Pak.j.stat.oper.res. Vol.19 No. 3 2023 pp 425-446 DOI: http://dx.doi.org/10.18187/pjsor.v19i3.4123

Pakistan Journal of Statistics and Operation Research

Modeling the Asymmetric Reinsurance Revenues Data using the Partially Autoregressive Time Series Model: Statistical Forecasting and Residuals Analysis

Salwa L. Alkhayyat[1,2], Heba Soltan Mohamed[3],
Nadeem Shafique Butt[4], Haitham M. Yousof[5,*] and Emadeldin I. A. Ali[6,7]

* Corresponding Author

[1]Department of Statistics, Faculty of Science, University of Jeddah, Kingdom of Saudi Arabia; slalkhayyat@uj.edu.sa
[2]Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Kafr El-Sheikh University, Egypt.; salwa.elkhayat@com.kfs.edu.eg
[3]Department of Statistics and Quantitative Methods, Faculty of Business Administration, Horus University, Damietta, Egypt; Hsultan@horus.edu.eg
[4]Department of Family and Community Medicine, King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia; nshafique@kau.edu.sa
[5]Department of Statistics, Mathematics and Insurance, Benha University, Egypt; haitham.yousof@fcom.bu.edu.eg
[6]Department of Economics, College of Economics and Administrative Sciences, Al Imam Mohammad Ibn Saud Islamic University, Saudi Arabia; EIALI@IMAMU.EDU.SA
[7]Department of Mathematics, Statistics, and Insurance, Faculty of Business, Ain Shams University, Egypt; i_emadeldin@yahoo.com

Abstract

The autoregressive model is a representation of a certain kind of random process in statistics, insurance, signal processing, and econometrics; as such, it is used to describe some time-varying processes in nature, economics and insurance, etc. In this article, a novel version of the autoregressive model is proposed, in the so-called the partially autoregressive (PAR(1)) model. The results of the new approach depended on a new algorithm that we formulated to facilitate the process of statistical prediction in light of the rapid developments in time series models. The new algorithm is based on the values of the autocorrelation and partial autocorrelation functions. The new technique is assessed via re-estimating the actual time series values. Finally, the results of the PAR(1) model is compared with the Holt-Winters model under the Ljung-Box test and its corresponding p-value. A comprehensive analysis for the model residuals is presented. The matrix of the autocorrelation analysis for both points forecasting and interval forecasting are given with its relevant plots.

Keywords: Time series; Statistical model; Forecasting; Residual analysis; Ljung-Box test; Simulation; Statistics and numerical data.

1. Introduction

Reinsurance refers to the coverage a company buys from another insurance company to protect itself (at least in part) from the possibility of a significant claim happening. Reinsurance is also referred to as stop-loss insurance or insurance for insurers. Reinsurance is the process through which insurers, through arrangement, transfer some of their risk portfolios to other parties in an effort to lessen the possibility that they could have to shoulder a sizable liability as a result of an insurance claim. The individual (company) who agrees to take on a portion of the possible responsibility in exchange for a piece of the insurance premium is known as the reinsurer. The ceding party is known as the diversified insurance portfolio party. Through reinsurance, insurers can continue to turn a profit by recouping all or part of the funds paid to claimants. In the event of significant or widespread losses, reinsurance offers protection from catastrophic losses and lowers the net liability for individual risks. The process also enables reinsurance seekers, or ceding businesses, to enhance their capacity for taking on all types of risks. In other terms, the reinsurer is the business that issues the reinsurance coverage. Reinsurance typically allows insurance firms to continue operating after major claims occurrences, such as severe catastrophes like hurricanes and wildfires. Reinsurance occasionally fulfils functions unrelated to its primary function in risk management, such as lowering the capital needs of the cascading company, tax mitigation, or other functions.

Skewed/asymmetric real data sets are those that have an asymmetric distribution, where the data points are not evenly distributed around the mean. Such data sets are common in many fields, including economics, finance, engineering, insurance, reinsurance, and other social sciences. Moreover, asymmetric data sets are often encountered in real-life scenarios, such as income distribution, sales data, or medical research. Analyzing such data sets can provide valuable insights into the underlying patterns and behaviors of the system under study. Skewed/asymmetric data sets are important in prediction and forecasting, as they often contain information about rare events or extreme values that can significantly impact the outcome. For example, in financial forecasting, the distribution of returns is often highly skewed, and analyzing this data can help predict the likelihood of extreme market events. For this purpose, we are motivated to analyze new asymmetric reinsurance data via a new modified time series model. This new model, as will be explained later, is capable of describing asymmetric time series data and dealing with it in analysis, segmentation, and future predictions. The ability of the model to do so stemmed from the new modified algorithms that were applied to estimate the model parameter completely differently from the well-known traditional methods. It is worth noting that the used reinsurance revenue data is skewed to the right (positive skew), while we may later find the reinsurance revenue data skewed to the left (negative skew), and due to the limitations of the model and the limitations of the study, we will rely solely on the skewed reinsurance revenue data left side. The new model's ability lies in the new algorithms, which do not require specific conditions in the data.

This paper provides a model called the partial autoregressive (PAR) model for estimating the revenues of reinsurance businesses. The new model stands out for its simple procedures and high degree of prediction accuracy. This model stands in contrast to several pricey ones with stringent requirements such as the autoregressive integrated moving average (ARIMA) models (see Box and Jenkins (1970) and Box et al. (2015)). One of the key areas of actuarial statistics is the forecasting of reinsurance businesses' revenue streams. There are numerous strategies, methods, and models for statistical prediction in statistical literature. It is challenging to declare one methodology, method, or model to be the finest of all time. Many other aspects and reasons, such as the needed accuracy, cost, and speed, affect this choice between approaches, methodologies, and models. To deliver trustworthy and accurate projections to an acceptable and sufficient degree, great care is devoted to balance all these aspects. Box and Jenkins methodology compares discrepancies between time series data points to determine results. By using seasonal differencing, moving averages models, and autoregressive models, patterns can be found that can be used to construct future projections. ARIMA models are used as the main implementation of the Box-Jenkins technique. The two names for these cars are occasionally used interchangeably. See Box et al. (2015) for additional information on time series analysis with future forecasting.

Nonetheless, research using ARIMA models was widely disseminated in the actuarial and statistical literature, see Cummins and Griepentrog (1985) for more results about combining the econometrics and the ARIMA models for predicting the costs of paid claims for vehicle insurance data, Jang et al. (1991) for analyzing and employing medical insurance plan based on the ARIMA model, Mohammadi and Rich (2013) for applying the ARIMA model to the dynamics of unemployment insurance claims, Hafiz et al. (2021) estimated the rate of the insurance penetration in Nigeria, and Kumar et al. (2020) for predicting the quantity of auto insurance claims using the ARIMA model. Several authors have focused on the autoregressive (AR), moving average (MA), and ARIMA models in the context of applied numerical techniques and future statistical prediction. For the prediction of electricity price see Jakaša et al. (2011). For forecasting oil seeds, see Darekar and Reddy (2017). For modelling total seeds of rice and wheat, see Sahu et al. (2015). For forecasting wheat production in India, see Nath et al. (2019)), identification of paddy crop phenological parameters see Palakuru et al. (2019), and Shrahili et al. (2021) for analyzing some left skewed insurance data via a new extension of the Chen model and via the AR model. Following Shrahili et al. (2021), this paper defines a new reinsurance claims PAR model as well as a new algorithm for estimating the model is presented and applied. To choose the optimum model, the technique essentially involves examining the reinsurance claims for each of the conceivable ARIMA models, this selection will depend on suitability for the reinsurance claims data, the significance of the model parameter is statistically checked. In general, it is better to use a model with fewer significant parameters. In order to create a particular Box-Jenkins model for the time series of reinsurance claims, it is necessary to first determine if the time series is stationary or not and whether any significant seasonality needs to be taken into account.

The Box-Jenkins model is identified before choosing the autoregressive model. The insurance claims are modelled using the PAR model. Its applicability is assessed through a few simulated exercises. The optimal parameter is chosen using artificial techniques. The new model was given this name because it involves a dynamic change in the method for estimating the model parameter. This change can be considered the development of the original AR model and the development of the traditional method popular in estimating the value of the model parameter. Following is a list of the main advantages of the updated model:

I. Since the future prediction process will be dependent on this ideal value, the upgraded model is depending on extensive simulation tests to select the best value for the model parameter. According to the methods we'll mention in their place, the statistical prediction procedure will take place.

II. For several businesses in the domains of manufacturing, marketing, insurance, engineering, and other industries, firms frequently want quick, accurate, and affordable forecasts for the near future. The proposed updated statistical model offers what companies and businesses require in the form of accurate predictions at the most affordable prices, and these predictions include qualities that are statistically significant.

III. This new model has been chosen precisely for its rapid statistical forecasting capabilities, not to mention how simple it is to use to statistical prediction as it is entirely devoid of the complexity that afflict so many contemporary models.

The remainder of the paper is structured as follows: The key statistical findings of the PAR model are presented in Section 2. Section 3 deals with evaluation and application to real data on reinsurance. Finally, Section 4 offers a few final remakes.

2.The PAR model: Structure

Before going into the details of the new model, a review of some details about the AR model is given. These details relate to the importance of the model and its applications in the field of insurance, reinsurance, economics, risk management, financing, agriculture and others. Here are some applications of AR models:

I. The AR models can be used to analyze stock market data and make predictions about future prices. By modeling the relationships between past prices and current prices, an AR model can provide valuable insights into market trends and volatility (see Idrees et al. (2019)).

II. The AR models can be used to forecast economic indicators such as inflation, and unemployment rates. By analyzing the historical data, an AR model can provide valuable insights into economic trends and help policymakers make informed decisions (see Bruneau et al. (2007)).

III. Due to Githeko and Ndegwa (2007), the AR models can be used to analyze and forecast climate data such as temperature, precipitation, and sea level. By modeling the relationships between past climate data and current data, an AR model can help researchers understand the factors that drive climate variability and change.

IV. The AR models can be used to forecast energy consumption and production, which is important for energy planning and policy making. By analyzing historical data, an AR model can provide insights into energy trends and help forecast future demand and supply (for more details, see Debnath and Mourshed (2018)).

V. According to Akbar (2019), the AR models can be used to forecast financial data such as stock prices, interest rates, and exchange rates. By modeling the relationships between past financial data and current data, an AR model can provide insights into market trends and help investors make informed decisions.

VI. The AR models can be used to analyze mortality data to understand trends and patterns in death rates over time (see Lee and Carter (1992)). By modeling the relationships between past mortality rates and current rates, an AR model can help actuaries estimate future mortality and assess the adequacy of reserves for life insurance and annuity products.

VII. Risk management, credit risk, insurance risk, and operational risk are just a few of the different sorts of hazards that the AR models can be used to forecast and model. An AR model can assist actuaries in estimating the likelihood and severity of possible losses and developing methods to reduce these risks by evaluating historical data (see Shrahili et al. (2021) and Mohamed et al. (2022a,b,c) for more details).

According to Yan and Hong (2015), reinsurance markets have received very little attention in empirical studies on asymmetric information, in contrast to primary insurance markets. By looking for a positive association between coverage and "ex post" risk in three significant reinsurance markets for the years 1995 to 2000, Yan and Hong (2015) looked into the possibility of asymmetric knowledge. They found that:

I. In the private passenger auto liability and homeowners' reinsurance markets, asymmetric information problems exist, but not in the market for product liability reinsurance.

II. Retention limits are frequently used to mitigate asymmetric information problems.

III. long-term contractual relationships are either rarely used or ineffective in controlling asymmetric information problems.

In several reinsurance markets, Yan and Hong (2015) found evidence of the use of retention restrictions to minimize asymmetric information issues, but little to no evidence of the use of long-term contractual ties to regulate asymmetric information. Additionally, Yan and Hong (2015) were unable to independently test for the existence of asymmetric information in the internal and external reinsurance markets due to a lack of data. In the actuarial literature, there are many useful works which deserve a huge attention and consultation in future works such as Lee and Carter (1992), Mohamed et al. (2022a,b,c), Yan and Hong (2015), Saikkonen and Teräsvirta (1985) among others.

These numerous uses gave the model significant weight in the statistical literature, and our decision to publish this new model was mostly driven by these uses. On the other hand, the ARIMA is a class of statistical models that forecasts future values by explaining a given time series based on its own previous values, such as its own lags and lagged forecast errors. Making the time series stationary is the first stage in creating a robust ARIMA model. The PAR model, where is the order of the PAR model, is a linear regression model that employs its own lags as predictors. When the predictors are unrelated and uncorrelated with one another, the linear regression models are suitable. One can analyze and ascertain the required number of AR terms by looking at the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. The PACF plots, however, are more precise than the ACF plots. Several simulated results are offered for the purpose of investigating the necessary number of AR terms. Following Shrahili et al. (2021), the new PAR model of order π(PAR(π)) can be expressed as

$$y_\varsigma = c + \varphi_1 y_{\varsigma-1} + \varphi_2 y_{\varsigma-2} \ldots + \varphi_\pi y_{\varsigma-\pi} + \varepsilon_\varsigma,$$

where ε_ς is the white noise, c is a constant, the lagged values of y_ς are the predictors, and $\varphi_1, \varphi_2, \ldots, \varphi_\pi$ are the unknown parameters. The autoregressive models are normally restricting to stationary data, where some constraints on the parameter values are required. For the PAR(1) model:

- when $\varphi_1 = 0$, then y_ς is equivalent to white noise model which is ARIMA model with parameters (0,0,0);
- when $\varphi_1 = 1$ and $c = 0$, then y_ς is equivalent to a random walk model with drift.
- when $\varphi_1 < 0$, then the ACF oscillates and has negative correlation at lag 1.

Further, one can write

$$y_\varsigma = c + \varphi_1 y_{\varsigma-1} + \varepsilon_\varsigma \mid -1 < \varphi_1 < 1, \forall\, \varsigma = 0, \pm 1, \pm 2, \ldots.$$

Here, the expected value of y_ς can be formulated as

$$E(y_\varsigma) = 0 | \varsigma = 0, \pm 1, \pm 2, \ldots,$$

and its variance can be expressed as

$$Var(y_\varsigma) = \gamma(0) | \varsigma = 0, \pm 1, \pm 2, \ldots,$$

where

$$\gamma(0) = \Delta(\varphi_1^2)\sigma_\varepsilon^2 | \Delta(\varphi_1^2) = \frac{1}{1-\varphi_1^2},$$

σ_ε^2 is the variance of the residuals, and the covariance $\mathrm{Cov}(y_\varsigma, y_{\varsigma-1})$ reduces to

$$Cov(y_\varsigma, y_{\varsigma-1}) = \Delta(\varphi_1^2)\varphi_1 \sigma_\varepsilon^2.$$

Analogously, the covariance $Cov\,(y_\varsigma, y_{\varsigma-2})$ has the form

$$Cov(y_\varsigma, y_{\varsigma-2}) = \Delta(\varphi_1^2)\varphi_1^2 \sigma_\varepsilon^2.$$

For the PAR(2) model (or the ARIMA(2,0,0) model)

$$y_\varsigma = c + \varphi_1 y_{\varsigma-1} + \varphi_2 y_{\varsigma-2} + \varepsilon_\varsigma|_{-1<\varphi_1<1, \varphi_2+\varphi_1<1 \text{ and } \varphi_2-\varphi_1<1}.$$

The PAR model in Equation (2) and some of its mathematical results will be used for statistical modeling of the claims payment data, and future prediction.

3. Assessment and forecasting

While Box-Jenkins models, also known as ARIMA models, are widely used and effective for time series analysis, they do have some disadvantages that should be considered. Here are the main disadvantages of Box-Jenkins models:

I. Stationarity assumption: Box-Jenkins models require the time series to be stationary, meaning that the mean, variance, and autocorrelation structure of the series remain constant over time. However, many real-world time series exhibit non-stationary behavior, such as trends or seasonality. In such cases, the series needs to be transformed or changed to achieve stationarity, which can complicate the modeling process and potentially introduce additional uncertainty.

II. Complexity and parameter estimation: Box-Jenkins models can become complex, particularly when dealing with higher-order autoregressive (AR) or moving average (MA) components. Estimating the parameters of these models can be challenging, especially when the dataset is small or contains missing values. The iterative process of identifying the appropriate model order and estimating parameters can be time-consuming and require significant computational resources.

III. Sensitivity to outliers: Box-Jenkins models can be sensitive to outliers or extreme values in the time series data. Outliers can significantly affect the model estimation process and bias the parameter estimates. Additionally, outliers might lead to incorrect model selection, resulting in suboptimal forecasting performance.

IV. Model selection: Box-Jenkins models involve selecting the appropriate model order, which includes determining the number of autoregressive (p), integrated (d), and moving average (q) terms. The model selection process typically relies on diagnostic tools, such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). However, selecting the optimal model order is not always straightforward, and different criteria may lead to different model choices. It requires careful judgment and expertise to strike the right balance between model complexity and goodness-of-fit.

V. Forecast uncertainty: While Box-Jenkins models can provide accurate forecasts under certain conditions, they may not fully capture the complexities and inherent uncertainty of real-world time series data. The assumption of constant parameters and linear relationships may limit their ability to accurately forecast in situations with nonlinear trends, sudden shifts, or structural breaks.

VI. Limited handling of complex patterns: Box-Jenkins models are primarily designed to capture linear dependencies and autocorrelations in time series data. They may struggle to effectively model complex patterns, such as long-term dependencies, nonlinear relationships, or non-Gaussian and heavy-tailed distributions. In such cases, alternative modeling techniques, such as state space models or machine learning approaches, may be more suitable.

VII. It's important to note that while Box-Jenkins models have these disadvantages, they still offer valuable insights and are widely used in time series analysis. However, it's essential to carefully consider the limitations and assess the suitability of these models for specific datasets and analysis goals.

Moreover, all Box-Jenkins models require numerical approximations of the solutions of a few nonlinear equations in order to estimate the parameters. It is typical to utilize statistical tools like R for this purpose. The nonlinear least squares and maximum likelihood methods are the two major methods for fitting all of these models. Although the second strategy is typically the most popular in statistical literature, it is excluded from our study since it is too complex for the full Box-Jenkins models (see Shrahili et al. (2021) and Mohamed et al. (2022a,b,c) for more details). Before dealing with the Box-Jenkins methodology (in general) and the PAC model (specifically) in prediction operations, it is inevitable to examine the behavior of the PACF, as the PACF plays a role no less important than the role of the ACF to use them in time series inactivity tests, as well as in identifying the appropriate model for time data. In modern analysis, the partial ACF takes various forms, just like the previously presented ACF. Sometimes it fades slowly, sometimes it fades quickly in an exponential form, and sometimes it gradually approaches zero in the form of waves of the sine function, or it breaks completely after a certain number of time gaps. Due to Makridakis and Hibon (1997), the following cases can be mentioned:

I. The PACF of the non-seasonal time series is completely cut off after a certain number of time slots.

II. It slowly fades into a calm exponential pattern.

III. It quickly fades into a fast exponential pattern.

IV. It approaches zero gradually in the form of sine function waves.

V. It gradually approaches zero in the form of a combination of exponential functions.

In this section, the results of the PAR model are compared with Holt-Winters model results under the Ljung-Box test and its corresponding p-value. There are two distinct iterations of the Holt-Winters method, each with a distinct seasonal component. The multiplicative technique is used when seasonal fluctuations change proportionally to the level of the series rather than when they are essentially constant across the course of the series. The Ljung-Box test determines whether an autocorrelation exists in a time series. The independent distribution of the residuals is hypothesized by H_0 (there is no autocorrelation in the data). The alternative theory is that the residuals are serially correlated and not independently distributed. The Ljung-Box test and its corresponding p-value are used for assessing the prediction accuracy (see Ljung and Box (1987)). The Ljung-Box test is a statistical test used to assess the presence of autocorrelation in a time series. It helps determine whether the residuals of a time series model are independent and uncorrelated. The test is based on the Ljung-Box statistic, which measures the overall autocorrelation in the residuals up to a certain lag. The null hypothesis of the Ljung-Box test is that there is no autocorrelation in the residuals (i.e., the residuals are independent). The alternative hypothesis suggests the presence of autocorrelation.

The p-value associated with the Ljung-Box test represents the probability of observing the computed test statistic (or a more extreme value) under the null hypothesis of no autocorrelation. If the p-value is below a specified significance level (e.g., 0.05), we reject the null hypothesis and conclude that there is evidence of autocorrelation in the residuals. In practice, the Ljung-Box test is often applied to a series of ACF values, and the p-values are calculated for each lag. The p-values are then examined to identify significant lags where autocorrelation is present. It's important to note that the Ljung-Box test is not the only test available for assessing autocorrelation in time series data. Other tests, such as the Durbin-Watson test or the Breusch-Godfrey test, may also be used depending on the specific requirements and characteristics of the data.

Given the importance of analyzing the residuals of any statistical model, it is preferred to return with the ACF and the PACF when analyzing errors and judging the adequacy and efficiency of the PAR model. The PAR model parameter estimation algorithm is based on a different approach from its traditional predecessor. These steps are formulated in algorithm 1, This algorithm is based mainly on the following steps:

Step 1. Calculate the autocorrelation coefficients (r_k) and partial autocorrelation coefficients (r_{kk}) for the maximum possible number of lags which are $n-1$.

Step 2. Use all autocorrelation coefficients and partial autocorrelation coefficients in evaluating the model and re-estimating the series values, such that:

$$\widehat{\varphi_1}|(-1<\varphi_1<1)=I_k|I_k=\{r_k,2r_k\},k=1,2,\dots,n-1,(-1<2r_k<1),$$

or

$$\widehat{\varphi_1}|(-1<\varphi_1<1)=J_{kk}|J_{kk}=\{r_{kk},2r_{kk}\},k=1,2,\dots,n-1,(-1<2r_{kk}<1).$$

Consequently, practitioner choose the unknown parameter values $\widehat{\varphi_1}$ that minimize the mean of square prediction errors (MSPE), where

$$\text{MSPE}=\sum_{t=0}^{T}\left(y_t-\hat{y}_{t|t-1}\right)^2.$$

Step 3. Use this value to predict new future observations.

Step 4. Evaluate predictions and analyze residuals to judge the adequacy and efficiency of the model.

This Section discusses the reinsurance revenue of a reinsurance company in the insurance industry in the United States. A monthly time series represents the reinsurance revenue. Fortunately, our data are recent time series data with a starting date of February 2015 and an ending date of April 2020. The data are available at: https://catalog.data.gov/datase, and for other new insurance data see Ibrahim et al. (2023), Khedr et al. (2023) and Yousof et al. (2023a,b,c). The information on reinsurance revenues must be looked at first. Both graphical and numerical approaches can be used to examine real data. A variety of graphical methods are offered for examining early fits of theoretical distributions including the normal, logistic, uniform, exponential, beta, lognormal, and Weibull, such as the skewness-kurtosis plot (also known as the Cullen and Frey plot). The bootstrapping findings are used and shown for greater accuracy. The picture by Cullen and Frey is a superb representation of the characteristics of the distributions, however it only compares distributions in the space of the squared skewness, kurtosis. Additional graphical tools include the "box plot" for finding the extreme reinsurance revenues and the "nonparametric Kernel density estimation (NKDE)" approach for assessing the initial shape of the empirical hazard rate function, and the "total time on test (TTT)" plot can be used for examining the initial shape of the empirical hazard rate function. The Cullen and Frey plot for the data of reinsurance revenues is shown in Figure 1. The NKDE plot for the data of reinsurance revenues is shown in Figure 2 (top left figure). Figure 2 (top right plot) displays the Q-Q plot for the data of the reinsurance revenues. The TTT plot for the data of the reinsurance revenues data is shown in Figure 2 (the bottom left plot), and Figure 2 (the bottom right figure) displays a box plot of the data for reinsurance revenues.

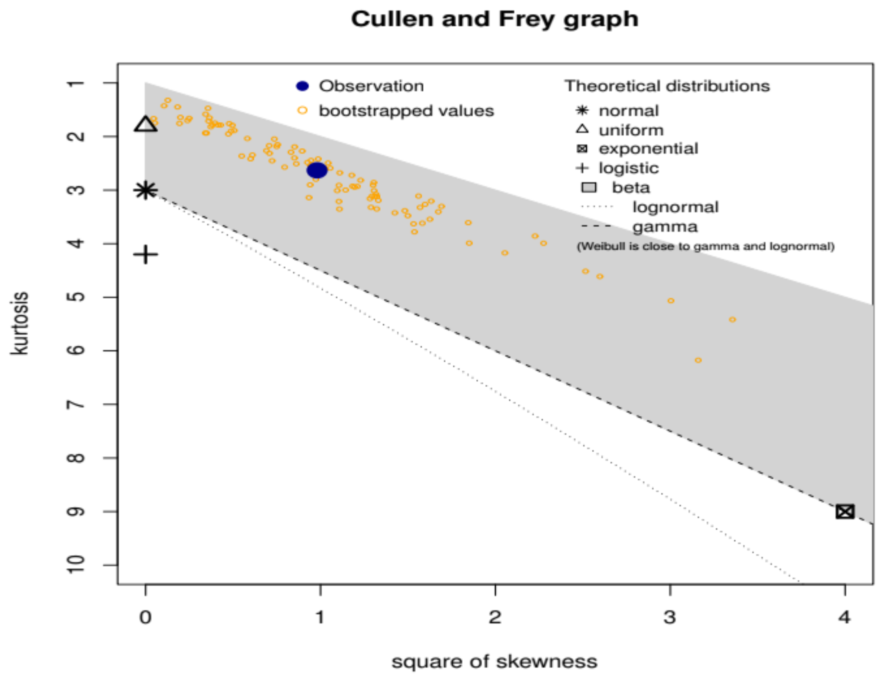

Figure 1: Cullen and Frey plot for the reinsurance claims data.

Due to Figure 2 (top left figure), the reinsurance revenues data are not very skewed and are close to being symmetric, they are of course not completely symmetrical, and this is confirmed by the results of Table 1, where the skewness=0.26682 and the kurtosis =2.26414. Table 1 presents a summery statistic for the reinsurance revenues data. No extreme observations are spotted based on Figure 2 (top right and bottom right plots) due to the reinsurance revenues data. Further, Figure 2 (bottom left plot) shows that the HRF for the reinsurance revenues data is monotonically increasing.

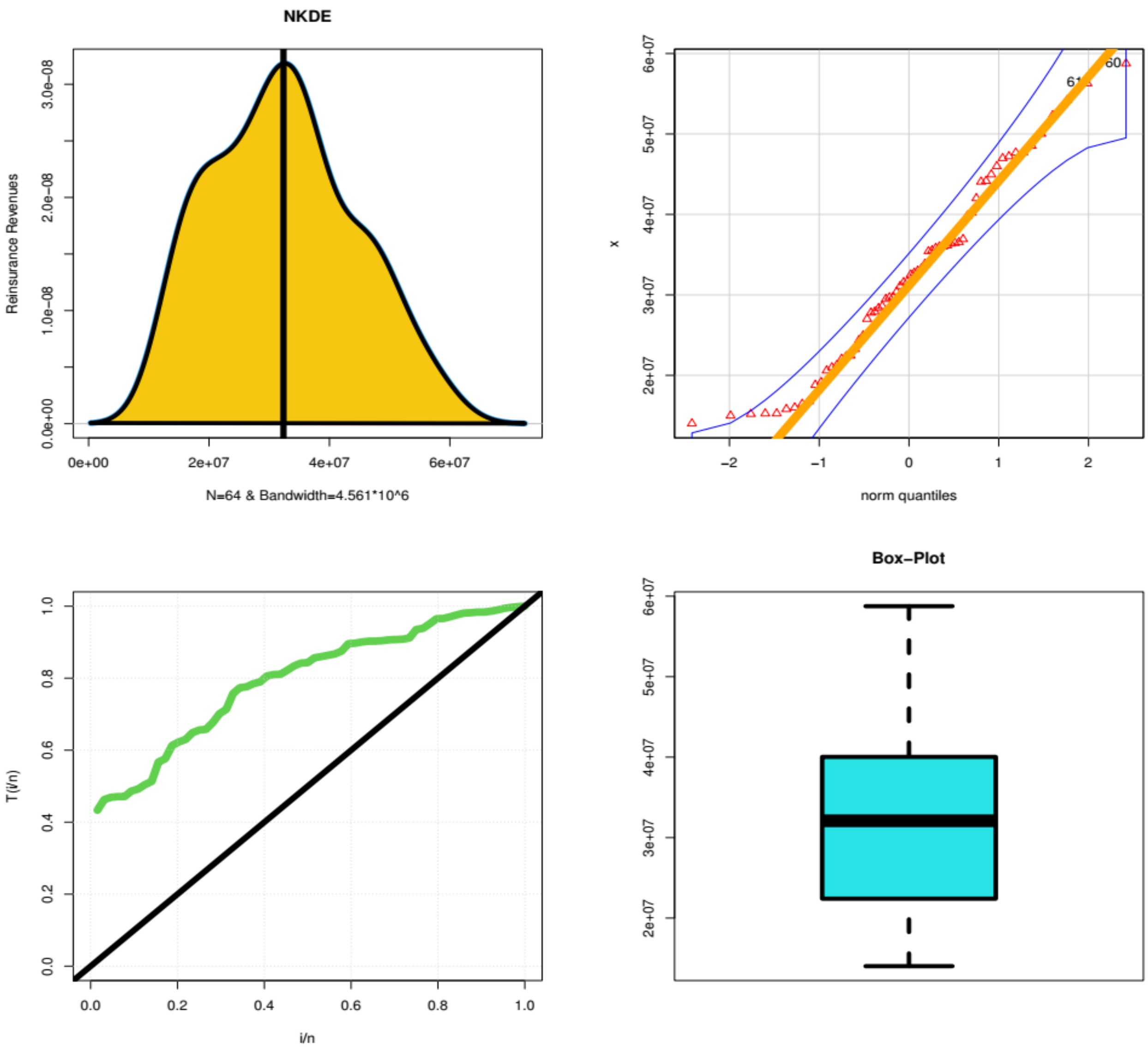

Figure 2: NKDE plot, Q-Q plot, TTT plot and box plot for the original insurance claims data.

Table 1: Some statistical measures for the reinsurance revenues data.

Measure	Result	Measure	Result
Minimum value	1402148	Maximum value	5875647
Standard deviation	1164150	25% Quantile	2242655
33% Quantile	2762846	75% Quantile	3992999
Mean	3236045	Median	3209088
Kurtosis	2.26414	Skewness	0.26682

This application is allocated for modeling the process of a time series of the revenues of reinsurance companies, and it is interested in knowing the shape of the spread by knowing the extent to which the values of the time series are interconnected with their previous values, and therefore the scattergram at lag $k = 1$ is drawn. The scatter gram is a diagram that shows points referencing two different variables. Two variables are observed and plotted on a graph to make a scattergram. The resulting display illustrates how the variables are related. Where the points are most closely grouped together, the link is stronger. Statistical surveys or laboratory test results are occasionally represented using scattergrams. The terms scatter plot, scatter diagram, scatter chart, and scatter graph can all be used to refer to a scattergram. Figure 3 shows the scattergrams (top right and left plots).

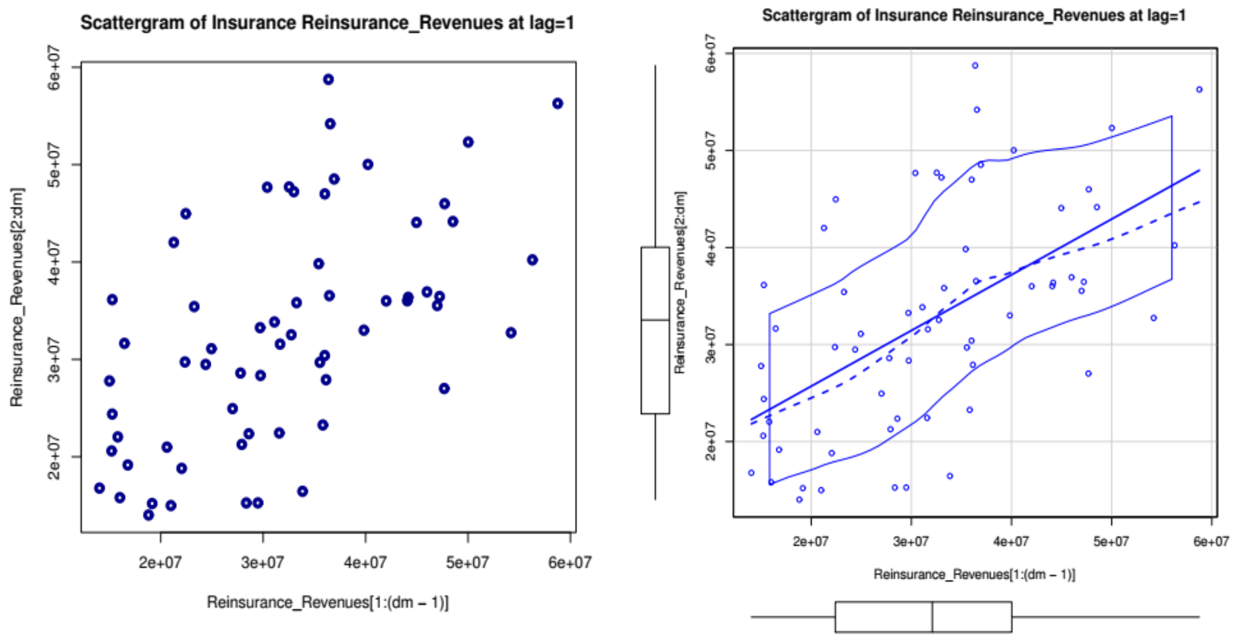

Figure 3: Scattergrams and ACF for the original insurance claims data.

Based on Figure 3, it is evident through the spread of the points that the data has a certain pattern, and this pattern can be considered an ascending pattern or what we can call in time series analysis the general trend pattern. Furthermore, the ACF, autocovariance function (ACOF), and PACF are provided. The ACF illustrates how a change in separation affects the correlation between any two signal values. Theoretical ACF assesses stochastic process memory in the time domain, not the frequency content of the process, and provides some information on the distribution of hills and valleys on the surface. The ACOF is defined as the sequence of covariances of a stationary process. The theoretical ACF is given in Figure 4 (left plot), the ACOF n Figure 4 (right plot) and theoretical PACF in Figure 4 (last plot) for the reinsurance revenues data under lag $k = 1$.

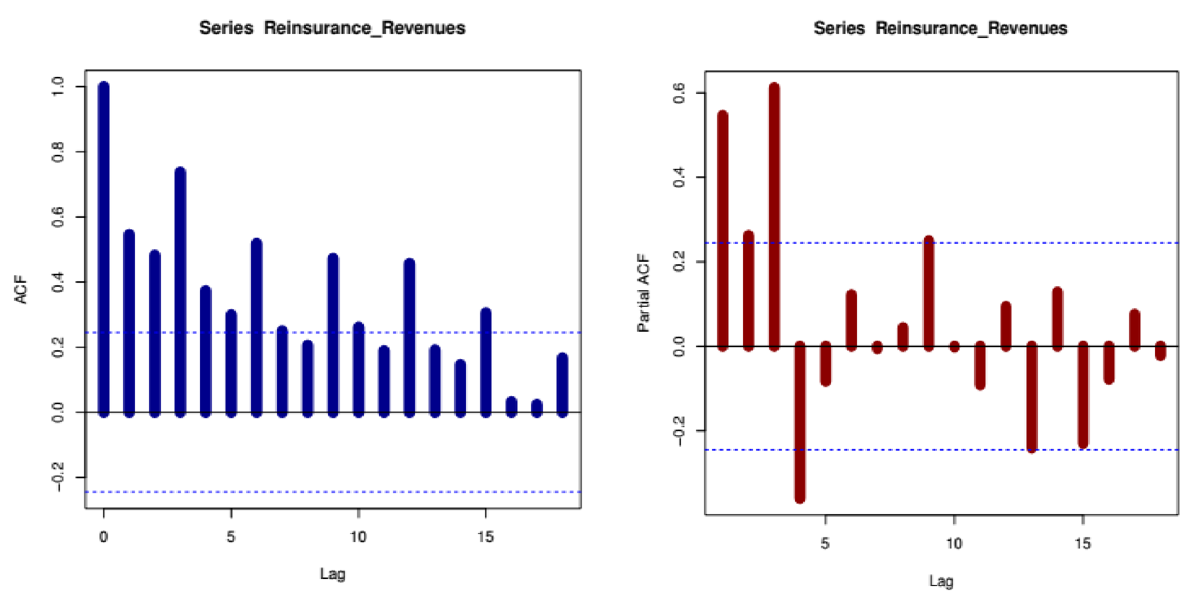

Figure 4: The ACF (left panel) and PACF (right panel) for the reinsurance revenues data.

Based on Figure 4 (the left panel), it is notted that the ACF is a positve funtion, that beeing said: all autocorrelation coefficients are positive. However, the partial autocorrelation coefficients can be positive and can be negative without any well-known pattern. Figure 5 presents the original time series plot (top left), the separated components (top right), the seasonally adjusted time series (bottom left) and trend adjusted time series (bottom right).

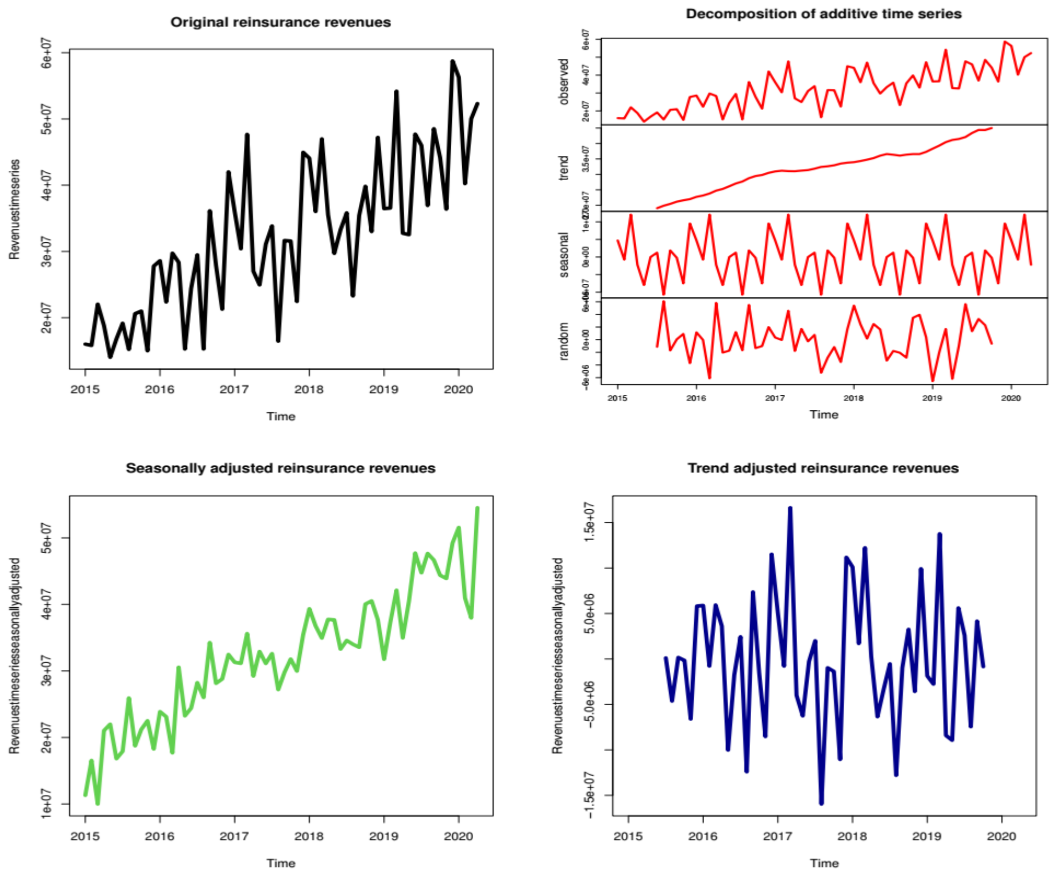

Figure 5: The original time series, the separated components, the seasonally adjusted time series and trend adjusted time series.

Based on Figure 5 (bottom left and bottom right), it is noted that excluding the seasonal effect did not affect stationary of the time series, but the general trend effect only had a significant impact on the stationary of the time series. Therefore, it is recommended excluding the trend component from the time series. So, for obtaining stationary reinsurance data, it is recommended to exclude the trend component from the time series. The ACF and PACF for the reinsurance revenues data from k=0 to k=18 is listed in Table 2. With k=0 , the ACF began with 1 and all subsequent ACF were all less than 1. The PACF accepts both positive and negative values, but not frequently, and it lacks the typical pattern that ARIMA models typically have that can be used to infer certain findings. Due to Table 2, its noted that r_0=1 and r_1=r_11=+0.546 this is in complete agreement with the theoretical results.

Table 3 lists the seasonal components for the reinsurance revenues data. It is noted that the seasonal components can have positive and negative values due to the seasonal status. Table 3 gives the seasonal components for 2015, 2016, 2017, 2018, 2019 and 2020 (up to April 2020). Table 4 gives the trend components for the reinsurance revenues data. It is seen that the trend components are all positive values, and this numerical result matches with the theoretical one. Table 4 gives the trend components for 2015 (from June), 2016, 2017, 2018, 2019 (up to April 2019).

Table 5 presents the random components for the reinsurance revenues data. It is seen that the random components can be positive and negative, and this numerical result matches the theoretical one. Table 4 gives the trend components for 2015 (from June), 2016, 2017, 2018, 2019 (up to April 2019).

Table 2: ACF, ACVF, PACF up to $k = 18$.

lag(k)	ACF	ACVF	PACF
0	1.000	1.33×10^{14}	-
1	0.546	7.29×10^{13}	+0.546
2	0.482	6.43×10^{13}	+0.262
3	0.737	9.84×10^{13}	+0.612
4	0.374	4.98×10^{13}	-0.360
5	0.299	3.99×10^{13}	-0.083
6	0.519	6.92×10^{13}	+0.121
7	0.250	3.34×10^{13}	-0.005
8	0.206	2.74×10^{13}	+0.044
9	0.473	6.31×10^{13}	+0.250
10	0.261	3.49×10^{13}	-0.002
11	0.189	2.53×10^{13}	-0.091
12	0.457	6.09×10^{13}	+0.094
13	0.191	2.55×10^{13}	-0.240
14	0.146	1.95×10^{13}	+0.128
15	0.305	4.07×10^{13}	-0.230
16	0.033	4.34×10^{12}	-0.078
17	0.024	3.19×10^{12}	+0.076
18	0.167	2.22×10^{12}	-0.022

Table 3: The seasonal components.

Season	Data
2015	4715247.02, -742028.390,12060973.980, -2212708.70,-7976948.32,-18955.410
	1242600.57, -10733295.64,1872643.01, -214220.570,-7524548.14,9531240.59
2016	4715247.020, -742028.390,12060973.98, -2212708.70,-7976948.32,-18955.410
	1242600.57, -10733295.64,1872643.01, -214220.570,-7524548.14,9531240.59
2017	4715247.020, -742028.390,12060973.98, -2212708.70,-7976948.32,-18955.410
	1242600.57, -10733295.64,1872643.01, -214220.570,-7524548.14,9531240.59
2018	4715247.020, -742028.390,12060973.98, -2212708.70,-7976948.32,-18955.410
	1242600.57, -10733295.64,1872643.01, -214220.570,-7524548.14,9531240.59
2019	4715247.020, -742028.390,12060973.98, -2212708.70,-7976948.32,-18955.410
	1242600.57, -10733295.64, 1872643.01, -214220.570,-7524548.14,9531240.59
2020	1242600.57, -10733295.64,1872643.010, -214220.570,-7524548.14,9531240.59
	1242600.570, -10733295.640,1872643.01, -214220.570.

Table 4: The trend components.

Season	Data
2015	19044119,19842077,20435866,21153695,21602843,21971379
2016	22718310,23151217,23801309,24737116,25287155,26141365
	27042538,27685683,28767734,29459504,29807304,30490634
2017	30952217,31183336,31045570,31010826,31212133,31384062
	31842900,32413052,32618942,32945625,33498678,33786615
2018	33959008,34325328,34766470,35267834,36051673,36585360
	36362366,36067310,36389348,36573107,36574756,37294785
2019	38320378,39313320,40428645, 41154962,41475990,42097259

Table 5: the random components.

Season	Data
2015	-1124826.940,6096203.01,-1704569.52,53400.2300, 915074.82,-3710811.510
2016	1168029.490,-42114.110,-6123674.50,5826601.03,-2045603.08,-1736751.30
	1201378.940,-1682270.30,5500650.36,-1330139.91,-1010706.55,1992284.89
2017	339916.2300,-44532.610,4571586.42,-1784152.20,1713659.74,-263760.970
	763321.9800,-5225089.24,-2841492.68,-1159198.74,-3527758.98,1648270.59

2018	5393266.24,2436987.07,168546.8900,2481361.95, 1624874.14, -3305339.08
	-1778431.44,-2065358.71,-2838501.39,3472678.96,3948937.70, 395803.010
2019	-6575664.980,-2024793.36,1709088.16,-6198263.80,-967383.83,5631398.33
	1345491.190, 3283448.97,2290846.95,-629806.82

Tables 6 report the predicted values ($\widehat{Y}_{\varsigma}|_{\varphi=0.546}$), the prediction errors (PE) ($E_{\varsigma}|_{\varphi=0.546}$), sum of prediction errors (SEs), mean of prediction errors (MEs), absolute prediction error (APE) ($|E_{\varsigma}||_{\varphi=0.546}$), sum of absolute prediction error (SAPE), mean of absolute prediction error (MAPE), square prediction errors (SPE) ($E^2_{\varsigma}|_{\varphi=0.546}$), sum of square prediction errors (SSPE) and MSPE. For the purpose of evaluating the PAR(1) model, the value of the series was re-estimated again as if it had not occurred. Thus, the values of the series starting are estimated from February 2015 to April 2020. After the evaluation process, predictions of future reinsurance revenues will be made.

Based on Tables 6 one concludes the following results:

I. MEs < MAPE, however the MEs in not a sufficiently accurate measure to rely on for serious comparisons. The reason for this is simply that there are negative and positive values, and many negative values may cause the size of the output of the MEs to be smaller than MAPE.

II. As mentioned, the MAPE scale is more accurate than the MEs scale, but the MSPE scale is the best for the ease of mathematical interpretation of its results.

III. The results presented in Table 6 are the best results obtained, all of them were built on the assumption that the estimated value of the parameter is 0.94955. Figure 6 presents the fitted time series plot (top left), the separated components (top right), the seasonally adjusted time series (bottom left) and trend adjusted time series (bottom right).

Based on Figure 6 (bottom left and bottom right), it is noted that excluding the seasonal effect did not affect stationary of the time series, but the general trend effect only had a significant impact on the stationary of the time series. Therefore, the trend component is removed from the fitted time series. After reviewing the original reinsurance data (see Table 2, Table 3, Table 4, Table 5 and Figure 6) then fitted values and calculating their PE, SEs, MEs, APE, SAPE, MAPE, SPE, SSPE and MSPE (in Table 6) and also drawing those fitted values (in Figure 6). This application is also interested in analyzing the residuals of the fitting process in the context of further statistical verification before starting the operations of predicting future reinsurance observations. Therefore, Figure 7, Figure 8 and Figure 9 are presented. Figure 7 gives the Cullen and Frey plots for PE, APE and SPE respectively. Figure 8 presents scattergram at lag k=1 for PE, APE and SPE respectively. Figure 9 shows the Q-Q plots for PE, APE and SPE, respectively. Based on Figure 7 (the Cullen and Frey plots for PE, APE and SPE respectively), it is seen that the PE and APE are normally distributed. Due to Figure 8 (the Scattergram at lag k = 1 for PE, APE and SPE respectively), it is noted that the PE and APE have a similar pattern of spread. According to Figure 9 (the Q-Q plots for PE, APE and SPE respectively), PE and APE are normally distributed without extreme values. However, the SPE has some extreme values, and this is an expected result. Generally, all these plots confirm beyond any doubt the adequacy and efficiency of the new model for modeling and forecasting reinsurance data.

Table 6: Assessing the PAR model.

Tim e↓ Fitting and criteria →	$\widehat{Y}_{\varsigma}\|_{\varphi=0.546}$	$E_{\varsigma}\|_{\varphi=0.546}$	$\|E_{\varsigma}\|\|_{\varphi=0.546}$	$E^2_{\varsigma}\|_{\varphi=0.546}$
2015				
Jan	-	-	-	-
Feb	16010022	-202434.55	202434.55	4.097975×10^{10}
Mar	15962617	6084529.02	6084529.02	3.702149×10^{13}
Apr	17388875	1425708.29	1425708.29	2.032644×10^{12}
May	17723083	-3701603.39	3701603.39	1.370187×10^{13}
Jun	16855301	-71372.480	71372.480	5.094031×10^{9}
Jul	16838660	2323232.19	2323232.19	5.397408×10^{12}
Aug	17383278	-2178293.70	2178293.70	4.744963×10^{12}
Sep	16872609	3731330.98	3731330.98	1.392283×10^{13}
Oct	17747200	3245674.78	3245674.78	1.053440×10^{13}
Nov	18508126	-3514756.34	3514756.34	1.235351×10^{13}
2016				
Jan	20053568	8548018.50	8548018.50	7.306862×10^{13}
Feb	22057390	309684.07	309684.07	9.590422e+10
Mar	22129988	7608620.57	7608620.57	5.789111×10^{13}

Apr	23913471	4437536.94	4437536.94	1.969173×10^{13}
May	24953611	-9689007.27	9689007.27	9.387686×10^{13}
Jun	22682478	1703180.08	1703180.08	2.900822×10^{12}
Jul	23081780	6404737.07	6404737.07	4.102066×10^{13}
Aug	24583112	-9312994.74	9312994.74	8.673187×10^{13}
Sep	22400029	13740998.56	13740998.56	1.888150×10^{14}
Oct	25621099	2294044.66	2294044.66	5.262641×10^{12}
Nov	26158864	-4886814.65	4886814.65	2.388096e+13
Dec	25013315	17000844.88	17000844.88	2.890287×10^{14}
		2017		
Jan	28998544	7008836.67	7008836.67	4.912379×10^{13}
Feb	30641400	-244624.62	244624.62	5.984120×10^{10}
Mar	30584179	17093951.73	17093951.73	2.922032e+14
Apr	34591198	-7577233.27	7577233.27	5.741446e+13
May	32814969	-7866124.30	7866124.30	6.187591×10^{13}
Jun	30971123	130222.540	130222.54	1.695791×10^{10}
Jul	31001560	2847262.23	2847262.23	8.106902×10^{12}
Aug	31669142	-15214475.04	15214475.04	2.314803×10^{12}
Sep	28102606	3547486.65	3547486.65	1.258466×10^{13}
Oct	28934198	2638007.62	2638007.62	6.959084×10^{12}
Nov	29552514	-7106142.97	7106142.97	5.049727×10^{13}
Dec	27886722	17079403.77	17079403.77	2.917060×10^{14}
		2018		
Jan	31890414	12177106.86	12177106.86	1.482819×10^{14}
Feb	34744890	1275397.16	1275397.16	1.626638×10^{12}
Mar	35043902	11952088.41	11952088.41	1.428524×10^{14}
Apr	37845569	-2309081.32	2309081.32	5.331857×10^{12}
May	37304320	-7604720.82	7604720.82	5.783178×10^{13}
Jun	35521686	-2260620.61	2260620.61	5.110406×10^{12}
Jul	34991796	834738.91	834738.91	6.967890×10^{11}
Aug	35187479	-11918823.79	11918823.79	1.420584×10^{14}
Sep	32393588	3029901.85	3029901.85	9.180305×10^{12}
Oct	33103807	6727758.70	6727758.70	4.526274×10^{13}
Nov	34680899	-1681753.79	1681753.79	2.828296×10^{12}
Dec	34286602	12935226.20	12935226.20	1.673201×10^{14}
Jan	31890414	12177106.86	12177106.86	1.482819×10^{14}
		2019		
Jan	37318880	-858919.91	858919.91	7.377434×10^{11}
Feb	37117553	-571054.34	571054.34	3.261031×10^{11}
Mar	36983677	17215029.72	17215029.72	$2.963572e\times10^{14}$
Apr	41019178	-8275188.39	8275188.39	6.847874×10^{13}
May	39079430	-6547772.46	6547772.46	4.287332×10^{13}
Jun	37544499	10165202.63	10165202.63	1.033313×10^{14}
Jul	39927432	6064709.57	6064709.57	3.678070×10^{13}
Aug	41348947	-4415281.98	4415281.98	1.949471×10^{13}
Sep	40313933	8212327.13	8212327.13	6.744232×10^{13}
Oct	42239200	1921216.18	1921216.18	3.691072×10^{12}
Nov	42689459	-6314502.51	6314502.51	3.987294×10^{13}
Dec	41209286	17547187.66	17547187.66	3.079038×10^{14}
Jan	37318880	-858919.91	858919.91	7.377434×10^{11}
		2020		
Jan	45322656	10965644.87	10965644.87	1.202454×10^{14}
Feb	47893166	-7667922.74	7667922.74	5.879704×10^{13}
Mar	46095763	3926402.23	3926402.23	1.541663×10^{13}
Apr	47016088	5304604.94	5304604.94	2.813883×10^{13}
Sum	**-**	**137573866**	**401556906**	$\mathbf{4.08448\times10^{15}}$
Mean	**-**	**2183712**	**6373919**	$\mathbf{6.483302\times10^{13}}$

DOI: http://dx.doi.org/10.18187/pjsor.v19i3.4123

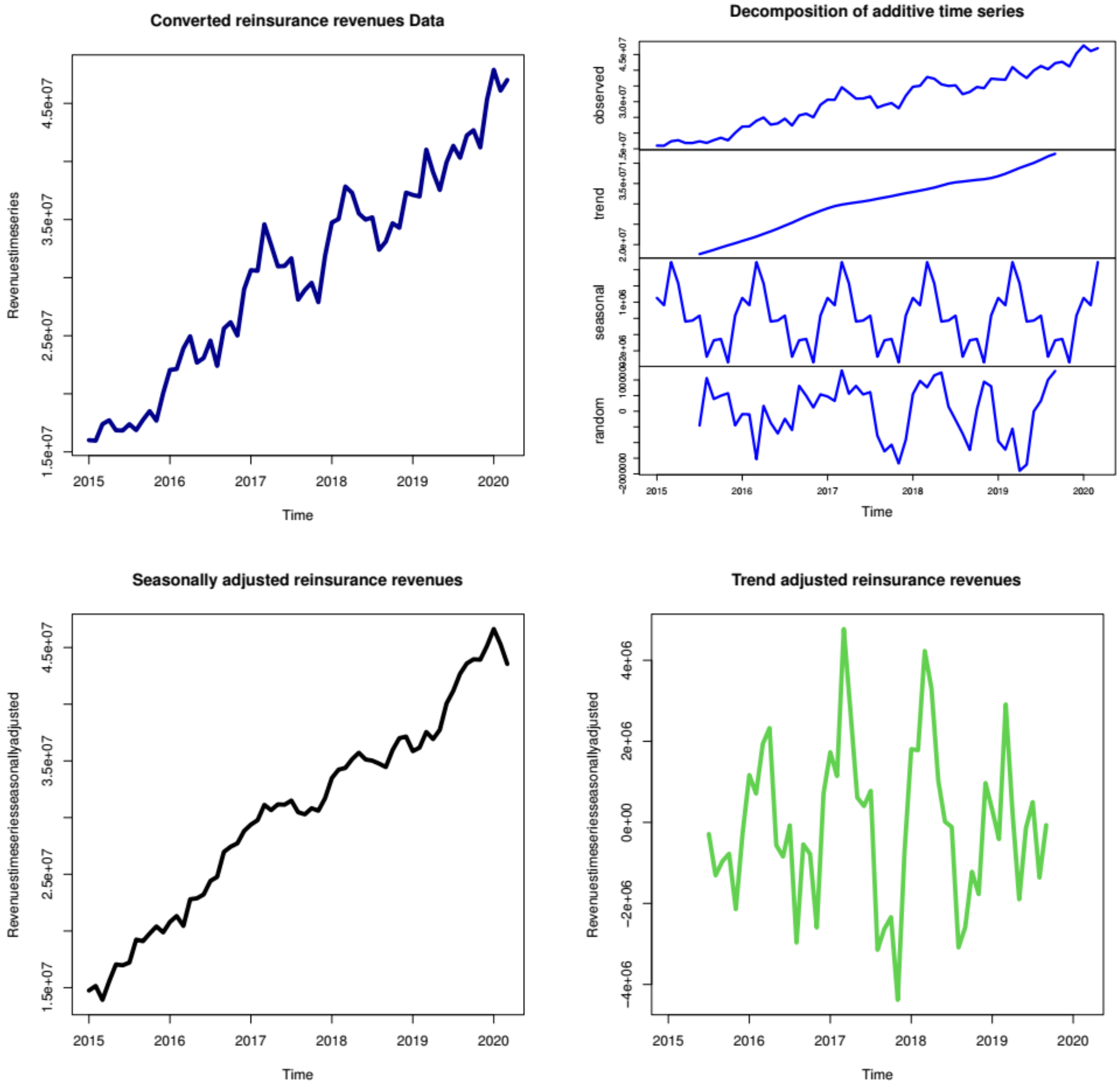

Figure 6: The fitted time series, the separated components, the seasonally adjusted time series and trend adjusted time series.

DOI: http://dx.doi.org/10.18187/pjsor.v19i3.4123

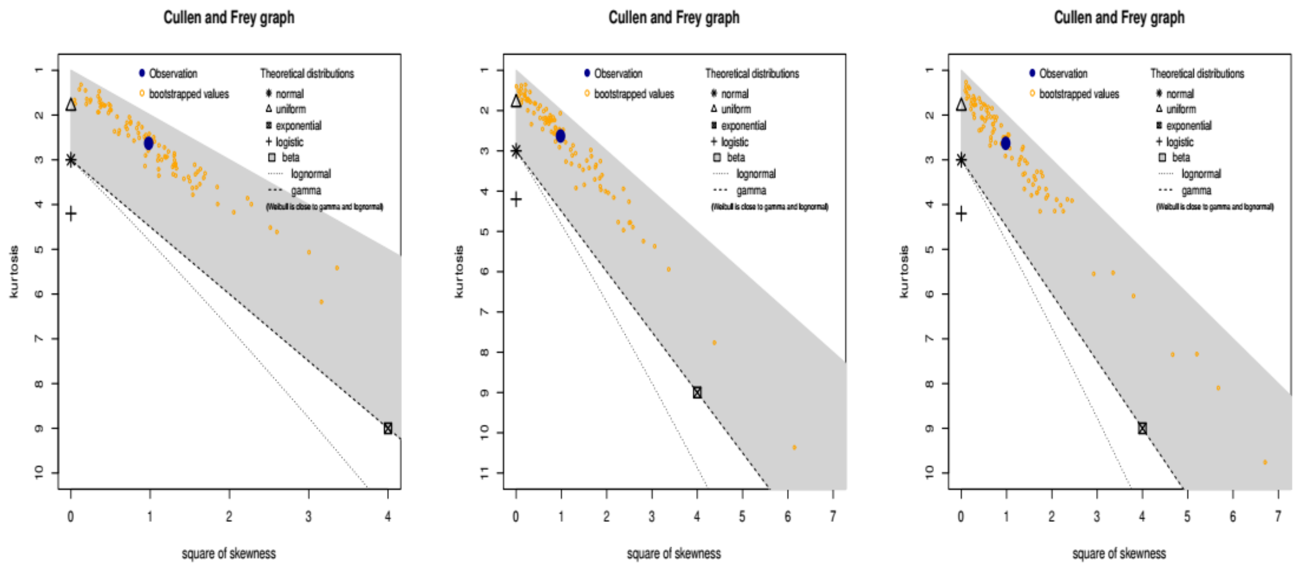

Figure 7: Cullen and Frey plots for PE, APE and SPE respectively.

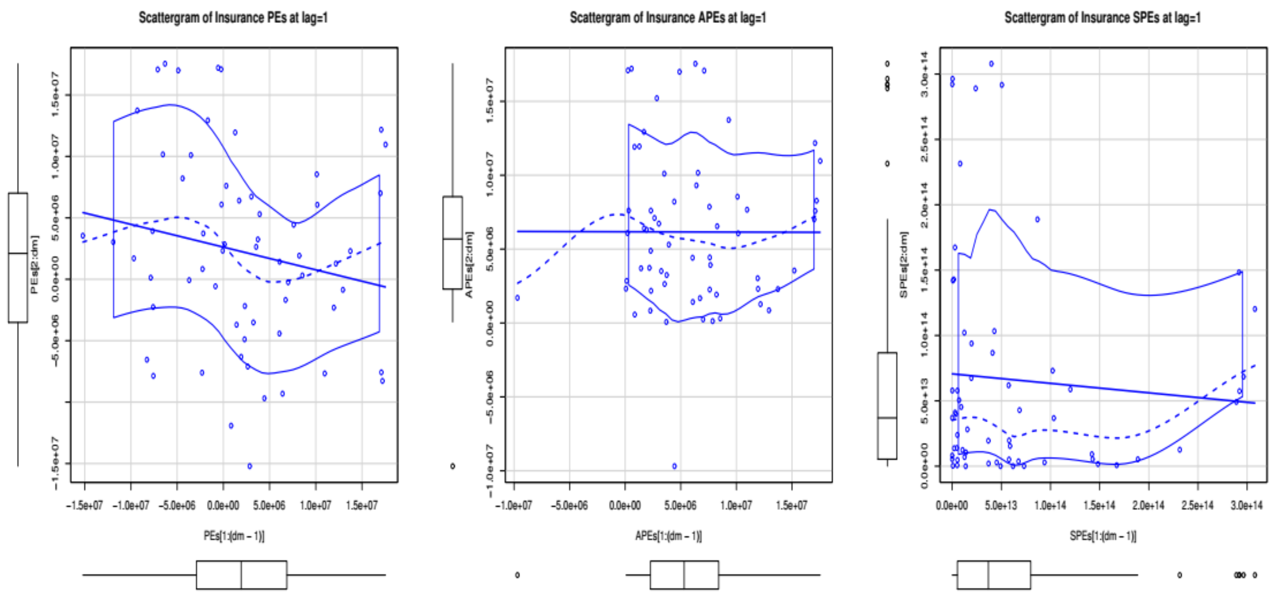

Figure 8: Scattergram at lag k = 1 for PE, APE and SPE respectively.

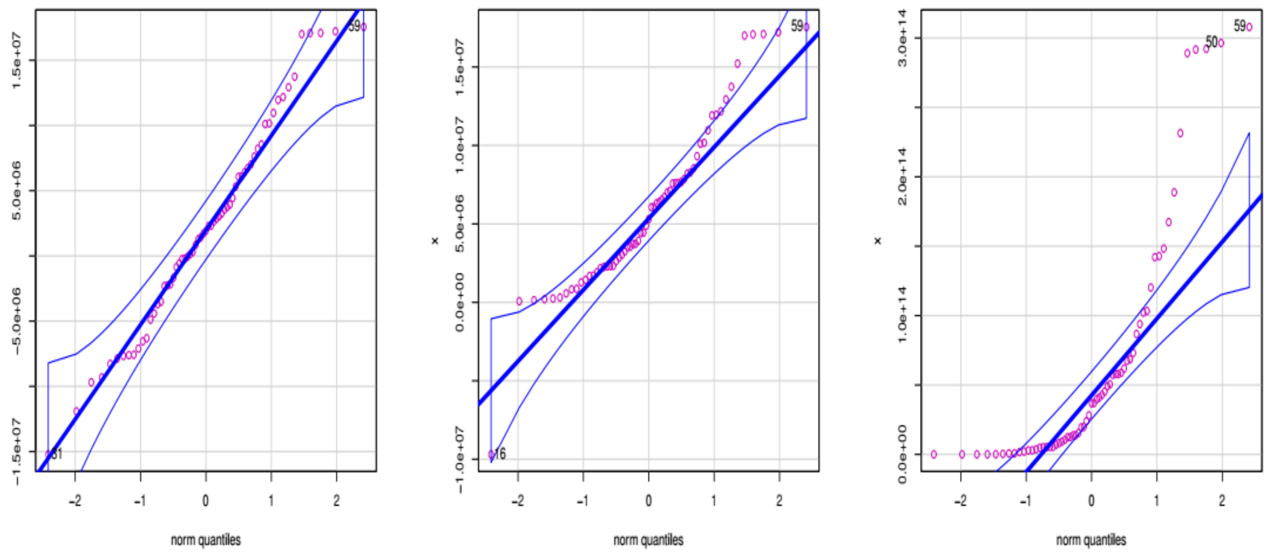

Figure 9: Q-Q plots for PE, APE and SPE respectively.

Now, after all these procedures and verifications, the process of forecasting the future views of the reinsurance data can be started. Knowing that the longer the forecast period, the greater the forecast error and the more short-term the forecast, the more accurate the forecast. This means that the time range has a significant effect on the values of forecast errors, and that it is an accepted fact in the field of forecasting in general, whatever the model, method, technique, or methodology that is being used in prediction. The 80% (Lower bound (LB), Upper bound (UB)) and 95% (LB, UB) for 2020 (beginning in May), 2021, and September 2022 for the PAR model and Holt-Winters model are reported in Table 7.

Table 7: Predictive values as predictive confidence intervals "80%(LB, UB), 95%(LB, UB)".

	$\hat{Y}_{T\|T-1}$ for 2020	
	PAR model	Holt-Winters model
May	(44622981,49230897),(43403339,50450539)	(38247556,58271470),(32947546,63571480)
Jun	(43856296,49997582),(42230795,51623082)	(37976161,58542865),(32532484,63986542)
Jul	(43245978,50607900),(41297395,52556483)	(37711747,58807279),(32128098,64390928)
Aug	(42723357,51130520),(40498115,53355762)	(37453802,59065224),(31733604,64785423)
Sep	(42258885,51594992),(39787767,54066111)	(37201871,59317155),(31348309,65170717)
Oct	(41836620,52017258),(39141967,54711910)	(36955555,59563472),(30971600,65547426)
Nov	(41446795,52407082),(38545782,55308095)	(36714492,59804534),(30602927,65916099)
Dec	(41082916,52770961),(37989277,55864600)	(36478360,60040666),(30241795,66277231)
	$\hat{Y}_{T\|T-1}$ for 2021	
Jan	(40740403,53113475),(37465449,56388429)	(3624687060272156),(2988776166631265)
Feb	(40415883,53437995),(36969138,56884740)	(3601975760499269),(2954042166978605)
Mar	(40106787,53747091),(36496416,57357461)	(3579678260722244),(2919941067319616)
Apr	(39811104,54042773),(36044209,57809669)	(3557772660941300),(2886439467654632)
May	(39527227,54326650),(35610057,58243821)	(3536239161156635),(2853506767983959)
Jun	(39253846,54600031),(35191956,58661921)	(3515059361368433),(2821114968307877)
Jul	(38989875,54864002),(34788248,59065630)	(3494216261576864),(2789238268626644)
Aug	(38734405,55119472),(34397540,59456337)	(3473694461782082),(2757852868940498)
Sep	(38486665,55367213),(34018653,59835224)	(3453479461984232),(2726936769249659)
Oct	(38245991,55607886),(33650575,60203303)	(3433557962183448),(2696469369554333)
Nov	(38011812,55842065),(33292430,60561448)	(3413917362379853),(2666431769854709)
Dec	(37783630,56070248),(32943454,60910424)	(3394546362573563),(2636806270150964)
	$\hat{Y}_{T\|T-1}$ for 2022	
Jan	(37561004,56292873),(32602978,61250899)	(3375433962764687),(2607576470443262)

Feb	(37343549,56510328),(32270409,61583468)	(3356570162953325),(2578726770731759)
Mar	(37130920,56722957),(31945221,61908656)	(3337945463139572),(2550242771016599)
Apr	(36922809,56931068),(31626943,62226934)	(3319551063323516),(2522110871297918)
May	36718940,57134937),(31315153,62538725)	(3301378563505241),(2494318471575842)
Jun	(36519064,57334813),(31009468,62844409)	(3283420063684826),(2466853371850493)
Jul	(36322955,57530923),(30709544,63144333)	(3265668363862343),(2439704472121982)
Aug	(36130407,57723471),(30415068,63438810)	(3248116264037864),(2412860872390418)
Sep	(35941233,57912645),(30125752,63728126)	(3230757364211453),(2386312672655900)
Ljung-Box	**17.4068**	**23.4914**
p-value	**0.7996**	**0.70634**

Generally, the 80% (LB, UB) is wider than the 95% (LB, UB) for all months because the forecast error increases if the time range of future predictions increases. Actually, the 95% (LB, UB) is preferable since it was established based on a higher confidence level. This Table's findings show that the larger the statistical prediction's time horizon, the greater its prediction error and the wider its confidence intervals. Drawing prediction values will not be as important to us as examining forecast errors of the new model is analyses beforehand. Table 7 gives all results of the predictive values as predictive confidence intervals (80% (LB, UB), 95% (LB, UB)) for the PAR model and Holt-Winters model. The Ljung-Box test and its corresponding p -value are used for assessing the prediction accuracy. It is noted that, the Ljung-Box test for the PAR model is 17.4068, however, the Ljung-Box test for the Holt-Winters model is 23.4914. It should be noted that whereas the p-value for the Holt-Winters model is 0.70634, the p-value for the PAR model is 0.7996. It is also a good idea to check if the forecast errors are regularly distributed with mean zero and constant variance to make sure the predictive model cannot be improved. Finally, Figure 10 presents the fitted reinsurance time series values against the original reinsurance one. Figure 11 presents the fitted reinsurance time series values against the original reinsurance one using the Holt-Winters model. A temporal plot of the in-sample forecast errors can be made, as shown in Figure 10, to show if the forecast errors in this case have constant variance. This graph demonstrates how the in-sample forecast error variance has remained stable throughout time.

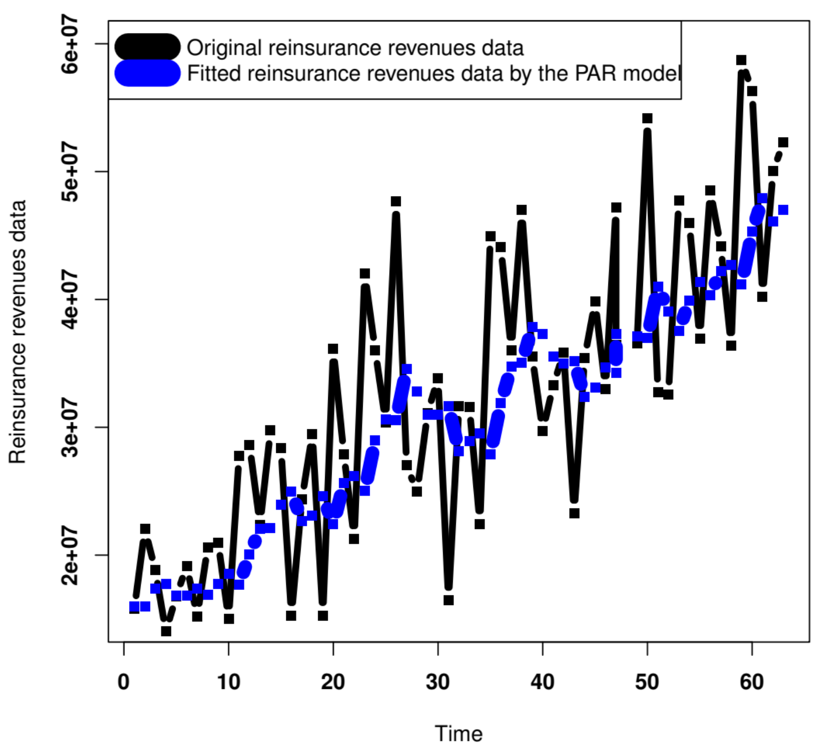

Figure 10: The fitted reinsurance time series values against the original reinsurance one via the PAR model.

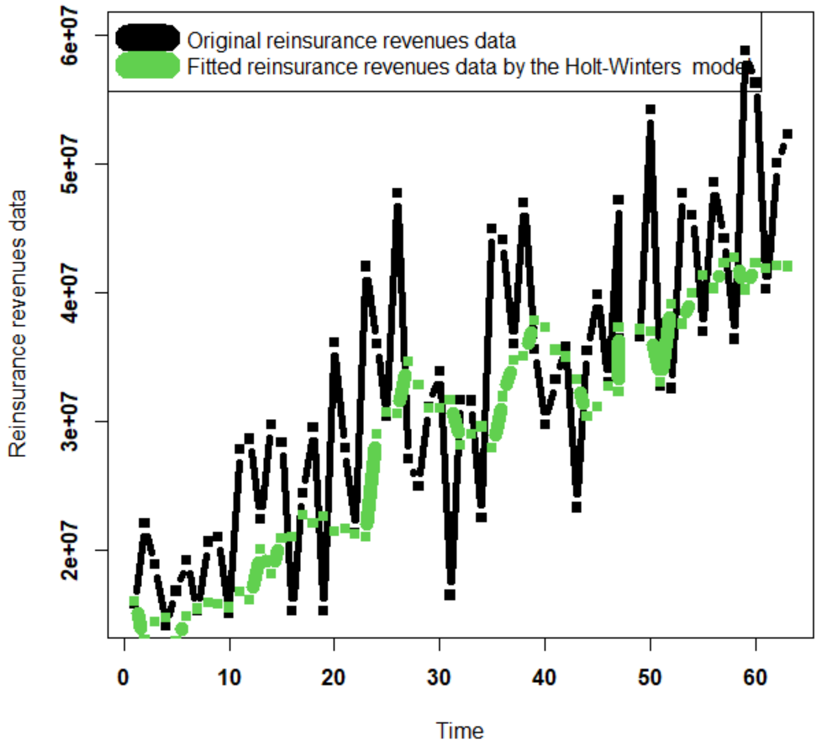

Figure 11: The fitted reinsurance time series values against the original reinsurance one via the Holt-Winters model.

Based on Figures 10 and 11, It is clear that the new PAR(1) model gives future estimates that are very close to the original time series, and this indicates the efficiency of the PAR(1) model and its adequacy to model those data and the possibility of being relied upon with a high degree of confidence in future predictions of reinsurance data.

Residuals analysis plays a crucial role in time series modeling. Time series models are used to forecast future values based on historical data patterns. Residuals, also known as errors or residuals, are the differences between the observed values and the predicted values obtained from the time series model. Analyzing these residuals helps to assess the model's performance and ensure that the underlying assumptions of the model are met. Here are some key reasons why residuals analysis is important in time series models:

I. Residuals analysis helps determine whether the model adequately captures the underlying patterns and dynamics of the time series data. If the residuals exhibit a random and independent behavior, it suggests that the model has captured most of the systematic components, and the model is considered adequate. On the other hand, if the residuals show a non-random pattern or exhibit serial correlation, it indicates that the model might be missing important information or structure in the data.

II. Time series models are based on certain assumptions, such as the absence of autocorrelation (independence of residuals), constant mean and variance, and normality of residuals. Residual analysis allows us to validate these assumptions. If the residuals exhibit autocorrelation, it suggests that the model does not adequately capture the temporal dependencies in the data. Violation of assumptions may lead to biased parameter estimates and unreliable forecasts.

III. Residuals analysis provides insights into the accuracy of the model's forecasts. By examining the residuals' patterns and properties, such as mean, variance, and distribution, one can assess the model's ability to capture the true underlying behavior of the time series. Large and systematic residuals indicate potential model inadequacies or missing explanatory variables, which may affect the forecast accuracy.

IV. Residuals analysis can guide model refinement and improvement. If the residuals exhibit systematic patterns, such as trends or cycles, it suggests that the model might benefit from including additional variables or adjusting the model's structure. By analyzing the residuals, one can identify areas where the model can be enhanced to better capture the complexity of the time series data.

V. Residuals analysis helps identify outliers or unusual observations that are not accounted for by the model. Outliers can significantly influence the model's performance and forecasts. By examining the residuals' magnitude and distribution, one can identify potential outliers and investigate the reasons behind their occurrence.

Overall, residuals analysis is an essential step in time series modeling as it allows us to assess model adequacy, validate assumptions, evaluate forecast accuracy, refine the model, and identify outliers. It helps ensure that the model is capturing the true dynamics of the data and provides reliable forecasts for future time points. Figure 12 gives the matrix of the ACF analysis for both point forecasting and interval forecasting. Based on Figure 12 the matrix of the ACF analysis is a skewed symmetric matrix, the elements of the diagonal represent the variance of the residues and the elements around the diagonal represent the covariance of the residues. Due to the first row of the matrix, there are no correlations between the residuals of the point forecasting and the lower bounds (80% & 95%), and the residuals of the point forecasting and the upper bounds (80% & 95%). Due to the first column of the matrix, there are no correlations between the residuals of the lower bounds (80% & 95%) and the point forecasting, and the upper bounds (80% & 95%) and the point forecasting. Figure 13 provides the checking the residuals normality vial three different shapes. The first shape (left panel) refers to the plot of the residuals, the second shape (middle panel) refers to the histogram of the residuals, the third shape (right panel) refers to the Q-Q of the residuals. Based on Figure 13 (left panel, middle panel, and right panel), its is noted that the results are normally distributed.

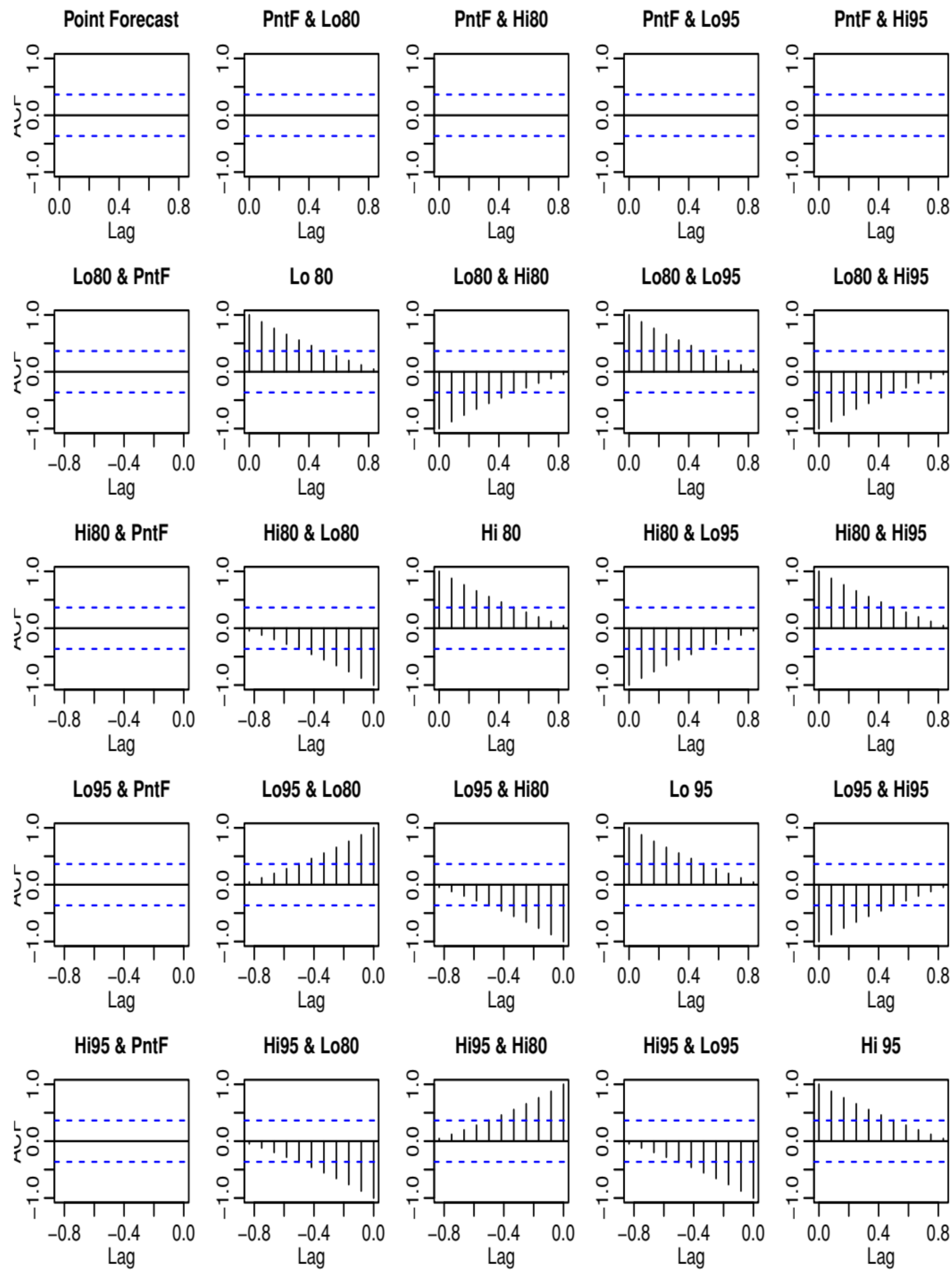

Figure 12: The ACF analysis for both point forecasting and interval forecasting.

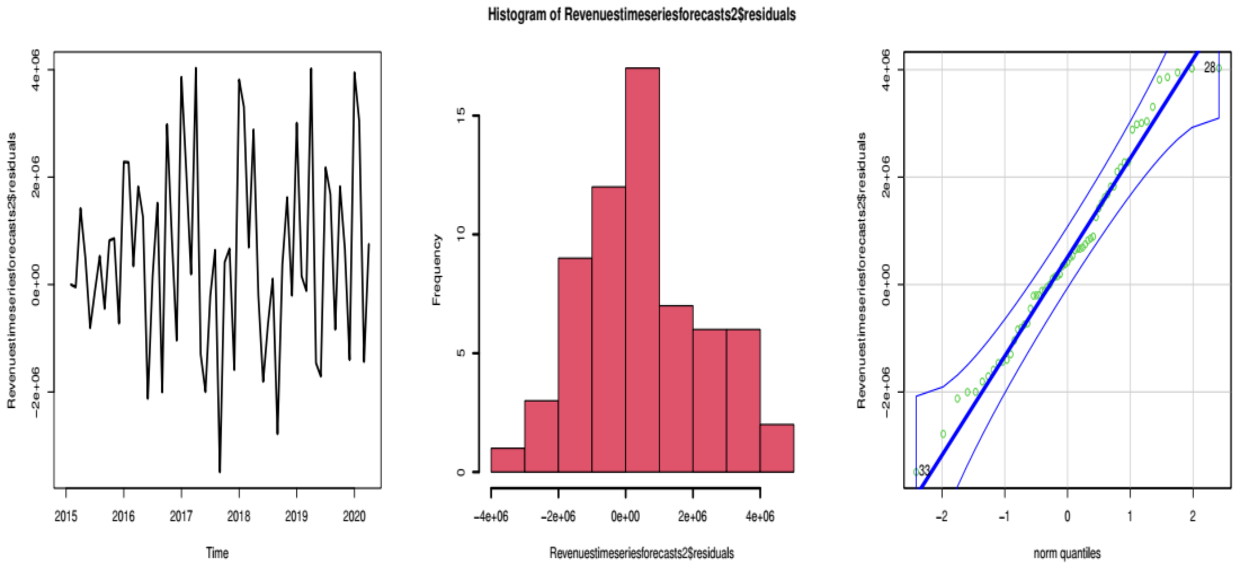

Figure 13: Checking the residuals normality.

4. Concluding remarks

In this paper, a new version of the autoregressive model is proposed, in the so-called a partial autoregressive (PAR) model. The PAR(π) approach is a linear regression model that uses its own lags as predictors, where π is the order of the PAR model. The linear regression models are adequate when the predictors are not correlated and are independent of each other. The results of the new model depended on a new algorithm that has been formulated to facilitate the statistical prediction process considering the rapid developments in time series models. The new algorithm depends on the values of the autocorrelation and partial autocorrelation functions in determining the estimated value of the parameter model subject to certain terms and conditions. For this main proposal, a monthly time series represents the reinsurance revenue has been analyzed. Fortunately, our data are recent time series data with a starting date of February 2015 and an ending date of April 2020. The new model is assessed by re-estimating the actual time series values. The PAR model was used to predict the future values of reinsurance revenues data, the results of the PAR model is compared with Holt-Winters model results under the Ljung-Box test and its corresponding p-value. Many graphical tools were also used to show the importance and applicability of the new model, in addition to a full analysis of the model's errors to show the adequacy and efficiency of the new model in forecasting operations. The new model is expected to draw a lot of interest from people who are interested in applied statistics, time series, actuaries, and other related fields. Given that the model's development will be continued in other upcoming works, the model's simplest situations (π =1) are taken into consideration in this work. It is noted that, the Ljung-Box test for the PAR model is 17.4068, however, the Ljung-Box test for the Holt-Winters model is 23.4914. It should be noted that whereas the p-value for the Holt-Winters model is 0.70634, the p-value for the PAR model is 0.7996.

References

1. Box, G. and Jenkins, G. (1970). Time series analysis: Forecasting and control. San Francisco: Holden-Day.
2. Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
3. Cummins, J. D. and Griepentrog, G. L. (1985). Forecasting automobile insurance paid claim costs using econometric and ARIMA models. International Journal of Forecasting, 1(3), 203-215.
4. Jang, K. P., Kam, S. and Park, J. Y. (1991). Trend and forecast of the medical care utilization rate, the medical expense per case and the treatment days per cage in medical insurance program for employees by ARIMA model. Journal of Preventive Medicine and Public Health, 24(3), 441-458.
5. Hafiz, U. A., Salleh, F., Garba, M. and Rashid, N. (2021). Projecting insurance penetration rate in Nigeria: An ARIMA approach. Revista Geintec Gestao Inovacao E Tecnologias, 11(3), 63-75.
6. Jakaša, T., Andročec, I. and Sprčić, P. (2011). Electricity price forecasting-ARIMA model approach. In 2011 8th International Conference on the European Energy Market (pp. 222-225). IEEE.
7.

8. Kumar, V. S., Satpathi, D. K., Kumar, P. P. and Haragopal, V. V. (2020). Forecasting motor insurance claim amount using ARIMA model. In AIP Conference Proceedings (Vol. 2246, No. 1, p. 020005). AIP Publishing LLC.
9. Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. Biometrika, 65, 297-303.
10. Darekar, A. and Reddy, A. (2017). Forecasting oilseeds prices in India: Case of groundnut. Forecasting Oilseeds Prices in India: Case of Groundnut (December 14, 2017). Journal Oilseeds Research, 34(4), 235-240.
11. Sahu, P. K., Mishra, P., Dhekale, B. S., Vishwajith, K. P. and Padmanaban, K. (2015). Modelling and forecasting of area, production, yield and total seeds of rice and wheat in SAARC countries and the world towards food security. American Journal of Applied Mathematics and Statistics, 3(1), 34-48.
12. Ibrahim, M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Yousof, H.M. Bayesian and Non-Bayesian Risk Analysis and Assessment under Left-Skewed Insurance Data and a Novel Compound Reciprocal Rayleigh Extension. Mathematics 2023, 11, 1593. https://doi.org/10.3390/ math11071593
13. Khedr, A. M., Nofal, Z. M., El Gebaly, Y. M., & Yousof, H. M. A Novel Family of Compound Probability Distributions: Properties, Copulas, Risk Analysis and Assessment under a Reinsurance Revenues Data Set. Thailand Statistician, forthcoming.
14. Nath, B., Dhakre, D. S. and Bhattacharya, D. (2019). Forecasting wheat production in India: An ARIMA modelling approach. Journal of Pharmacognosy and Phytochemistry, 8(1), 2158-2165.
15. Palakuru, M., Yarrakula, K., Chaube, N. R., Sk, K. B. and Satyaji Rao, Y. R. (2019). Identification of paddy crop phenological parameters using dual polarized SCATSAT-1 (ISRO, India) scatterometer data. Environmental Science and Pollution Research, 26(2), 1565-1575.
16. Shrahili, M.; Elbatal, I. and Yousof, H. M. (2021). Asymmetric density for risk claim-size data: Prediction and bimodal data applications. Symmetry 2021, 13, 2357.
17. Idrees, S. M., Alam, M. A. and Agarwal, P. (2019). A prediction approach for stock market volatility based on time series data. IEEE Access, 7, 17287-17298.
18. Bruneau, C., De Bandt, O., Flageollet, A. and Michaux, E. (2007). Forecasting inflation using economic indicators: the case of France. Journal of Forecasting, 26(1), 1-22.
19. Githeko, A. K. and Ndegwa, W. (2001). Predicting malaria epidemics in the Kenyan highlands using climate data: a tool for decision makers. Global Change and Human Health, 2(1), 54-63.
20. Debnath, K. B. and Mourshed, M. (2018). Forecasting methods in energy planning models. Renewable and Sustainable Energy Reviews, 88, 297-325
21. Akbar, M., Iqbal, F. and Noor, F. (2019). Bayesian analysis of dynamic linkages among gold price, stock prices, exchange rate and interest rate in Pakistan. Resources Policy, 62, 154-164.
22. Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. Journal of the American Statistical Association, 87(419), 659-671.
23. Makridakis, S. and Hibon, M. (1997). ARMA models and the Box-Jenkins methodology. Journal of forecasting, 16(3), 147-163.
24. Mohamed, H. S., Ali, M. M. and Yousof, H. M. (2022a). The Lindley Gompertz Model for Estimating the Survival Rates: Properties and Applications in Insurance, Annals of Data Science, 10.1007/s40745-022-00451-3
25. Mohamed, H. S., Cordeiro, G. M. and Yousof, H. M. (2022b). The synthetic autoregressive model for the insurance claims payment data: modeling and future prediction. Statistics, Optimization & Information Computing, forthcoming.
26. Mohamed, H. S., Cordeiro, G. M., Minkah, R., Yousof, H. M. and Ibrahim, M. (2022c). A size-of-loss model for the negatively skewed insurance claims data: applications, risk analysis using different methods and statistical forecasting. Journal of Applied Statistics, forthcoming
27. Mohammadi, H. and Rich, D. P. (2013). Dynamics of unemployment insurance claims: an application of ARIMA-GARCH models. Atlantic Economic Journal, 41(4), 413-425.
28. Saikkonen, P. and Teräsvirta, T. (1985). Modelling the dynamic relationship between wages and prices in Finland. The Scandinavian Journal of Economics, 102-119.
29. Yan, Z. and Hong, L. (2015). Testing for asymmetric information in reinsurance markets. The Geneva Papers on Risk and Insurance-Issues and Practice, 40, 29-46.
30. Yousof, H. M., Ansari, S. I., Tashkandy, Y., Emam, W., Ali, M. M., Ibrahim, M., Alkhayyat, S. L. (2023). Risk Analysis and Estimation of a Bimodal Heavy-Tailed Burr XII Model in Insurance Data: Exploring Multiple Methods and Applications. Mathematics. 2023; 11(9):2179. https://doi.org/10.3390/math11092179
31. Yousof, H.M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Minkah, R.; Ibrahim, M. A Novel Model for Quantitative Risk Assessment under Claim-Size Data with Bimodal and Symmetric Data Modeling. Mathematics 2023, 11, 1284. https://doi.org/10.3390/math11061284
32. Yousof, H.M.; Tashkandy, Y.; Emam, W.; Ali, M.M.; Ibrahim, M. A New Reciprocal Weibull Extension for Modeling Extreme Values with Risk Analysis under Insurance Data. Mathematics 2023, 11, 966. https://doi.org/10.3390/ math11040966