of the distributional assumption which is given by:

$$\Theta = \left\{ \theta \in \mathbb{R}^{p+q+r+1} : \beta_0 > 0, \beta_1, \ldots., \beta_p, \alpha_1, \ldots.., \alpha_q, \eta_1, \ldots., \eta_r \geq 0, \sum_{k=1}^{p} \beta_k + \sum_{l=1}^{q} \alpha_l < 1 \right\}.$$

The intercept β_0 is essential to be positive while all other parameters must be nonnegative to ensure positivity of the conditional mean, λ_t. For the log-linear model, the parameter space is taken to be

$$\Theta = \left\{ \theta \in \mathrm{R}^{p+q+r+1} : |\beta_1|, \ldots., |\beta_p|, |\alpha_1|, \ldots.., |\alpha_q| < 1, \left| \sum_{k=1}^{p} \beta_k + \sum_{l=1}^{q} \alpha_l \right| < 1 \right\}.$$

According to [22], the estimation of the negative binomial parameter does not rely on the additional dispersion parameter, ϕ. This allows utilizing a quasi-maximum likelihood approach based on the Poisson likelihood to estimate the regression parameter, θ. The QMLE approach is preferred for simplicity and its practicality on deriving consistent estimators when the model for λ_t has been correctly specified. The conditional quasi log-likelihood function up to a constant is as follow:

$$\ell\,(\theta) = \sum_{t=1}^{n} \log p_t\,(y_t; \theta) = \sum_{t=1}^{n} \left(y_t ln\,(\lambda_t\,(\theta)) - \lambda_t\,(\theta) \right)$$

where $p_t(y;\theta) = \mathrm{P}(Y_t = y\,|\mathcal{F}_{t-1})$ is the probability density function of a Poisson distribution. In the case of the poison assumption it holds $\sigma^2 = 0$ and in the case of the negative binomial assumption $\sigma^2 = 1/\phi$, where σ^2 is variance and ϕ is dispersion parameter. The QMLE of θ is the solution of the non-linear constrained optimization problem;

$$\hat{\theta} = \arg max_{\theta\in\Theta}\ell(\theta).$$

For NB-INGARCH, the σ^2 is related to the dispersion parameter ϕ of the negative binomial distribution by $\phi = 1/\sigma^2$. The dispersion parameter ϕ of the negative binomial distribution is estimated by solving the equation:

$$\sum_{t=1}^{n}\frac{\left(Y_t-\hat{\lambda}_t\right)^2}{\hat{\lambda}_t+\hat{\lambda}_t^2/\hat{\phi}} = n-(p+q+r+1).$$

4.3 Model Assessment

The assessment of the model has been evaluated from Pearson residuals, the probability integral transform (PIT) histogram, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). Pearson residuals was explained by [23] as:

$$r_t^P = \left(y_t-\hat{\lambda}_t\right)/\sqrt{\hat{\lambda}_t+\hat{\lambda}_t^2\hat{\sigma}^2}.$$

For $t = 1, \ldots.., n$ the empirical autocorrelation function (ACF) of these residuals is valuable for diagnosing serial dependency. A plot of residuals versus time can reveal changes of the data generating process over time. Meanwhile, to access the probabilistic calibration of the predictive distribution, the probability integral transform (PIT) is used. PIT can be given by:

$$F_t(u|\,y) = \begin{cases} 0, u \le P_t(y-1) \\ \dfrac{u-P_t(y-1)}{P_t(y)-P_t(y-1)} \\ 1, u \ge P_t(y) \end{cases} , P_t(y-1) < u < P_t(y).$$

The mean PIT is then given by:

$$\overline{F}(u) = \frac{1}{n}\sum_{t=1}^{n}F_t(u|y_i), \quad 0 \le u \le 1..$$

order to examine whether $\overline{F}$ is the cumulative distribution function of a uniform distribution or not, [24] suggested to plot a histogram with H bins, where bin h has the height of

$$f_j = \overline{F}(h/H) - \overline{F}((h-1)/H),\ h, \ldots, H.$$

A U-shape the histogram designates an under-dispersion of the predictive distribution whereas an upside-down U-shape of histogram designates an over-dispersion. Other popular tools of the model selection criteria are the log-likelihood value, Akaike's information criterion (AIC) and

the Bayesian information criterion (BIC). The model with the lowest value of the respective information criterion is preferable. According to [23], the AIC and BIC are given by:

$$AIC = -2\hat{\ell}(\hat{\theta}, \hat{\sigma}^2) + 2df$$

$$BIC = -2\hat{\ell}(\hat{\theta}, \hat{\sigma}^2) + \log(\eta_{eff})df,$$

where df is the total number of parameters and η_{eff} is the number of effective observations

4.4 Performance Measures

The asthma data was divided into two parts; in-sample (training data) and out-sample (testing data). The period of in-sample data is from 2012 to 2014 which is used as model fitting and out-sample data is 1 year period (2015) which is used to validate the forecasting accuracy of the model. The performance measures are based on MAPE and RMSE that are given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)i}, \quad MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}\right|}{}i,$$

where Y_i is the predicted data, $\hat{Y}_i$ is the actual data and n is the length of the data.

5 Result and Discussion

5.1 Statistical Descriptive

In this study, we analyse the weekly number of asthma patients in Johor Bahru, Malaysia. The weekly data was obtained from the Health District Department of Johor Bahru. The data used for the study is from January 2012 to December 2015. The time series plot of asthma diseases is shown in Figure 2.

Figure 2 shows the time series plot of the number of asthma patient versus time (week) in Johor Bahru, Malaysia. There are two sudden changes from the plot in July 2012 and July 2013. This outbreak represents the highest number of asthma cases for the two consecutive years. One of the reasons that may contribute to this outbreak is the extremely bad air pollution due to Indonesian forest fires. Therefore, this intervention effect will be incorporated into our models. The descriptive statistics of asthma data is shown in Table 1.

Table 1: Descriptive Statistics of Asthma Data

Mean	**Median**	**Std deviation**	**Skewness**	**Kurtosis**
18	17	8.80594	1.090045	4.796721

The above descriptive statistics revealed the average number of people who are diagnosed with asthma in Johor Bahru is 18 people per week. The median of asthma data is 17 is almost equal to the mean. This means that the distribution can be assumed to be approximately symmetrical. The standard deviation of 8.8059 indicating the summary measure

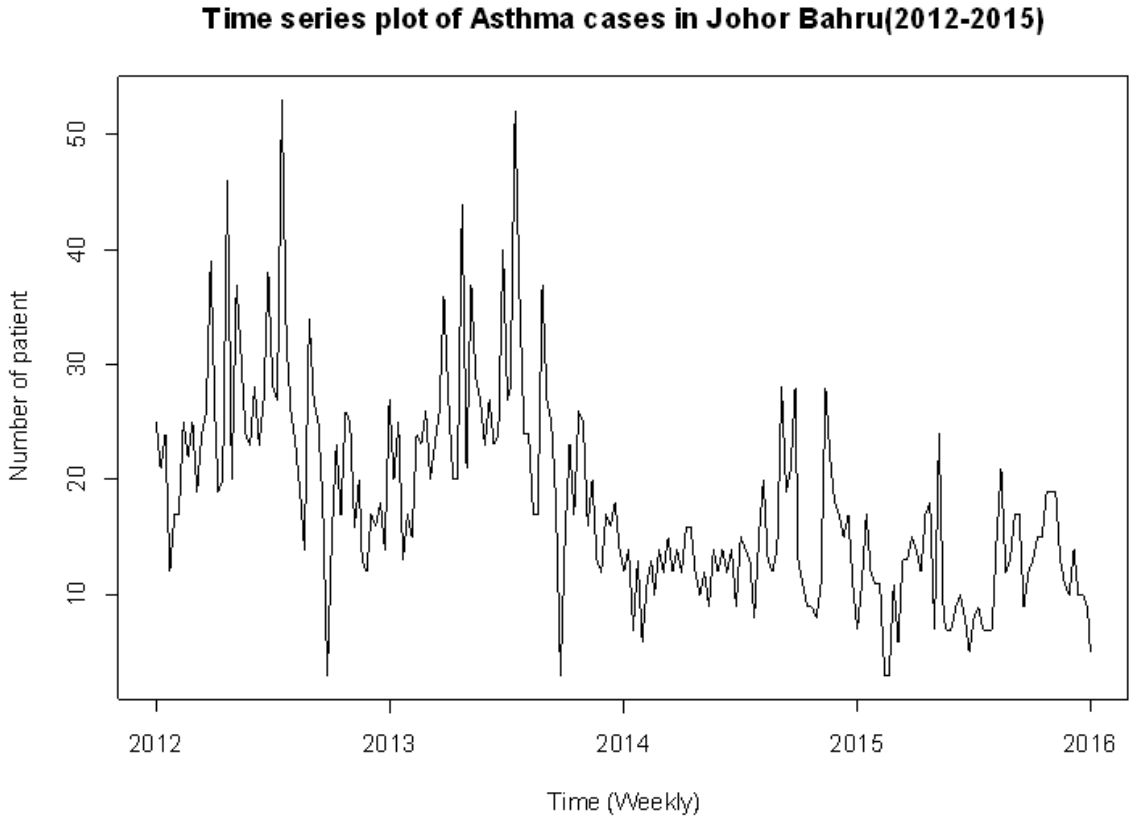

Figure 2: The Time Series Plot of Asthma Diseases in Johor Bahru for the Period of 2012 to 2015

of the differences of each observation from the mean of asthma diseases. Another important aspect in scrutinizing the behaviour of the data are skewness and kurtosis. These are the moment parameters that offer relevant information on the symmetry and shape of the related distribution [24]. The asthma data shows positive skewness with 1.0900 and it has a long tail in the positive direction. Kurtosis value show 4.7967 which is greater than 3. This indicates that the data has tails that asymptotically approach zero more slowly than a Gaussian and therefore produces more outliers than the normal distribution.

5.2 Modelling Asthma with Poisson-INGARCH and NB-INGARCH

We fit our data with Poisson INGARCH and NB-INGARCH with identity and log link function. We include a regression on the previous observation to integrate the serial dependence in the model and we also include two intervention effects in the model which happened at the period of July 2012 and July 2013. These outbreaks may be due to bad air pollution because of forest fires in Indonesian that usually happen around June to September. The climate change may also contribute to the increment of asthma diseases. Parsimonious parameterized Poisson INGARCH (1, 6) and NB-INGARCH (1,6) which is models with order 1 and 6 was chosen based on the lag spike of the autocorrelation function (ACF) of the asthma data. The parameter estimation of the models was determined by QMLE and Table 2 show the summary of the results.

Table 2: Parameter Estimation of Poisson INGARCH and NB-INGARCH

Link function	**Model**		β_0	β_1	β_6	α_6	I_1	I_2	σ^2
Identity	Poisson-INGARCH(1,6)	Estimation	5.44	0.352	0.386	$3{,}42\text{x}10^{-10}$	$4.41\text{x}10^{-8}$	24.2	
		Std. error	2.1350	0.0534	0.0666	0.1117	1.1352	7.1318	-
	NB-INGARCH	Estimation	4.56	0.369	0.412	$4.44\text{x}10^{-9}$	$2.01\text{x}10^{-4}$	10.4e	0.0571
		Std. error	3.1946	0.0832	0.1001	0.1692	1.7090	10.7060	-
Log	Poisson-INGARCH	Estimation	0.9092	0.3545	0.4144	-0.0557	-0.0699	0.7070	-
		Std. error	0.3033	0.0560	0.0673	0.1079	0.0535	0.1435	-
	NB-INGARCH	Estimation	0.9092	0.3545	0.4144	-0.0557	-0.0699	0.7070	0.0584
		Std. error	0.4480	0.0838	0.0981	0.1601	0.0830	0.2831	-

Therefore, based on parameter estimation above, the fitted Poisson-INGARCH (1,6) model with identity link function for asthma disease Y_t in time period t can be written as:

$$Y_t|\,\mathcal{F}_{t-1} \sim Poisson\,(\lambda_t)$$

$$\begin{aligned}\lambda_t = {} & 5.44 + 0.352Y_{t-1} + 0.386Y_{t-6} + (3.42 \times 10^{-10})\,\lambda_{t-6} + (4.41 \times 10^{-8})\,\mathrm{I}_1\,(t = 23) \\ & + 24.2\mathrm{I}_2\,(t = 29)\,, \quad t = 1, \ldots, n\end{aligned}.$$

Whereas the fitted Poisson-INGARCH (1,6) model with log link function for asthma disease Y_t in time period t is given by:

$$Y_t|\,\mathcal{F}_{t-1} \sim Poisson\,(\lambda_t)$$

$$\begin{aligned}\log(\lambda_t) = {} & 0.9092 + 0.3545\log(Y_{t-1} + 1) + 0.4144\log Y_{t-6} - 0.0557\log(\lambda_{t-6}) \\ & - 0.0699\mathrm{I}_1\,(t = 23) + 0.7070\mathrm{I}_2\,(t = 29)\,, \quad t = 1, \ldots, n\end{aligned}.$$

The fitted NB-INGARCH(1,6) model with identity link function for asthma disease Y_t in time period t and over dispersion coefficient, $\sigma^2 = 5.71 \times 10^{-2}$where $(\phi = 1/\sigma^2)$ is given by:

$$Y_t|\,\mathcal{F}_{t-1} \sim NegBin\,(\lambda_t, 17.51)$$

$$\begin{aligned}\lambda_t = {} & 4.56 + 0.369Y_{t-1} + 0.412Y_{t-6} + (4.44 \times 10^{-9})\,\lambda_{t-6} + (2.01 \times 10^{-4})\,\mathrm{I}_1\,(t = 23) \\ & + 10.4\mathrm{I}_2\,(t = 29)\,,\ t = 1, \ldots, n\end{aligned}.$$

The fitted NB-INGARCH(1,6) model with log link function for asthma disease Y_t in time period t and over dispersion coefficient, $\sigma^2 = 0.0584$ where $(\phi = 1/\sigma^2)$ given by:

$$Y_t|\,\mathcal{F}_{t-1} \sim NegBin\,(\lambda_t, 17.12)$$

$$\begin{aligned}\log(\lambda_t) = {} & 0.9092 + 0.3545\log(Y_{t-1} + 1) + 0.4144\log Y_{t-6} - 0.0557\log(\lambda_{t-6}) \\ & - 0.0699\mathrm{I}_1\,(t = 23) + 0.7070\mathrm{I}_2\,(t = 29)\,, \quad t = 1, \ldots, n\end{aligned}.$$

5.3 Diagnostic Checking

The diagnostic plot of Poisson INGARCH (1,6) with identity link function displays in Figure 3. Consequently, the empirical autocorrelation function (ACF) of Pearson residuals as shown in Figure 3(a) does not exhibit any serial correlation which indicates an adequate model. Meanwhile, the cumulative periodogram of Pearson residuals exhibits the frequency of the residuals in the acceptable range. The dashed lines give approximately 95% confidence limits of a Kolmogorov-Smirnov test on a constant spectral density which are used as a graphical checking for uncorrelated residuals. This indicates that this model is appropriate. However, Figure 3(c) shows Probability Integral Transform (PIT) histogram with no uniform shape indicating that the residuals are not normal and Poisson distribution is not adequate for model fitting. Hence, Poisson INGARCH (1,6) with identity link function is unable to describe the observed autocorrelation structure and cannot elucidate the volatility of the asthma data.

The diagnostic plot for Poisson INGARCH (1,6) with log link function shows in Figure 4. The empirical autocorrelation function (ACF) and cumulative periodogram of Pearson residuals exhibit the same results as Poisson INGARCH with identity link function. However, the PIT histogram tends to perform an upside-down U-shape indicates over dispersion are not taking account in the data. Thus this model is not adequate in representing the asthma data. Therefore, to overcome the shortcoming, we try to fit the asthma data with NB-INGARCH.

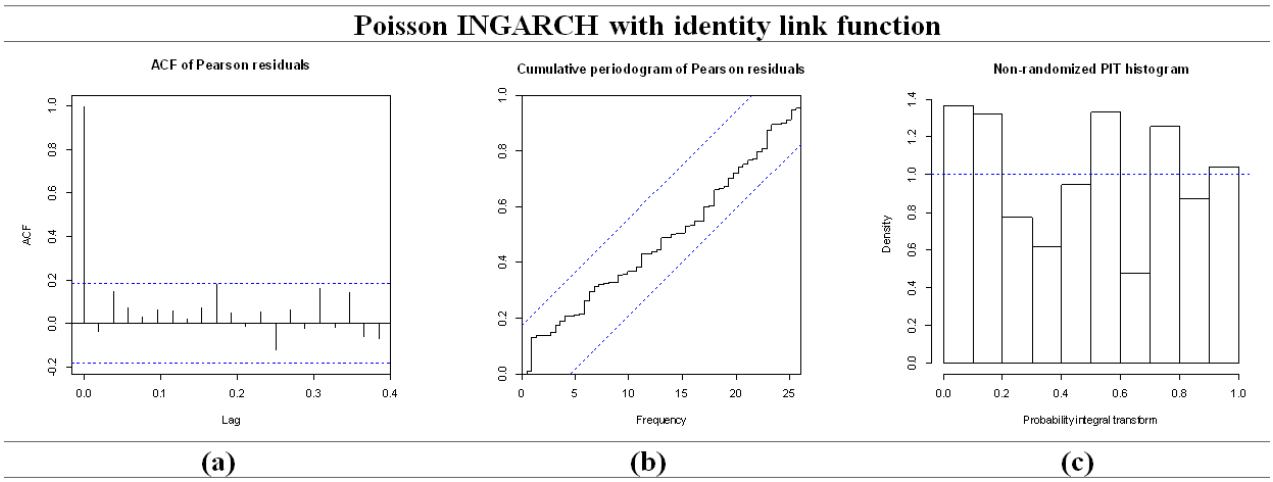

Figure 3: Diagnostic Checking Based on (a) Pearson Residuals, (b) Cumulative Periodogram and (c) PIT

Based on the previous study, NB-INGARCH was proven able to capture the over dispersion that exists in the data [5]. The conditional dispersion parameter was taking into account in NB-INGARCH model. Figure 5 and Figure 6 shows the diagnostic plot of NB-INGARCH (1,6) with identity link function and log link function. The PIT histogram corresponds to the negative binomial distribution appears to approach uniformity better than Poisson distribution. Therefore, the probabilistic calibration of the negative binomial model is adequate.

Both NB-INGARCH models are adequate in modelling asthma data. Models with additional dispersion parameter lead to a considerable improvement compared to the Poisson INGARCH (1, 6). We proceed with another diagnostics tools; the log likelihood, AIC, and BIC to select the best model. Table 3 shows the value of log likelihood, AIC and BIC. The higher the loglikelihood value, the lowest AIC and BIC will be preferred in the model selection. Based on table 3, NB-INGARCH with identity link function shows the higher value of log likelihood and lowest value of AIC and BIC. Thus, as compared with other models, NB-INGARCH with identity link function is outperformed.

Table 3: The Log Likelihood, AIC and BIC Value of the Models

Model	**Loglikelihood**	**AIC**	**BIC**
Poisson INGARCH with identity link function	-402.002	816.004	832.5256
Poisson INGARCH with log link function	-404.995	821.99	838.5115
NB- INGARCH with identity link function	-384.7532	783.5064	802.7815
NB- INGARCH with Log link function	-385.6438	785.2875	804.5626

Based on the diagnostic checking, from the four models, only two models are adequate namely NB-INGARCH with identity link function and NB- INGARCH with Log link function. However in terms of log likelihood, AIC and BIC NB- INGARCH with identity link function performed better.

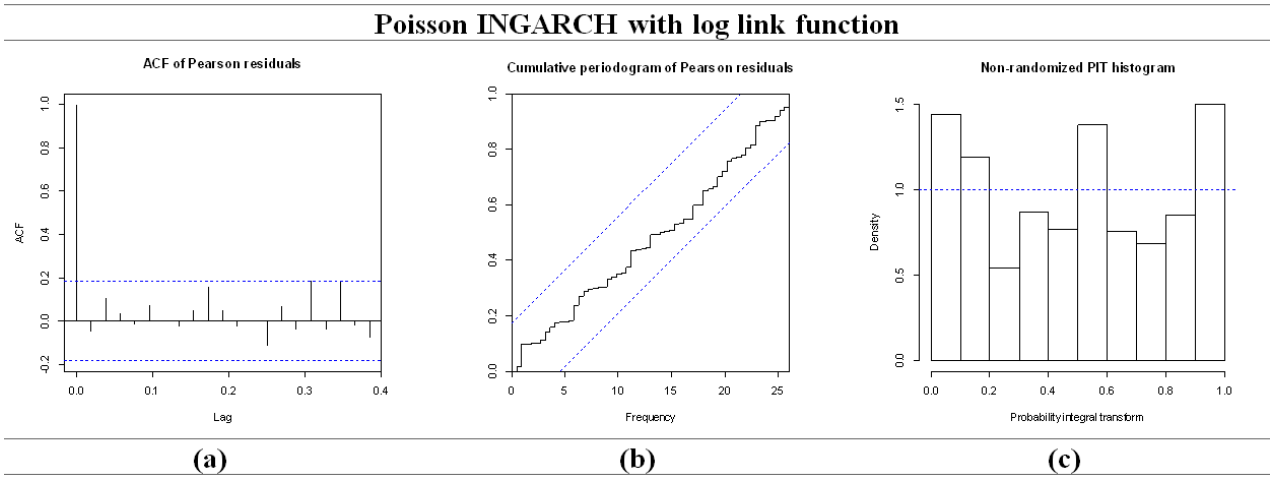

Figure 4: Diagnostic Checking Based on (a) Pearson Residuals, (b) Cumulative Periodogram and (c) PIT

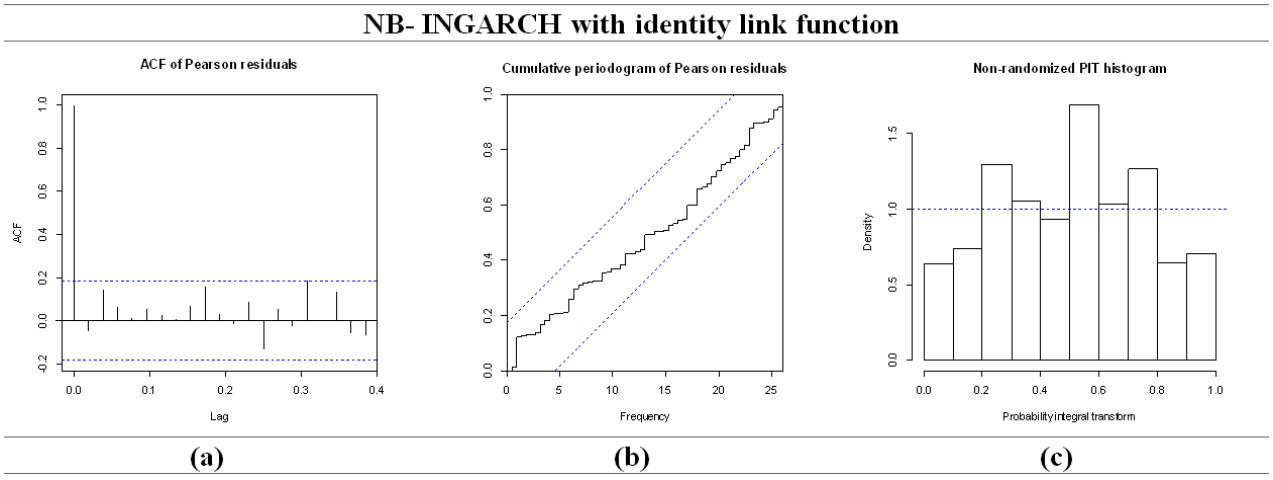

Figure 5: Diagnostic Checking Based on (a) Pearson Residuals, (b) Cumulative Periodogram and (c) PIT

5.4 Performance Measures

To reconfirm this, the performance measures were evaluated based on RMSE and MAPE value. These measurements are used to determine the ability of the models to forecast in-sample and out-sample data. Table 4 shows the performance measure of the models.

As we can see, NB- INGARCH with identity link function exhibits lowest RMSE and MAPE in out sample data as compare with NB- INGARCH with Log link function. This means that NB- INGARCH with identity link function is able to capture the real observation better than NB- INGARCH with log link function. Incorporating negative binomial distribution in the INGARCH model is satisfactory as the over dispersion is taken into account in the model. This finding motivates our contribution, in the sense that PIT based on negative binomial distributions shows the better fit of the data than the corresponding PIT based on Poisson. This study at the same time could solve the limitation found in the previous study, as proposed

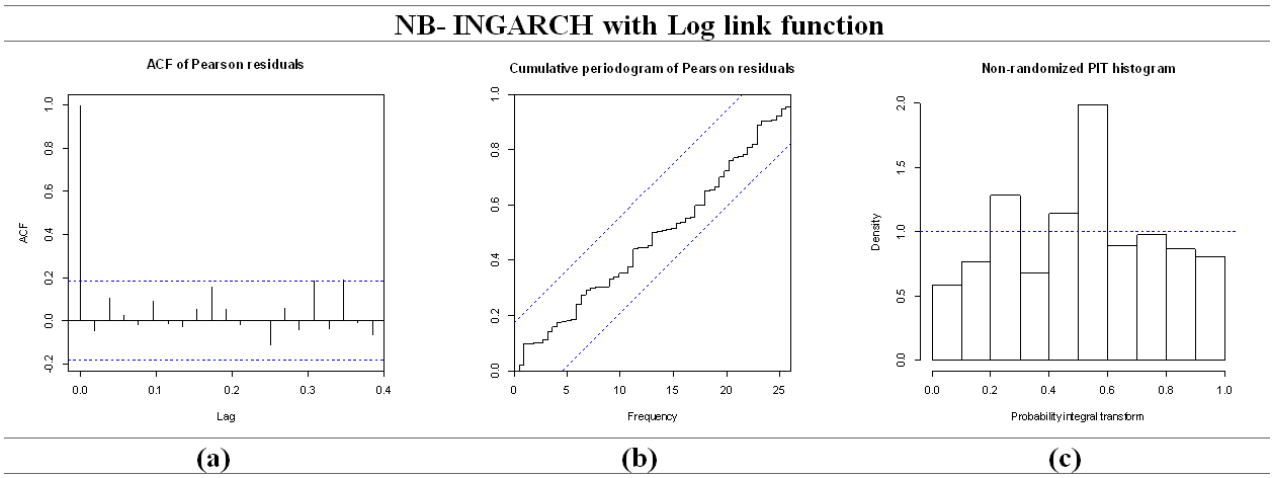

Figure 6: Diagnostic Checking Based on (a) Pearson Residuals, (b) Cumulative Periodogram and (c) PIT

Table 4: Performance Measures of the Models

Model	**In-sample**		**Out-sample**	
	RMSE	MAPE	RMSE	MAPE
NB- INGARCH with identity link function	6.9546	0.3153322	8.542325	0.958437
NB- INGARCH with Log link function	6.8292	0.3172256	13.20872	1.455826

by [9]. Another advantage of this study is by using QMLE we avoid the complicated likelihood functions and at the same time it is still possible to obtain consistent estimators whose standard error can be robustly estimated [15].

This is the initial finding of the asthma analysis. Further analysis is still required to improve the performance of the models especially the accumulation of the covariates or exogenous variables in the models. Hopefully, this finding can be a benchmark for any asthma time series model and provide some knowledge to the authorities especially the health sector in Johor Bahru. For future research we intend to incorporate the influential factors as the covariates or exogenous variables such as pollution and climate change such as temperature, rainfall and humidity in the model.

6 Conclusion

Count analysis approach has been applied in modelling and forecasting the number of weekly asthma patients in Johor Bahru for the period of 2012 to 2015. The drawback of the previous study by [9] is the model proposed failed to fulfill the normality assumptions Thus dealing with count or discrete data approach, Poisson INGARCH and NB-INGARCH is able to overcome

the shortcoming. NB-INGARCH with identity link function is outperformed as compared with other models. Although our method yielded reasonably good results, further studies are still required to refine our method for more general cases of time series of counts. For example, as pointed out by one referee, extensions can be made to other models such as zero-inflated generalized Poisson and COMPoisson [14] study the outlier effect [25] and study the causality in INGARCH model [26]. This recommendation will be our intention in future research.

Acknowledgement

The authors would like to thank Universiti Teknologi Malaysia (UTM), Ministry of High Education (MOHE) and grant vote no: 20H14 for the financial funding.

References

[1] Amancio, C.T. & Nascimento, L.F.C. Asthma and air pollutants: a time series study. *Reviata de Assaciacao Medica Brasileira*. 2012. 58(3): 302-307

[2] Souza, A. De, Kofanovski, A. Z., Sabbah, I., Débora, A., & Santos, S. Allergy & Therapy asthma and environmental indicators: a time-series study. 2016. 7(1): 1–5.

[3] Xu, Z., Huang, C., Su, H., Turner, L.R. Qiao, Z. & Tong, S. Diurnal temperature range and childhood asthma: a time series study. *Environmental health*. 2013. 12(12): 1-14.

[4] Zhang, Y., Peng, L., Kan, H., Xu, J., Chen, R., Liu, Y. & Wang, W. Effects of meteorological factors on daily hospital admissions for asthma in adults: a time series analysis. *PLoS ONE*. 2014. 9(7): E102475

[5] Chen, K., Glonek, G., Hansen, A., William, S., Tuke, J., Salter, A. & Bi, P. The effect of air pollution on asthma hospital admissions in Adelaide, South Australia, 2003-2013: time series and case-crossover analyses. *Clinical & Experimental Allergy*. 2016. 47: 1-15

[6] Ding. L., Zhu, D., Peng, D. & Zhao, Y. Air pollution and asthma attacks in children: A case-crossover study analysis in the city of Chongqing, China. *Environmental Heatlh*. 2016. 220: 348-353.

[7] Guilbert, A., Cox, B., Bruffaerts, N., Hoebeke, L., Packeu, A., Henrick, M., Cremer, K.D., Bladt, S., Brasseur, O., & Nieuwenhuyse, A.V. Relationships between aeroallergen levels and hospital admissions for asthma in the Brussels-Capital Region: a daily time series analysis. *Environmental Health*. 2018. 17:35.

[8] Lenick, C.R.O., Wiquist, A., Chang, H.H., Kramer, M.R., Mullholland, J.A., Grundstein, A. & Sarnat, S.E. Evaluation of individual and area level factors as modifiers of the association between warm-season temperature and pediatric asthma morbidity in Atlanta, GA. *Environmental research*. 2017. 156: 132-144.

[9] Jamaludin, A. R., Yusof, F., Lokoman, R. M., Noor, Z. Z., Alias, N., Aris, N. M. Correlational study of air pollution-related diseases (asthma, conjunctivitis, URTI and dengue) in Johor Bahru, Malaysia. *Malaysian Journal of Fundamental and Applied Sciences*. 2017: 354-361.

[10] Alzahrani, N., Neal, P., Simon, E.F.S., McKinley, T.J., & Touloupou, P. Model selection for time series of count data. *Computational Statistics and Data Analysis*. 2018. 122: 33–44.

[11] Ferland, R., Latour, A. & Oraichi, D. Integer-valued GARCH process. *Journal of time series analysis*. 2006. 27(6): 923-942.

[12] Heinen, A. Modeling time series count data: An autoregressive conditional Poisson model. 2003. MPRA paper, University library of Munich, Germany.

[13] Rydberg, T. H. & Shephard, N. *A Modeling Framework for the Prices and Times of Trades Made on the New York Stock Exchange*. Nuffield College Working Paper. W99-14. 2000.

[14] Zhu, F. Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling*. 2011. 56(10): 191–203.

[15] Christou, V. & Fokianos, K. Quasi-likelihood inference for negative binomial time series models. *Journal of time series Analysis*. 2014. 35: 55-78.

[16] Ahmad, A. & Francq, C. Poisson QMLE of count time series models. *Journal of Time series Analysis*. 2016. 37: 291-314.

[17] Zhu, F. Modelling over dispersed or under dispersed count data with generalised Poisson interger-valued GARCH models. Journal of Mathematical Analysis and Applications. 2012. 389: 58-71.

[18] Christou, V. & Fokianos, K. On count time series prediction. *Journal of Statistical Computation and Simulation*. 2013. 2: 1-18

[19] Liboschik, T., Kersche, P., Fokianos, K. & Fried, R. Modelling interventions in INGARCH processes. *International Journal of Computer Mathematics*. 2014. 93(4): 640–657.

[20] Relationship between aeroallergen levels and hospital admissions for asthma in the Brussels-capital region: a daily time series analysis. *Environmental Heatlh*. 2018. 17(35): 1-12

[21] Liboschik, T., Fokianos, K. & Fried, R. Tscount: An R package for analysis of count time series following Generalised Linear Models. *Journal of statistical software*. 2017.

[22] Tobias, M., Ekkehard, G., Heinz, S. & Tim, F. Group Sequential Designs for Negative Binomial Outcomes. Cornell University Library. 2017.

[23] Czado, C., Gneiting, T. & Held, L. Predictive model assessment for count data. *Biometrics*. 2009. 65: 1254-1261

[24] Goncalves, E., Lopes, N.M. & Silva, F. Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis*. 2015. 36(4): 503-527

[25] Kitromilidou and Fakianos, K. Robust estimation methods for a class of log-linear count time series models. *Journal of Statistical Computational Simulation*. 2016. 86: 740-755.

[26] Lee, Y. & Lee, S. On casuality test for time series of counts based on Poisson INGARCH models with application to crime and temperature data. *Communications in Statistics-Simulation and Computation*. 2018. 1-11