THE INTERNATIONAL JOURNAL OF HUMANITIES & SOCIAL STUDIES

The Sustainability of Old Pension Benefits Using Statistical Analysis

Alhaji Inusah Mahama
Lecturer, Department of Statistics, Bolgatanga Polytechnic, Ghana

Abstract:
The role of insurance is to reimburse. Pension policies are instituted to secure the future income of individuals once they are out of active service. The Social Security and National Insurance Trust (SSNIT) has the primary responsibility to collect contributions from its member's whiles they are in active service, invest it and pays it back to them at the due time. The research sought to evaluate the sustainability of the old age pension benefit of SSNIT using statistical analysis. The Historical research design was adopted in this study. The number of pensioners and the total monthly payment made to them from 2011 to 2017 was used for the study. Analysis revealed that, the number of Pensioners follows the Negative Binomial Statistical Distribution whiles the total monthly payments also follows the Lognormal (3P) Distribution. Both distributions are positively skewed. In the quest to forecast, stationary was established through differencing of the data, the sample ACF and the sample Partial Autocorrelation Function (PACF) of the differenced data were considered to generate possible models after which the AIC, AICc and BIC of the candidate models under the various data were examined and those candidate models with the smallest AIC, AICc and BIC were chosen as the best-fit models among the candidate models and used for forecasting. Forecast for three years was done for both number of Pensioners and total monthly Payments. Both Forecasted values showed an increasing trend.

Keywords: *Queues, Queuing models, services*

1. Background of Study

The basic responsibility of pension schemes is to pay benefits to pensioners. The pension system in Ghana prior to 2004 was characterised with many problems which made the payment of the retirement benefit a failure in Ghana. The sustainability of a pension scheme needs to be looked at in terms of future burden of financing on all those working in the national economy in order to pay the promised benefits to those who are no longer working.

A mix of statistical studies and experienced judgments is used to valuate pension schemes to ascertain its sustainability over time. Economic, demographic and actuarial assumptions are made in that regard to estimate these future liabilities. However, since assumptions are often derived from long-term data, unusual short- term uncertainties and unanticipated trends can sometimes cause problems. It is usually easy to value the assets of a social security scheme since it usually holds liquid securities such us stocks and bonds as compared to liabilities which can be very difficult to value. Aside the economic, demographic and actuarial assumptions, statistical analysis and some assumption must be made to determine the total value of pension benefits that must be paid out in the future. Secondly, an assumption must be made on the expected growth of the scheme's assets which will allow it meet those obligations. There are (2) two measurement methods to determine the value of pay-outs that must be made in the future. Namely:

- *The Solvency Value- A Market-Based:* Under this method, Measurement is based on the amount needed to fulfil all benefit obligations when invested in a portfolio of securities free of default risk whose cash flows match the future benefit payments. It is intended to fulfil the benefit obligation without additional funds.
- *The Budget Value- An Expected Return:* For this method, Measurement is based on the anticipated amount that is expected to be sufficient to pay all benefits when due, thus, if that amount is invested and earns the anticipated return of the plan's investment portfolio. When the portfolio is diversified and the return is uncertain, additional funds may be needed when the actual returns are less than the expected returns, and surplus assets may develop when the actual returns are greater than the expected return.

However, there is no defined statistical method to predict or determine the future benefit payments. In Ghana, the Social Security and National Insurance Trust (SSNIT) is a statutory public Trust charged under the PNDC Law 247 and National Pensions Act 2008 Act 766 with the administration of Ghana's Basic National Social Security Pension Scheme and to cater for the first tier of the contributory three-tier scheme in the Act 766. It operates the Defined Benefit type of pension compared with the Defined Contribution type which is managed by privately owned fund managers. The Trust is currently the largest non-bank financial institution in the country.

The primary responsibility is to replace part of lost income of Ghanaian workers or their dependents The Pension Scheme as administered by SSNIT has a registered membership of approximately 1,307,882 million as at August 2017 with over 184,761 pensioners who regularly receive their monthly pensions from SSNIT. The annual absolute growth of pensioners is over 12,000.

For the Scheme to adapt to a changing environmental and social landscape, to have good governance, and to manage risks effectively over both the short and long term, it has to accurately predict various aspects of its operations. Some of these areas include: the number of contributors, amount of revenue and expenditure, pensioners and its associated commitments etc. Not forgetting the economic and social indicators of the country at large. (www.ssnit.org.gh)

2. Problem Statement

The primary objective of a pension scheme is to pay benefits to its qualified members as and when the time is due and as such, there is the need for it to meet its current and future obligations. The literature reviewed on the current three-tier pension scheme indicates that less attention has been given to the evaluation and the sustainability of the scheme in Ghana. There are numerous concerns that have been raised by public sector workers over the inadequacies inherent in the level of pension to sustain a respectable life for the aged upon retirement. Most workers highlighted the low pension received under the Social Security and National Insurance Trust (SSNIT) compared to workers under CAP 30 of the 1950 British Colonial Ordinance (Pension Ordinance No.42) as discriminatory. It is usually easy to value the assets of a social security scheme since it usually holds liquid securities such us stocks and bonds as compared to liabilities which can be very difficult to value. In view of this, this study seeks to analyse and estimate the value of future liability of the scheme. Aside the economic, demographic and actuarial assumptions, statistical analysis and some assumption must be made to determine the total value of pension benefits that must be paid out in the future. Secondly, an assumption must be made on the expected growth of the scheme's assets which will allow it meet those obligations.

3. Research Objectives

- To determine the distribution of the number of benefits that is paid on old age pension.
- To determine the distribution of the amount of benefits paid to old age pensioners.
- To estimate the parameters: the number of old age pensioners and the amount of benefit.

3.1. Research Method

3.1.1. Data Source

The study made use of secondary data in gathering information. This data was retrieved from the SSNIT head office in Accra.

3.2. Methods of Data Analysis

The study was done on old age pension benefits of Social Security and National Insurance Trust (SSNIT) in Ghana. The number of members of the scheme who are under this benefit scheme and the total amounts that are being paid to this category of people will analysed and also determine the distribution that best fits. This will bring additional knowledge/ understanding/biases based on the distribution assumptions. The number of people who receives these benefits will follow a discrete distribution whiles the amount of benefits that are being paid will follow a continuous distribution. Examples of the discrete and continuous distributions that these are likely to follow are discussed below respectively.

3.2.1. Poison Distribution

Poison distribution in statistics is a distribution function useful for characterizing events with very low probabilities of occurrence within some definite time or space. It is used to model the number of events occurring within a given time interval. A random variable X is said to be a Poisson random variable with parameter λ >0 if its probability mass function has the form

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \cdots$$

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0, 1, 2, ... \qquad (3.1)$$

Where λ indicates the average number of successes per unit time or space.

Let X be a binomial random variable with parameters n and p. If $n \to \infty$ and $p \to 0$ so that $np = \lambda = E(X)$ remains constant then X can be approximated by a Poisson distribution with parameter λ.

3.2.2. Negative Binomial Distribution

In probability theory and statistics, the negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted r) occurs. For example, if we define a 1 as failure, all non-1s as successes, and we throw a die repeatedly until 1 appears the third time (r = three failures), then the probability distribution of the number of non-1s that appeared will be a negative binomial distribution.

3.3. Exponential Distribution

In probability theory and statistics, exponential distribution (also known as negative exponential distribution) is the probability distribution that describes the time between events in a poison point process, i.e., a process in which events occur continuously and independently at a constant average rate (en.m.wikipedia.org).

An exponential random variable with parameter λ >0 is a random variable with pdf

$$\left.\begin{aligned} f(x) &= \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \qquad F(X) = 1 - e^{-\lambda x}, \quad x > 0. \\ E(X) &= \frac{1}{\lambda} \\ \operatorname{Var}(X) &= \frac{1}{\lambda^2}. \end{aligned}\right\} \qquad (3.2)$$

Exponential random variables are often used to model arrival times, waiting times, and equipment failure times.

The moment generating function of an exponential distribution with parameter λ is given by

$$M_X(t) = \frac{1}{1 - \lambda t}. \qquad (3.3)$$

3.4. Log Normal Distribution

A lognormal (log-normal or Galton) distribution is a probability distribution with a normally distributed logarithm. A random variable is log normally distributed if its logarithm is normally distributed. Skewed distributions with low mean values, large variance, and all-positive values often fit this type of distribution. Values must be positive as log(x) exists only for positive values of x.

The probability density function is defined by the mean μ and distributions with low mean values, large variance, and all-positive values often fit this type of distribution. Values must be positive as log(x) exists only for positive values of x.

The probability density function is defined by the mean μ and standard deviation, σ:

$$\mathcal{N}(lnx; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(lnx - \mu)^2}{2\sigma^2}\right], \ x > 0. \qquad (3.4)$$

The shape of the lognormal distribution is defined by three parameters: σ, the shape Parameter. Also, the standard deviation for the lognormal, this affects the general shape of the distribution. Usually, these parameters are known from historical data. Sometimes, you might be able to estimate it with current data. The shape parameter doesn't change the location or height of the graph; it just affects the overall shape. m, the scale parameter (this is also the median). This parameter shrinks or stretches the graph. Θ (or μ), the location parameter, which tells you where on the x-axis the graph is located.

3.5. Characteristics of Lognormal Distribution

- It is skewed to the right
- The pdf starts at zero, increases to its mode, and decreases thereafter.
- The degree of Skewness increases as $\sigma^{'}$ increases, for a given $\mu^{'}$

After analysing the distributions, another statistical tool (Trend Analysis) will also be used to forecast the number of people likely to go on pension and the total future liabilities.

3.6. Trend Analysis

Trend analysis refers to techniques for extracting an underlying pattern of behaviour in time series. The trend may be linear, quadratic or exponential in nature. From Minitab, the trend equations are given as follows;

$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$ (3.5) *Linear Trend Model*

$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$ (3.6) *Quadratic Trend Model*

Y=$\beta_0 \times \beta^t{}_1 \times \varepsilon_t$ (3.7) *Exponential Trend Model*

Where;

Y= dependent variable, β_0= intercept or coefficient, β_1, β_2= independent variables, ε = error term

t = index of time period

The trend with the minimum mean absolute percentage error (MAPE), Mean absolute deviation (MAD) and Mean squared Deviation (MSD) is the best trend for the series. To be able to come up with an adequate model and also to make accurate forecast, the concept of time series will also be employed.

3.7. Differencing

A time series that is non-stationary can be made stationary by taking the first-difference. The first-difference is simply the difference of the value of the series at times *t* and *t – 1*

$y_t = x_t - x_{t-1}$

Where x_t is the original time series and y_t is the first-differenced series. The number of observations in the differenced series y_t would be one less than the number of observations in the original series. But if the series is also not stationary in the rate of change of the mean (i.e. slope); stationary can be achieved by taking the second difference.

3.8. Stationarity and Non-Stationarity

Stationary series vary around a constant mean level, neither decreasing nor increasing systematically over time, with constant variance. Non-stationary series have systematic trends, such as linear, quadratic, and so on. A non-stationary series that can be made stationary by differencing is called "non-stationary in the homogenous sense".

Stationary is used as a tool in time series analysis, where the raw data are often transformed to become stationary. For example, economic data are often seasonal or dependant on a non-stationary price level. Using non-stationary time series produces unreliable and spurious results and leads to poor understanding and forecasting. The solution to the problem is to transform the time series data so that it becomes stationary. If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by differencing. Differencing the scores is the easiest way to make anon-stationary mean stationary (flat). The number of times you have to difference the scores to make the process stationary determines the value of *d.* If d = 0, the model is already stationary and has no trend. When the series is differenced once, d=1 and linear trend is removed. When the difference is then differenced, d =2 and both linear and quadratic trend are removed. For non-stationary series, d values of 1 or 2 are usually adequate to make the mean stationary. If the time series data analysed exhibits a deterministic trend, the spurious results can be avoided by detruding. Sometimes the non-stationary series may combine a stochastic and deterministic trend at the same time and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend.

A non-stationary process with a deterministic trend becomes stationary after removing the trend, or detrending. For example, Y_t = α + β_t+ $\varepsilon_t$is transformed into a stationary process by subtracting the trendβ_t: Y_t - β_t = α +ε_t. No observation is lost when detrending is used to transform a non-stationary process to a stationary one. Non-stationary data, as a rule, are unpredictable and cannot be modelled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the non-stationary data needs to be transformed into stationary data. In contrast to the non-stationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time.

3.9. Autocorrelation Function (ACF)

Autocorrelation is the correlation that exists between successive values of the same variables. Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also called "lagged correlation" or "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. Positive autocorrelation might be considered a specific form of "persistence", a tendency for a system to remain in the same state from one observation to the next. For example, the likelihood of tomorrow being rainy than if today is dry. Autocorrelation complicates the application of statistical tests by reducing the number of independent observations. Autocorrelation can also complicate the identification of significant covariance or correlation between time series. Autocorrelation can be exploited for predictions: an auto correlated time series is predictable, probabilistically, because future values depend on current and past values. Three tools for assessing the autocorrelation of a time series are;

- The time series plot
- The lagged scatter plots
- The autocorrelation functions

An important guide to the persistence in a time series is given by the series of quantities called the sample autocorrelation coefficients, which measure the correlation between observations at different times. The set of autocorrelation coefficients arranged as a function of the separation in time is the sample autocorrelation function. The first-order autocorrelation coefficient is the sample coefficient of the first *N—1* observations, *t=1,2,…,N—1* and the next *N—1* observations, *t=2,3,…,N*. The correlation is given by

$$\rho_k = \frac{\epsilon\{(y_t-\mu)(y_{t+k}-\mu)\}}{\epsilon\{(y_t-\mu)^2\}} = \frac{\gamma(k)}{\gamma(0)} \qquad (3.8)$$

The quantity is called the autocorrelation coefficient at lag k. the collection of the values of ρ_k for k = 0, 1, 2… is referred to as the autocorrelation function (ACF).

The plot of the autocorrelation function of the lag is also called the correlogram. The autocorrelation function can be used for the following two purposes;

- To detect non-randomness in data
- To identify an appropriate time series model if the data are not random.

3.10. Partial Autocorrelation Function

Partial autocorrelation function gives the partial correlation of time series with its own lagged values, controlling for the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags. The partial autocorrelation function (PACF) denoted by the set of partial autocorrelations at various lags k are defined by (k=1, 2, 3…). The set of partial autocorrelations at various lags k are defined by;

$$\pi_k = \frac{\rho_k - \sum_{j=1}^{k-1}(\pi_{k-1,j})}{1-\sum_{j-1}^{k-1}(\pi_{k-1,j})(\rho_{k-j})}, k > 1, j = 1,2,3,\dots,k-1 \qquad (3.9)$$

Partial autocorrelation function plays an essential role in data analyses aimed at identifying the extent of the lag in an autoregressive model. The partial autocorrelation of an AR(ρ) process is zero at lag ρ+ 1 and greater. The approximate 95% confidence interval for the partial autocorrelations is at $\pm\frac{1.96}{\sqrt{n}}$ or $\pm\frac{2}{\sqrt{n}}$ where n is the record length (number of points) of the time series being analyzed.

3.11. Arima Model

ARIMA under time series analysis is an autoregressive integrated moving average model which is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). They are also applied in some cases where data depicts evidence of non-stationary, where an initial differencing steps (corresponding to the "integrated" part of the model) can be applied or introduced non-stationary. Non-seasonal ARIMA models are generally denoted by ARIMA (*ρ, d, q*) where the parameters *ρ*, *d*, and q are non-negative integers, *p* is the order of the autoregressive model, *d* is the degree of differencing, and *q* is the order of the moving average model. The process should be stationary after differencing a non-seasonal process *d* time. The ARIMA (*ρ, d, q*) is given by;

$$(1-B)^d y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \qquad (3.10)$$

In terms of the backward shift operator, the model can be written as

$$(1-\mu_1\beta - \mu_2\beta^2 + \cdots + \mu_p\beta^p) \qquad (3.11a)$$

$$(1-B)^d y_t = (1+\theta_1\beta + \theta_2\beta^2 + \theta_3\beta^3 + \cdots + \theta_q\beta^q)\varepsilon_t \qquad (3.11b)$$

3.12. SARIMA Model

The seasonal ARIMA model is a general class of models used to forecast a time series entirely from its own history. The model uses autoregressive terms and moving average terms to factor in the seasonality of the data trend and differences from time period to time period when making a prediction. This implies that, the SARIMA model extends the ARIMA model to capture seasonal and non-seasonal behaviour. The model is usually written as ARIMA (*ρ, d, q*) × (*P, D, Q*) where; *p, d, q* are the orders of non-seasonal AR, differencing and MA respectively. *P, D, Q* is the orders of seasonal AR, differencing and MA respectively and *s* represents the seasonal order.

3.13. Model Selection

When fitting models, there is the tendency of two or more models competing and for that reason it is appropriate to use a good model selection criterion to select the most adequate model. The possible models are determined based on the data pattern. In this study, AIC and BIC were the measures of the goodness-of-fit that were employed to select the most adequate model. For a given data set, several competing models may be ranked according to their AIC and BIC values with the one having the lowest information criterion value being the best. These information criterion attempts to find the model that best explains the data with the minimum of free parameters but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages over fitting (Aidoo, 2010). The following criterions are given as;

$$\left.\begin{array}{l} \text{AIC} = 2k - 2\log L \\ \text{OR} \\ \text{AIC} = 2k + n\log(\frac{RSS}{n}) \\ \text{AIC}_C = AIC + \frac{2K(K+1)}{n-k-1} \\ \text{BIC} = -2\log L + k\log n \\ \text{where} \end{array}\right\} \qquad (3.12)$$

K: is the number of parameters in the statistical model.
L: is the maximized value of the likelihood function for the estimated model.
RSS: is the residual sum of square of the estimated model.
n: is the number of observation or the sample size

The AIC$_c$ is a modification of the AIC by Hurvich and Tsai (1989) and it is AIC with the second order of correction for small sample sizes. Burnham & Anderson (1998) insist that since AIC$_c$ converges to AIC as *n* gets large, AIC$_c$ should be employed regardless of the sample size (*n*).

3.14. Model Diagnostics (Goodness of Fit)

Ideally, a model should extract all systematic information from the data. The part of the data unexplained by the model (that is the residuals) should be small. The diagnostics check is used to determine the adequacy of the chosen model. These checks are usually based on the residuals of the models. One assumption of ARIMA (*p, d, q*) is that, the residuals of the model should be white noise. A series $\{\varepsilon_t\}$ is said to be white noise if $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variable with constant mean usually assumed to be zero and constant variance.

Also, if $\{\varepsilon_t\}$ is normally distributed with mean zero and variance $\sigma_\varepsilon{}^2$, then $\sigma_\varepsilon{}^2$ is called a Gaussian White noise. For a white noise series, all the ACF are zero. In practice, if the residuals of the model are white noise, then the ACF of the residuals are approximately zero. In order to use the developed model to draw any meaningful conclusion, a statistical tool such as Ljung-Box Q statistic can be used to determining whether the series is independent or not (Smart, 2013).

3.15. Ljung-Box Test

The Ljung box test is used to test for serial correlation in the residuals of the model. The hypothesis to be tested is given by:

H_0 : There is no serial correlation in the residuals of the model.
H_1 **:** There is serial correlation in the residuals of the model.
The test statistic is given by;

$$Q_m = T(T+2)\sum_{k=1}^{m}(T-K)^{-1}\gamma_k{}^2 \qquad (3.13)$$

where
$\boldsymbol{\gamma}$**:** is the sample autocorrelation at lag k
T**:** is the number of observations
K**:** is the lag
When the p-value associated with is large, the model is considered adequate.

3.16. Arch-Lm Test

The ARCH-LM test is used to test for conditional variance or heteroscedasticity in the model residual. The hypothesis is given by:

- H_0 **:** There is no conditional heteroscedasticity in the residuals of the model.
- H_1**:** There is conditional heteroscedasticity in the residuals of the model.

The test statistic is given by:

$$LM = TR^2 \qquad (3.14)$$

T**:** is the number of observations
R^2 **:** is the coefficient of determination computed from the auxiliary residual regression.

3.17. Forecasting

After a model has passed the entire diagnostic test, it becomes adequate for forecasting. The ARIMA model as described by several researchers has proven to perform well in terms of forecasting as compared to other complex models. To choose a final model for forecasting, the accuracy of the model must be higher than that of all the competing models. Forecasting with this system is straight forward with the expected forecast value evaluated at a particular point in time. Confidence intervals may also be easily derived from the standard error of the residuals.

3.18. Statistical Software Appropriate for the Analysis

- Easyfit software was used to run all the distributions to know the one that best fit.
- Mnitab was also used for the preliminary analysis, trend analysis and normality test.
- The *R* studio statistical software was also used in the analysis, ACF and PACF, time series graph, fitting ARIMA models and forecasting.

4. Results and Discussions

4.1. Best Fit Distribution Results for Number of Pensioners

#	Distribution	Kolmogorov Smirnov		Anderson Darling	
		Statistic	Rank	Statistic	Rank
1	D. Uniform	0.12241	2	19.668	3
2	Geometric	0.53793	3	26.596	4
3	Logarithmic	0.83414	5	85.018	5
4	Neg. Binomial	0.11644	1	2.2494	2
5	Poisson	0.54762	4	-29.951	1
6	Bernoulli	No fit (data max > 1)			
7	Binomial	No fit			
8	Hypergeometric	No fit			

Table 1: Goodness of Fit – Summary

Neg. Binomial					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	84 0.11644 0.18937 1				
A	0.2	0.01	0.05	0.02	0.01
Critical Value	0.11508	0.13148	0.14605	0.16331	0.17523
Reject?	Yes	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	84 2.2494 2				
A	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	No	No	No

Table 2: Goodness of fit details

From table 4.1, Negative Binomial ranked first and second in Kolmogorov Smirnov and Anderson Darling test of best fit respectively. Hence, it is the distribution that best fits the data.

Table 2 Goodness of fit details

From Table 2 above, the Kolmogorov-Smirnov test revealed that, the data should be rejected only at confidence level of 80% but should not be rejected at 90, 95, 98 and 99 percent confidence level. Also, the Anderson-Darling test also revealed that, it should be rejected at 80% and 90% confidence level but should not be rejected at 95, 98 and 99 percent level of confidence. Below is diagram of the probability density function (Negative Binomial) of the data.

Figure 4.1 Negative Binomial Distribution

4.2. Descriptive Statistics for Number of Pensioners at SSNIT

Statistic	Value	Percentile	Value
Sample Size	84.00	Min	108730.00
Range	80815.00	5%	110130.00
Mean	140840.00	10%	112520.00
Variance	602490000.00	25% (Q1)	118890.00
Std. Deviation	24546.00	50% (Median)	134800.00
Coef. of Variation	0.17	75% (Q3)	160280.00
Std. Error	2678.10	90%	178210.00
Skewness	0.46	95%	184760.00
Excess Kurtosis	-1.07	Max	189550.00

Table 3: Descriptive Statistics of Responses

4.3. Results Discussion

Table 2 revealed that, the best statistical distribution that best fits the data (Number of pensioners) is Negative Binomial Distribution. Generally, Negative Binomial Distribution is a discrete probability Distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified number of failures occurs. Relating this to the data, the success is the number of people who will go on pension out of the total number of contributors of the SSNIT pension scheme in each month and the failure is the number of people who do not qualify for old age pension.

Also, this distribution is rightly skewed. This means that, for most months the number of pensioners on the scheme is less than the mean, since in a rightly skewed distribution of a data, the mode and the median is always less than the mean (average), nonetheless, there are still some small extreme values which are greater than the mean for some few months.

From the descriptive statistics of the number of pensioners at SSNIT from 2011 to 2017, It could be observed that, the minimum number of pensioners was 108730 and the maximum was 189550. This means that, the number of people on pension at SSNIT each month ranges from 108730 to 189550. It appears that, the average number of pensioners each month is 140840 which implies that the least expected number of people who will be on pension is 140840 with Median 134800 depicting that at SSNIT, the number of people likely to be on pension is more or less than 134800.

4.3.1. Best Fit Distribution for Total Monthly Payment to Pensioners at SSNIT (Continuous)

Analysis revealed that, Lognormal (3P) distribution ranked best as compared to the other continuous statistical distribution for the Kolmogorov Smirnov, Anderson Darling and Chi-Square test of fit.

4.3.2. Parameters of the Lognormal (3p) Distribution

Distribution	**Shape**	**Location**	**Scale**
Lognormal	σ = 1.2921	μ = 16.899	γ = 1.6781E+7

Table 4: Parameters

4.3.3. Goodness of Fit Results of Lognormal (3p)

Kolmogorov-Smirnov					
Sample Size	84				
Statistic	0.11948				
P-Value	0.16747				
Rank	5				
⍰	0.2	0.1	0.05	0.02	0.01
Critical Value	0.11508	0.13148	0.14605	0.16331	0.17523
Reject?	Yes	No	No	No	No
Anderson-Darling					
Sample Size	84				
Statistic	2.1619				
Rank	12				
⍰	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	No	No	No
Chi-Squared					
Deg. of freedom	5				
Statistic	4.7803				
P-Value	0.44327				
Rank	5				
⍰	0.2	0.1	0.05	0.02	0.01
Critical Value	7.2893	9.2364	11.07	13.388	15.086

Table 5: Goodness of Fit

From the goodness of fit table above, the Kolmogorov-Smirnov test revealed that, the data should be rejected only at confidence level of 80% but should not be rejected at 90, 95, 98 and 99 percent confidence level. Also, the Anderson-Darling test also revealed that, it should be rejected at 80% and 90% confidence level but should not be rejected at 95, 98 and 99 percent level of confidence. Nonetheless, the Chi-Square showed that it should not be rejected for all 80%, 90%, 95%, 98% and 99% level of confidence.

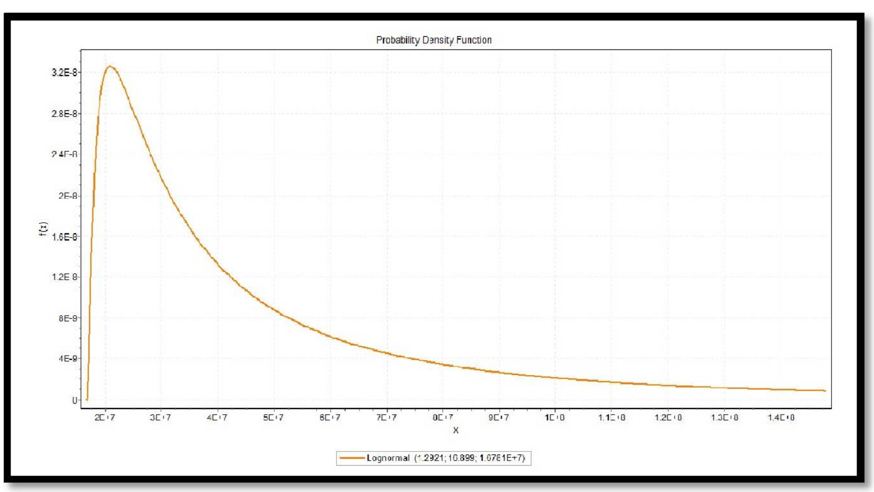

Figure 1: Probability Density Function for Lognormal Distribution

4.4 Descriptive Statistics for Total Monthly Pension Payments

Statistic	Value	Percentile	Value
Sample Size	84.00	Min	18312435
Range	107210000.00	5%	18527000.00
Mean	56877047	10%	19150000.00
Variance	1265486000000000.00	25% **(Q1)**	24770000.00
Std. Deviation	35574000.00	50% **(Median)**	47957000.00
Coef. of Variation	0.63	75% **(Q3)**	92083000.00
Std. Error	3881400.00	90%	116700000.00
Skewness	0.65	95%	122300000.00
Excess Kurtosis	-1.01	Max	125521739

Table 6: Descriptive Statistics

4.5. Discussion of Results

Table 4 revealed that, the Lognormal (3P) Distribution is the best fitted distribution for the data. Generally, a Lognormal distribution results if the variable is the product of a large number of independent, identically distributed variables. From the diagram (Figure 1), it can be observed that the data is rightly Skewed. This implies that the amount paid to pensioners is fairly symmetrical and values extend towards more positive values, meaning the data of amount paid to pensioners has small values and few large values.

From Table 5 it could be observed that, the minimum amount paid out was GHC 18,312,435 and the maximum was GHC 125,521,739. This means that, the amount paid by SSNIT to pensioners each month ranges from GHC 18,312,435 to GHC 125,521,739. It appears that, the average amount paid each month is GHC 56,877,047 which imply that the least expected amount to be paid is GHC 56,877,047 with Median GHC 47957000.00 depicting that at SSNIT, the amount likely to be paid to pensioners is more or less than GHC 47957000.00.

4.6. Trend Analysis

This shows the models that will best suit the number of pensioners and the amount paid by studying the measures of accuracy of the linear, exponential and quadratic models.

Model	MAPE	MAD	MSD
Linear	3	4122	21822006
Quadratic*	1*	978*	1576231*
Exponential Growth	2	2641	9157702

Table 7: Measure of Accuracy for Number of Pensioners
*NB: * Implies Best Model*

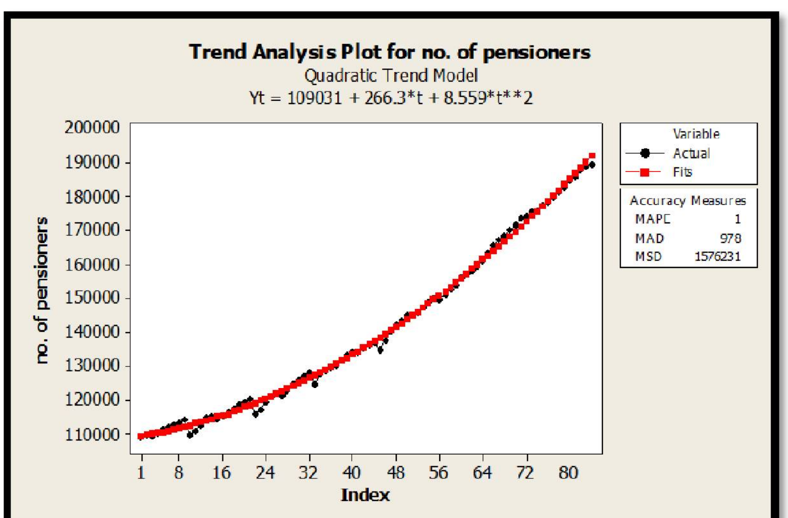

Figure 2: Trend analysis of number of pensioners

Model	MAPE	MAD	MSD
Linear	2.61607E+01	9.42682E+06	1.92328E+14
Quadratic*	1.39388E+01*	5.39635E+06*	1.18262E+14*
Exponential Growth	1.43454E+01	6.22903E+06	1.20400E+14

Table 8: Measure of Accuracy for Total Amount Paid
*NB: * Implies Best Model*

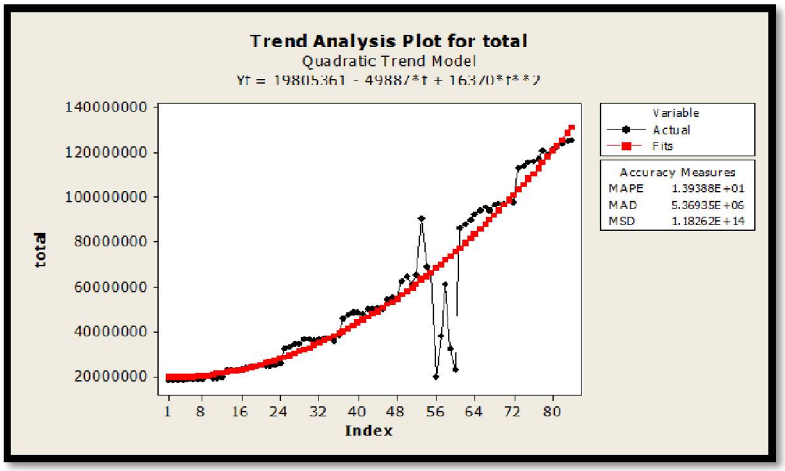

Figure 3: Trend Analysis of Total Monthly Payments

From Tables 4.6 and 4.7, it can be observed that the quadratic trend has the least Mean Absolute Percentage Error (MAPE) of 1 and 1.39388E+01 respectively, Mean Absolute Deviation (MAD) of 978 and 9.42682E+06 respectively and Mean Square Deviation (MSD) of 1576231 and1.18262E+14 for number of Pensioners and amount paid respectively. This implies that, the data follows a quadratic trend. Therefore, the fitted models for number of pensioners and total monthly payments are given respectively as 4.1 and 4.2 below;

$Yt = 109031 + 266.3t + 8.559t^2$ (4.1)

$Yt = 19805361 – 49887t + 16370t^2$ (4.2)

4.6.1. Time Series Plot of Raw Data of the Number of Pensioners

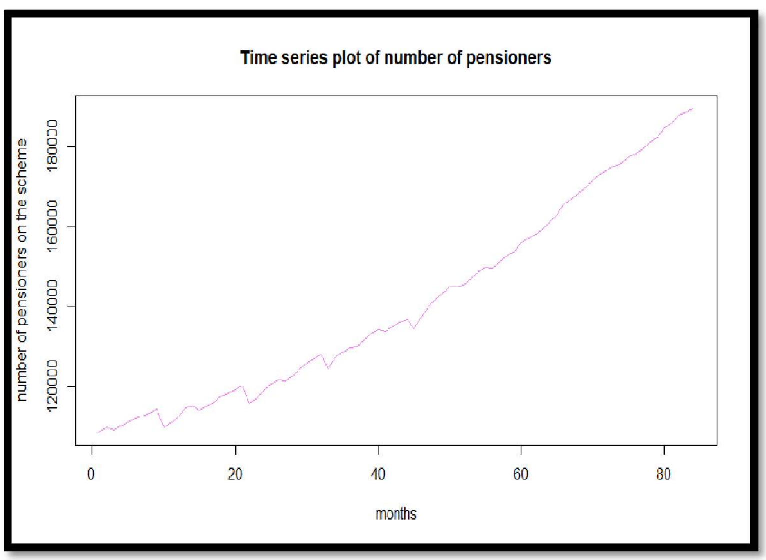

Figure 4: Plot for Number of Pensioners

The plot in figure 4.8.1 above is plot of the number of pensioners in SSNIT from 2011 to 2017. The plot shows an increasing trend in the number of pensions on the scheme. By default, one would expect that this data is not stationary and involves no seasonality by a mere pictorial inspection of the graphs, but there is the need to perform further test to really ascertain whether the data is stationary or non-stationary.

4.6.2. Test for Stationarity of the Raw Data

	ADF Test (*)	KPSS Test (**)
Test Statistic	-1.2764	2.819
P-value	0.8727	0.01
Stationary?	NO	NO
(a); H0: The data is non-stationary		
(b); H0: The data is stationary		

Table 9: Stationarity Test

Table 8above reveals that the raw data on number of pensioners at SSNIT is not naturally stationary by default. Hence the data requires differencing in order to make it stationary.

4.6.3. Finding the First Difference of Raw Data and testing for stationarity

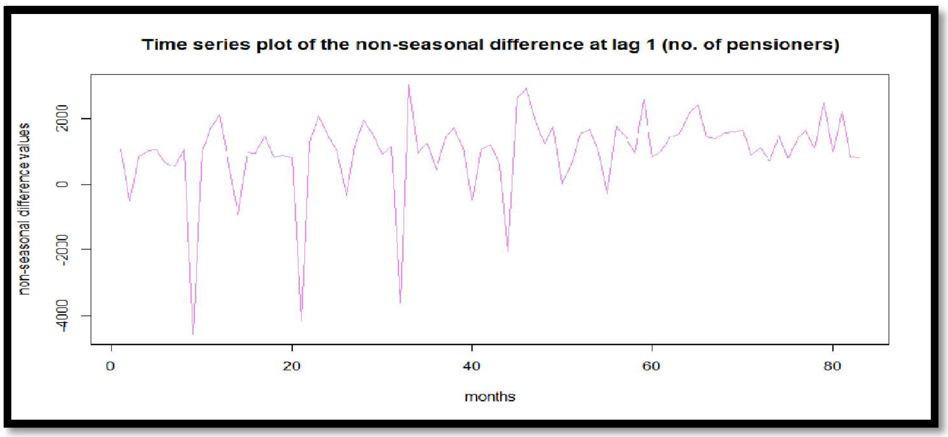

Figure 5: Plot of first difference

It will be difficult to really tell if the graph shown in Figure 5 is stationary or not. It possesses characteristics of possible stationarity and non-stationarity as the values seem to be distributed around zero, suggesting the possibility of them having a zero mean but some unusual spikes detected in the left tail also suggest possible non-stationarity. For this reason, further test needs to be performed at this point also to check for stationarity in the first difference.

4.6.4. Test for Stationarity of the First Difference (Number of Pensioners)

Test type	ADF Test (*)	KPSS Test (**)
Test Statistic	-5.493	0.7658
P-value	0.01	0.1
Stationary?	YES	NO
*; H0: The data is non-stationary		
**; H0: The data is stationary		

Table 10: Test for Stationarity

The results from table 4.8.4 shows conflicting responses from the ADF test and the KPSS test. The ADF test, we test the null hypothesis that the first differenced series is non-stationary. The table shows a test-statistics of -5.493 and a *p*-value of 0.01 which is less than the alpha value (α) of 0.05. This tells us to fail to reject the alternative hypothesis at 5% significance level; hence the differenced series is stationary. But the KPSS test v rather reports the presence of non-stationarity in the first difference, hence there is the need to take the second difference of the data until the two tests of stationarity agree.

4.6.5 Finding the Second Difference of Raw Data and Testing for Stationarity

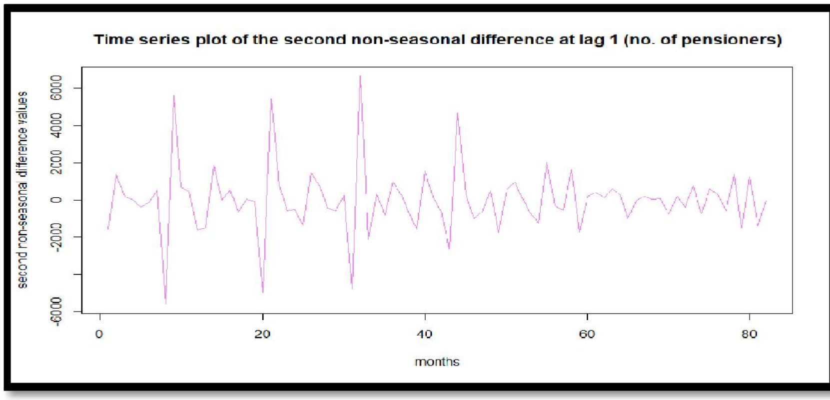

Figure 6: Plots for Second Difference

At this point in Figure 4.8.5, one can suspect stationarity since the values seem to be grouped around zero. Higher spikes are accompanied by other higher spikes in the opposite direction which stands to cancel themselves. However, results from further empirical stationarity tests will prove or disprove this fact.

4.6.6. Test For Stationarity of The Second Difference (Number of Pensioners)

Test type	ADF Test (*)	KPSS Test (**)
Test Statistic	-7.4405	0.017245
P-value	0.01	0.1
Stationary?	**YES**	**YES**
*; H0: The data is non-stationary		
**; H0: The data is stationary		

Table 11: Stationarity Test

This time in table 4.11, the two stationarity tests agree as they all reveals stationarity in the second difference. Hence, the second difference of the data on number pensioners at SSNIT is to be worked with.

4.7. Model Selection

ARIMA MODEL	AIC	AICc	BIC
ARIMA (1,2,0)	1448.17	1448.32	1452.98
ARIMA (1,2,1)	1412.82	1413.12	1420.04
ARIMA (2,2,0)	1441.85	1442.16	1449.07
ARIMA (2,2,1)	1412.7	1413.22	1422.33
ARIMA (3,2,0)	1433.61	1434.13	1443.24
ARIMA (3,2,1)	1410.7	1411.49	1422.73
ARIMA (4,2,0)	1426.39	1427.18	1438.42
ARIMA (4,2,1)	1411.55	1412.67	1425.99
ARIMA (6,2,0)	1421.53	1423.04	1438.37
ARIMA (6,2,1)	1414.8	1416.77	1434.05

Table 12: Arima Models

Table 12reports the ARIMA (3,2,1) to be the best model for forecasting the number of pensioners at SSNIT since it recorded the smallest AIC, AICc and BIC values.

4.7.1. Model Diagnostics (Ljung-Box and Arch-Lm Test)

Ljung-Box test is performed to determine whether there is serial correlation in the residual of the model. Also, ARCH-LM test is performed to prove heteroscedasticity in the residual of the model.

Test	Lag	Test Statistic	P- Value
L-Jung Box	1	0.2432	0.6219
L-Jung Box	4	2.3278	0.6757
ARCH-LM	4	0.488	0.9747
ARCH-LM	8	1.346	0.995

Table 13: Parameters

Table 13, shows that the *p*-values of the Ljung Box test at both lags 1 and 4 are greater than alpha (α) value of 0.05. The Ljung box test was used to test the null hypothesis of absence of serial correlation in the residual of the model. Now since all the *p*-value is greater than the alpha value of 0.05, we fail to reject the null hypothesis and conclude that there is no serial correlation in the residuals of the model. Hence the model is adequate.

Also, the ARCH-LM test was used to test for the heteroscedasticity in the residual of the model. Now, since it has p-values at both lags 4 and 8 which are higher than the alpha (α) value of 0.05, we fail to reject the null hypothesis and conclude that there is no heteroscedasticity in the residual of the model. Hence the model is adequate.

4.7.2. Forecasting with ARIMA (3,2,1)

Now the suitable model (ARIMA (3, 2, 1)) that fits the data has been identified and has passed through the necessary model test, the next thing will be to forecast to see how future trend of the number of pensioners will be like. The model was then used to forecast for monthly number of pensioners for the next three years of SSNIT as shown in the table below.

		80% Confidence level		90% Confidence level	
Month	**Point Forecast**	**Low**	**High**	**Low**	**High**
Jan-19	207371.9	201232	213511.7	197982	216762
Feb-19	208723.6	202180	215267.3	198716	218731.3
Mar-19	210075.5	203120	217030.7	199438	220712.5
Apr-19	211428.2	204054	218802.3	200151	222705.8
May-19	212781.1	204981	220580.9	200852	224709.8
Jun-19	214133.7	205901	222365.9	201544	226723.8
Jul-19	215486.1	206815	224157.6	202224	228748.1
Aug-19	216838.6	207721	225956.2	202895	230782.8
Sep-19	218191.2	208621	227761.6	203555	232827.9
Oct-19	219543.8	209514	229573.7	204204	234883.2
Nov-19	220896.3	210400	231392.3	204844	236948.5
Dec-19	222248.9	211280	233217.4	205474	239023.7
Jan-20	223601.4	212154	235048.9	206094	241108.8
Feb-20	224954	213021	236886.8	206704	243203.7
Mar-20	226306.6	213882	238731	207305	245308.2
Apr-20	227659.1	214737	240581.5	207896	247422.2
May-20	229011.7	215585	242438.1	208478	249545.7
Jun-20	230364.2	216428	244300.9	209050	251678.5
Jul-20	231716.8	217264	246169.7	209613	253820.6
Aug-20	233069.3	218094	248044.5	210167	255971.9
Sep-20	234421.9	218919	249925.2	210712	258132.2
Oct-20	235774.5	219737	251811.8	211247	260301.5
Nov-20	237127	220550	253704.2	211774	262479.7
Dec-20	238479.6	221357	255602.4	212293	264666.6

Table 14: Forecasted Values for Number of Pensioners
The Table Above Shows the Point Forecasts of the Number of Pensioners at SSNIT for the Next 24 Months

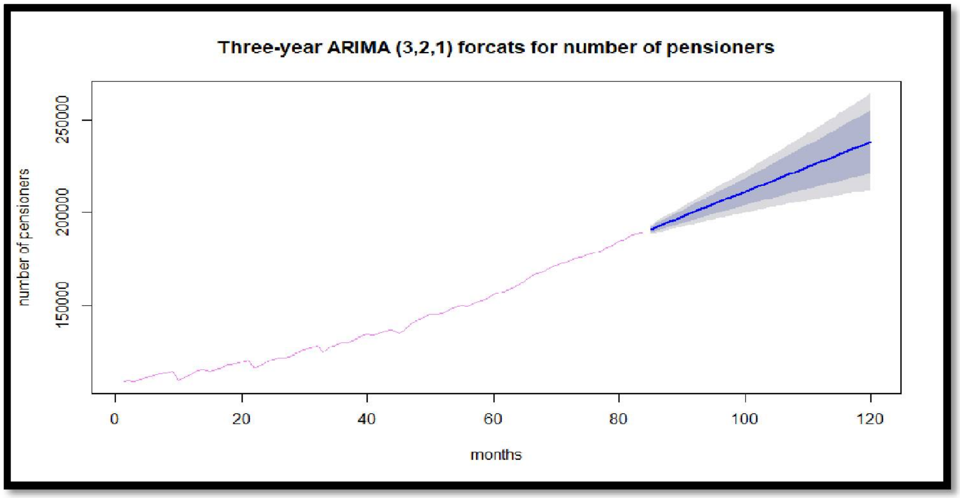

Figure 7: Graph 0f Forecast Values (No of Pensioners)

The forecast graph in Figure 9 shows the number of pensioners at SSNIT is going to increase from month to month for the next 36 months as it is already the case in the original data.

4.8. Model Selection Process (For total Amount)

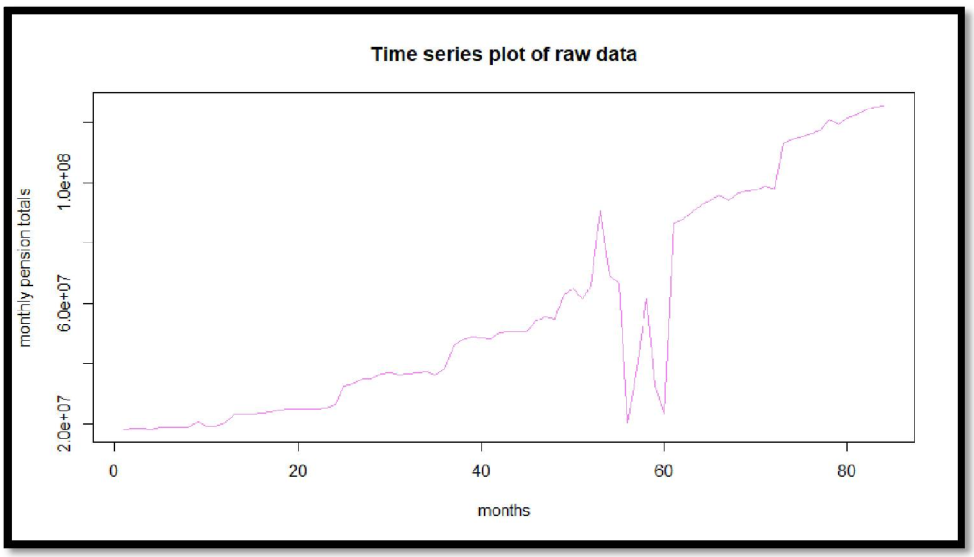

Figure 8: Time Series Plot of Raw Data for Monthly Payment

Figure 8 shows the time series plot of the raw data for monthly pension total amounts. The raw data does not show any seasonality in the data. The plot generally shows an increasing trend and by default one can suspect that there is no stationarity in the data. However, in order to be firm on the fact that there is no stationarity, there is the need to perform further empirical test.

4.8.1. Unit Root and Stationarity Test

This test is performed to further prove the stationarity of the data. It is achieved by testing for the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS).

4.8.2. Test for Stationarity in the Raw Data for Total Monthly Pension Payments

	ADF Test (*)	**KPSS Test (**)**
Test Statistic **Lag order**	-2.8454 4	2.5524 2
P-value	0.2291	0.01
Stationary?	**NO**	**NO**
(*); H0: The data is non-stationary		
(**); H0: The data is stationary		

Table 15: Test for Stationary on Raw Data (Pension Total)

From Table 15, using the ADF test, we test the null hypothesis that the original series is non-stationary against the alternative of stationary. The test results show a test statistic of -2.8454 and *p*-value of 0.2291. Since the P-value is more than the alpha value (α) of 0.05. We therefore fail to reject the null hypothesis which states that the original series is non-stationary.

Also using the KPSS test, we test the null hypothesis that the original series is stationary against the alternative of non-stationary. The results depict a test statistic of 2.5524 and a *p*-value of 0.01. Since the *p*-value is less than the alpha value of 0.05, we fail to reject the alternative hypothesis which shows non-stationary in the original series.

The above reveals by empirical test that the data at this stage is not stationary by default. Hence, there is the need to difference the data.

4.8.3. Taking the First Difference of the Original Data and Testing for Stationarity

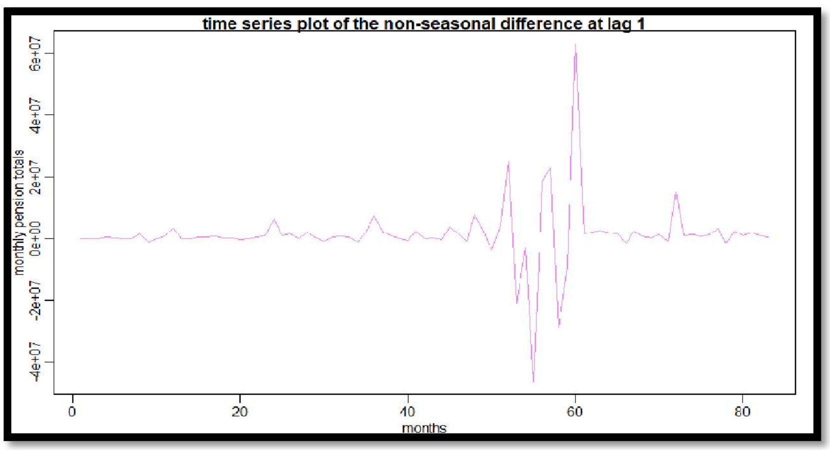

Figure 9: Plot for First Difference

From Figure 9, one can now begin to suspect stationarity in the first non-seasonal difference of the data. However, some long spikes are seen towards the right end of the plot which rather suggest possible non-stationarity even in the first difference. Hence, at this stage, there is the need to perform stationarity test on the first difference to actually determine whether it is stationary or non-stationary.

4.8.4. Test for Stationarity of First Difference for Monthly Totals

Test type	ADF Test (*)	KPSS Test (**)
Test Statistic	-4.414	0.080478
P-value	0.01	0.1
Stationary?	YES	YES
*; H0: The data is non-stationary **H0: The data is stationary		

Table 16: Stationarity of First Difference

From Table 4.16, using ADF test, we test the null hypothesis that the first differenced series is non-stationary. The table shows a test-statistic of -4.414 and a *p*-value of 0.01 which is less than the alpha value (α) of 0.05. This tells us to fail to reject the alternative hypothesis at 5% significance level; hence the differenced series is stationary. Using the KPSS test,

we test the null hypothesis that the first difference series is stationary. From all indications shows a *p*-value higher or greater than alpha value (α) of 0.05 telling us of failing to reject the null hypothesis and conclude that the first differenced series is stationary.
This time, the test for stationarity reveals that the data for pension monthly totals is stationary in its first difference. Hence, time series model can be built only when the data has been differenced in the first place.

4.9. Determining the Lags of the Arima Model

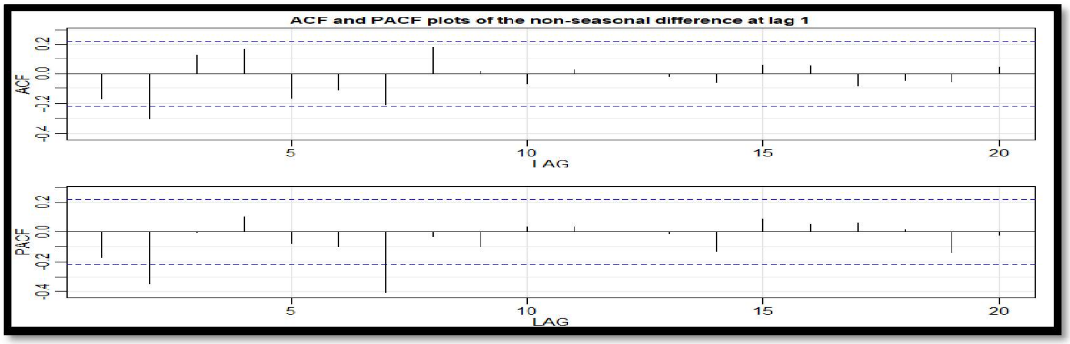

Figure 10: Plots for First Difference

Figure 10 shows the ACF and PACF plots of the non-seasonal difference, from the ACF plot, significant spike that goes beyond the confidence region is seen at lag 2. However, the PACF plot clearly shows that those spikes are found at lags 2 and 7. Hence, Autoregressive and Moving Average components of our ARIMA model will most likely be at lag 2 or 7. The difference (integrated) component is to be at lag 1 since only the first difference taken.

4.10. Model Selection

ARIMA MODEL	AIC	AICc	BIC
ARIMA (2,1,0)	2917.94	2918.24	2925.2
ARIMA (2,1,2)	2912.73	2913.51	2924.83
ARIMA (7,1,0)	2915.55	2917.49	2934.9
ARIMA (7,1,2)	2919.25	2922.3	2943.43

Table 17: Arima Models

Based on the AICs, AICc, and BICs found in Table 17, the ARIMA (2,1,2) was chosen to be the best ARIMA model since it has the least values. Hence, it will be implemented in this project and will be used for forecasting monthly total pension payments at SSNIT.

4.10.1. Model Diagnostics (Ljung-Box and Arch-Lm Test)

Test	**Lag**	**Test Statistic**	**P- Value**
L-Jung Box	1	0.0070593	0.933
L-Jung Box	4	1.6637	0.7973
ARCH-LM	4	5.05	0.2825
ARCH-LM	8	10.33	0.2427

Table 18: Test Parameters

Table 18, shows that the *p*-values of the Ljung Box test at both lags 1 and 4 are greater than alpha (α) value of 0.05. The Ljung box test was used to test the null hypothesis of absence of serial correlation in the residual of the model. Now since all the *p*-value is greater than the alpha value of 0.05, we fail to reject the null hypothesis and conclude that there is no serial correlation in the residuals of the model. Hence the model is adequate.
Also, the ARCH-LM test was used to test for the heteroscedasticity in the residual of the model. Now, since it has p-values at lags 8 and 4 which are higher than the alpha (α) value of 0.05, we fail to reject the null hypothesis and conclude that there is no heteroscedasticity in the residual of the model. Hence the model is adequate.

4.10.2. Forecasting with ARIMA

Now the suitable model (ARIMA (2, 1, 2)) that fits the data has been identified and has passed through the necessary model test, the next thing will be to forecast to see how future trend of total amounts paid to pensioners will be like. The model was then used to forecast for the total monthly payments to pensioners of SSNIT for the next three years of SSNIT as shown in the table below.

		80% Confidence Interval		95% Confidence interval	
Month	Point Forecast	Low	High	Low	High
Jan-19	120240804	81675647	158805961	61260486	179221122
Feb-19	120655902	80634901	160676903	59449063	181862742
Mar-19	121192533	79626581	162758485	57622895	184762171
Apr-19	121207999	78179548	164236450	55401661	187014336
May-19	120824282	76492096	165156469	53024054	188624511
Jun-19	120584571	75030466	166138676	50915580	190253563
Jul-19	120722853	73929906	167515800	49159216	192286490
Aug-19	120980675	72920191	169041159	49159216	194482843
Sep-19	121032880	71736968	170328791	45641289	196424471
Oct-19	120875175	70416187	171334162	43704812	198045537
Nov-19	120743084	69171745	172314424	41871527	199614642
Dec-19	120779937	68104911	173454962	40220437	201339436
Jan-20	120898587	67116461	174680714	38645923	203151252
Feb-20	120942225	66070611	175813839	37023333	204861117
Mar-20	120881322	64957118	176805525	35352633	206410011
Apr-20	120813121	63868181	177758060	33723351	207902890
May-20	120817121	62865679	178768563	32188038	209446204
Jun-20	120869439	61917690	179821188	30710519	211028359
Jul-20	120897649	60958500	180836799	29228631	212566668
Aug-20	120876133	59971183	181781082	27730051	214022215
Sep-20	120842671	58992946	182692395	26251679	215433662
Oct-20	120838508	58057749	183619268	24823623	216853394
Nov-20	120860532	57157839	184563226	23435670	218285394
Dec-20	120876688	56262904	185490472	22058433	219694944

Table 19: Forecasted Values

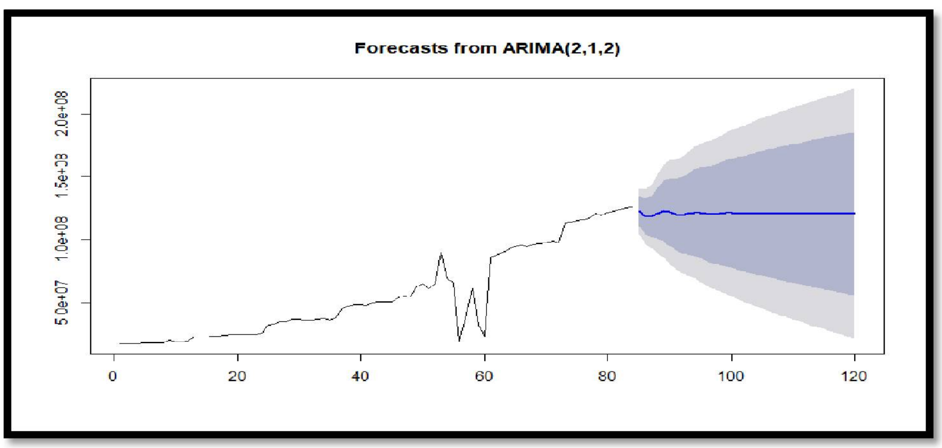

Figure 11: Graph of Forecast Values

Figure 4.16 above is the graph of the three-year forecast of ARIMA (2,1,2), with its 80% and 95% confidence intervals. The forecast reveals that monthly pension totals will reduce shortly and then rise again and after become constant for the next three years. Actually, the point forecast values are not constant throughout the period as the graph is suggesting a constant forecast. They vary in thousands which do not clearly show on the forecast graph.
On this note, it can be concluded that the two ARIMA models built in this project are good time series models forecasting the number of pensioners and monthly pension payments on the scheme for the next 36 months, hence the two models should be implemented.

5. Summary and Recommendation

The improvement in technology, health and pension schemes in line with social protection policies have become pre-occupation of humans due to the unpredictability of life at old age. Due to this, many developed and developing countries have introduced various pensions based on the ability of the countries to project the age and growth rates of their citizens. The study considered the sustainability of old age pension benefit using statistical analysis. The study was based on three objectives namely; to determine the distribution of the number of benefits that is paid on old age pension, to determine the distribution of the amount of benefits paid to old age pensioners, to estimate the parameters: the number of old age pensioners and the amount of benefit.

The introduction of pension in Ghana was to promote loyalty and efficiency in the 1940s within the colonial service. The problem statement is to analyse and estimate the value of future liability of the scheme. The theoretical frameworks and methodology clearly explain the general frameworks behind the study and the approach in achieving the objectives of the study. In order to achieve the purpose of the study, a secondary data was retrieved from the head office of SSNIT in Accra which entailed the number of pensioners and the amount paid out to them from 2011 to 2017 on monthly basis.

The first and second research objective investigated the best statistical distributions that best fits or best describes the nature of the number of pensioners and the total monthly payment of the scheme and its implications. It was discovered that, both number of pensioners and monthly payments followed positively skewed distributions which implied that, there were values that were below the mean for most of the months. Nonetheless, there were also some extreme values which were greater than the mean. The last objective also sought to analyse the trend, best fitted models and forecasted values of both number of pensioners and total monthly payments of the scheme. This brought to bare the total future liability on old age pension benefit which the scheme is likely to experience and its nature. It was revealed based on the forecasted values that, the scheme will experience increasing values on old age pension benefit for both number and amount of payment.

6. Recommendations

In view of the findings of the study, the following recommendations are proposed;

- Since the distribution of the number of pensioners (Negative Binomial) is positively skewed, there is a probability that for some period they may experience extreme increase in the number of pensioners though the probability is less. In view of this, they should try as much as possible to increase the enrolment (number of contributors) so that their contributions can be used to offset these pensioners in case of massive increase.
- The distribution for monthly payments (Lognormal (3P)) is also positively skewed and hence there is a likelihood of them experiencing payment of extremely huge amounts to pensioners at certain periods. As a result, they should invest some of their funds in more liquid investments so that in the case of massive increase they can fall on these investments to still fulfil their mandate/ responsibility (paying Benefit to pensioners).
- Monthly payments can also increase even though the number of pensioners might not have increased due to inflation and indexation. They must be careful how and the choice of investment since for some more risky investments, in the case of inflation the investment will lose its value as well as the returns. In this case when SSNIT experiences a massive increase for a season and receives low returns, they will not be able to fulfil their obligation.

7. References

i. Adjei, E. N. A. (1999). *Pension Schemes in Africa: The National Experience of Ghana. Social Security Documentation*, African Series No. 21, International Social Security Association.
ii. Agyeman, A. (2011). *An assessment of the returns on employee pension fund investment and their impact on future benefit payments*; case study of ssnit.
iii. Ashidam, B. N. A. (2011) *The problem of Cap 30 Pension Scheme in Ghana*
iv. Bar, C & Diamond Z. (2006). *Life span issues, Growth through old Age*: Journal of Development Studies, 42(2), 19-201
v. Bar, C. & Diamond Z. (2006). *Life Span Issues, Growth through Old Age*; Journal of Development studies, 45(2), 19-201.
vi. Bauer, R. and Kicken, L. (2008), "*The pension fund advantage: are Canadians overpaying their mutual funds?*", Journal of Pension Management, Vol. 1, pp. 64-71.
vii. Beebower, G.L. and Bergstrom, G.L. (1977), "*A performance analysis of pension and profit-sharing portfolios: 1966-1975*", Financial Analysts Journal, Vol. 33, pp. 31-42.
viii. Boateng, k. (2015). *Evaluating the Performance of Pension Funds*; case study of SSNIT
ix. Bodie, Z. (1991), "*Shortfall risk and pension fund asset management*", Financial Analysts Journal, Vol. 47, pp. 57-61.
x. Cannon, E& Ian T. (2012)."*The Value and Risk of Defined Contribution Pension Schemes:* Evidence". Journal of Risk and Insurance. doi:10.1111/j.1539-6975.2011.01456.x.
xi. Darko, E. (2016). *The effectiveness of the current three- tier pension scheme in providing adequate Social security for Ghanaians*.
xii. DEI Henry (2001). Public Pension Fund Management in Ghana
xiii. Dewotor, F. (2004) "*Towards a sustainable National Pension system, modern Ghana News*", (http://www.modernghana.com/news), accessed 2010 December 16)

xiv. Harrison, M.J. and Sharpe, W.F. (1983), "*Optimal funding and asset allocation rules for defined benefit pension plans*", in Bodie, Z. and Shoven, J.B. (Eds), Financial Aspects of US Pension System, Chapter 4, University of Chicago Press, Chicago, IL.

xv. Health Care, Education and Your Retirement, Mississauga, Ontario: Wiley, p. 210, ISBN 978-1118098738

xvi. http:// en.m.wikipidia.org/wiki/Exponential_distribution.

xvii. http:// en.m.wikipidia.org/wiki/Negative Binomial_distributionwww.statisticshowto.com/lognormal-distribution/

xviii. Hurvich, & Tsai. (1989). *Regression and time series model selection in small samples.*

xix. Korakye, K.A. (2007, September 14). Retrieved from Graphic Online: http:// www.graphic.com.gh

xx. Model selection and inference. 1998. By Burnham, Kenneth P., Anderson, David R. http://www.springer.com.

xxi. Research Dept. of SSNIT (2001): *Annual Statistical Report. Social Security and National Insurance Trust, Ghana*.

xxii. The National Pensions Act (2008). Act 766.