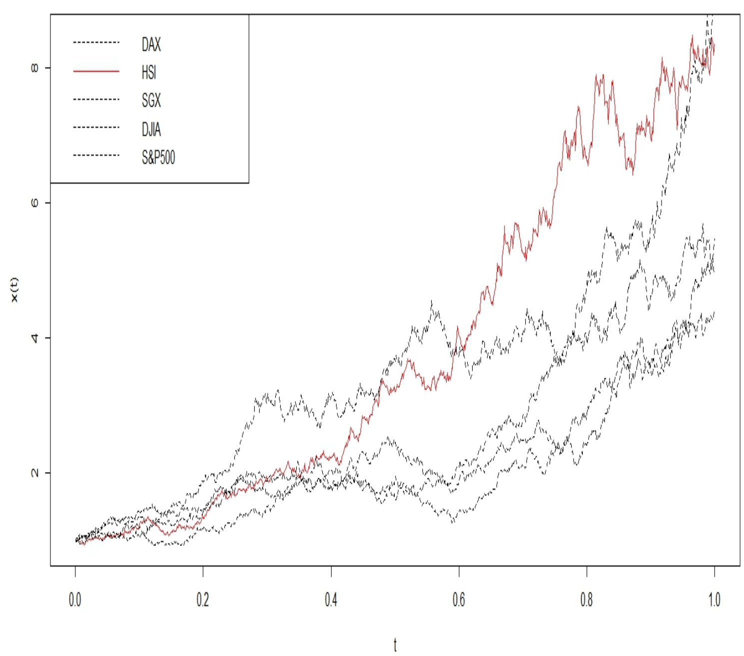
Time Series Project

A project report
Submitted in Fulfilment of the Requirement for Stat3002

of
The University of the West Indies

by
Brittanique Augustine,Tamrah Brown, Serue Edwards, Kamoya Graham, Justine Powell & Jevaun Walker
2015



Department of Mathematics
Faculty of Science and Technology
Mona Campus

1. Data observation

The Hang Seng Index (HSI) is a free-float methodology[1] index in Hong Kong. It is used to record and monitor the daily changes of the largest companies of the Honk Kong stock market. Fig.1 shows the daily stock index from March 1,1990 to November 17,1997.

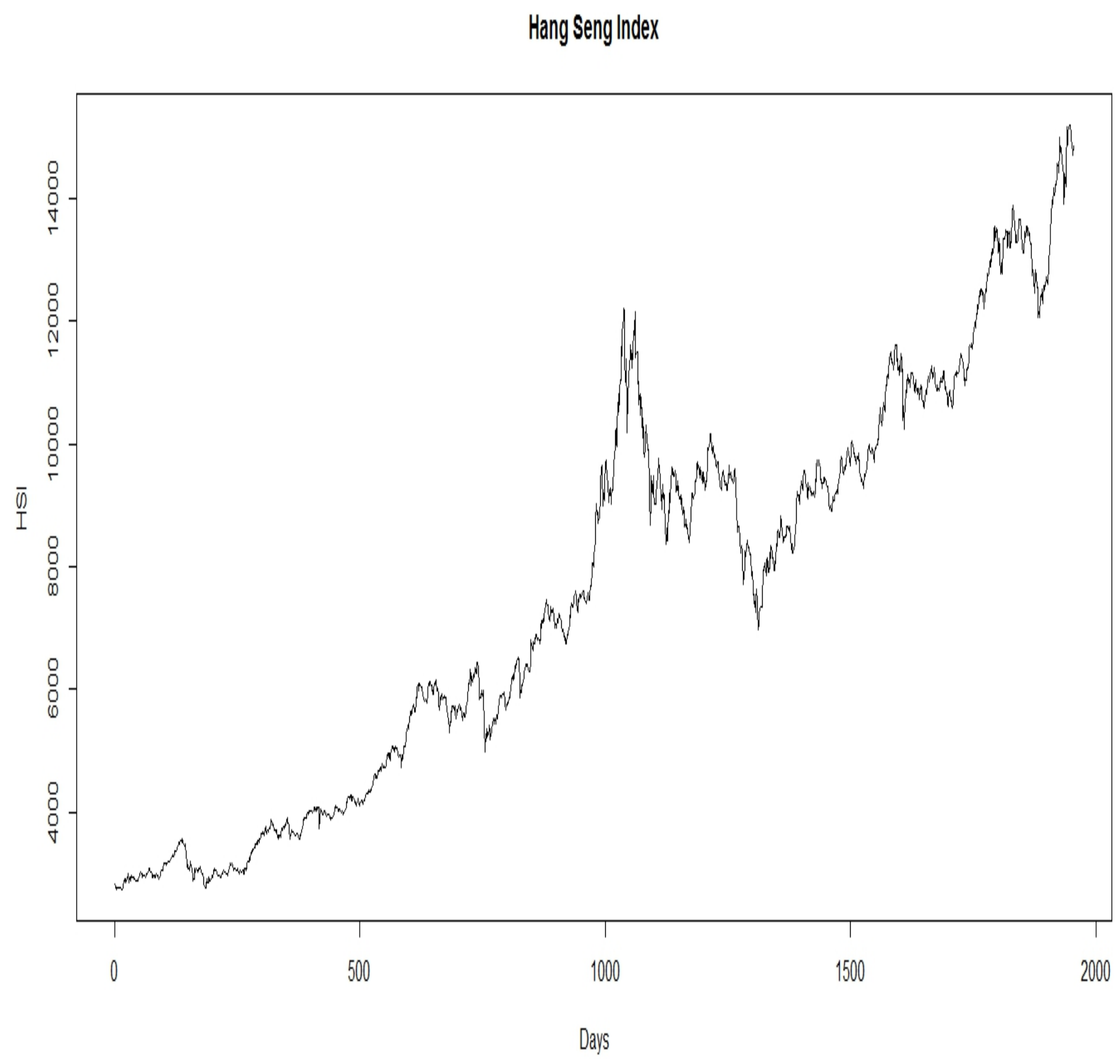


Figure 1. Hang Seng Index(HSI) in days dating from March 1,1990 to November 17,1997

[1]A method by which the market capitalization of an index's underlying companies is calculated. It is seen as a better method in contrast to full market capitalization

For analytical purposes, a time series plot of the HSI versus the years is useful. This is illustrated by Fig.2.

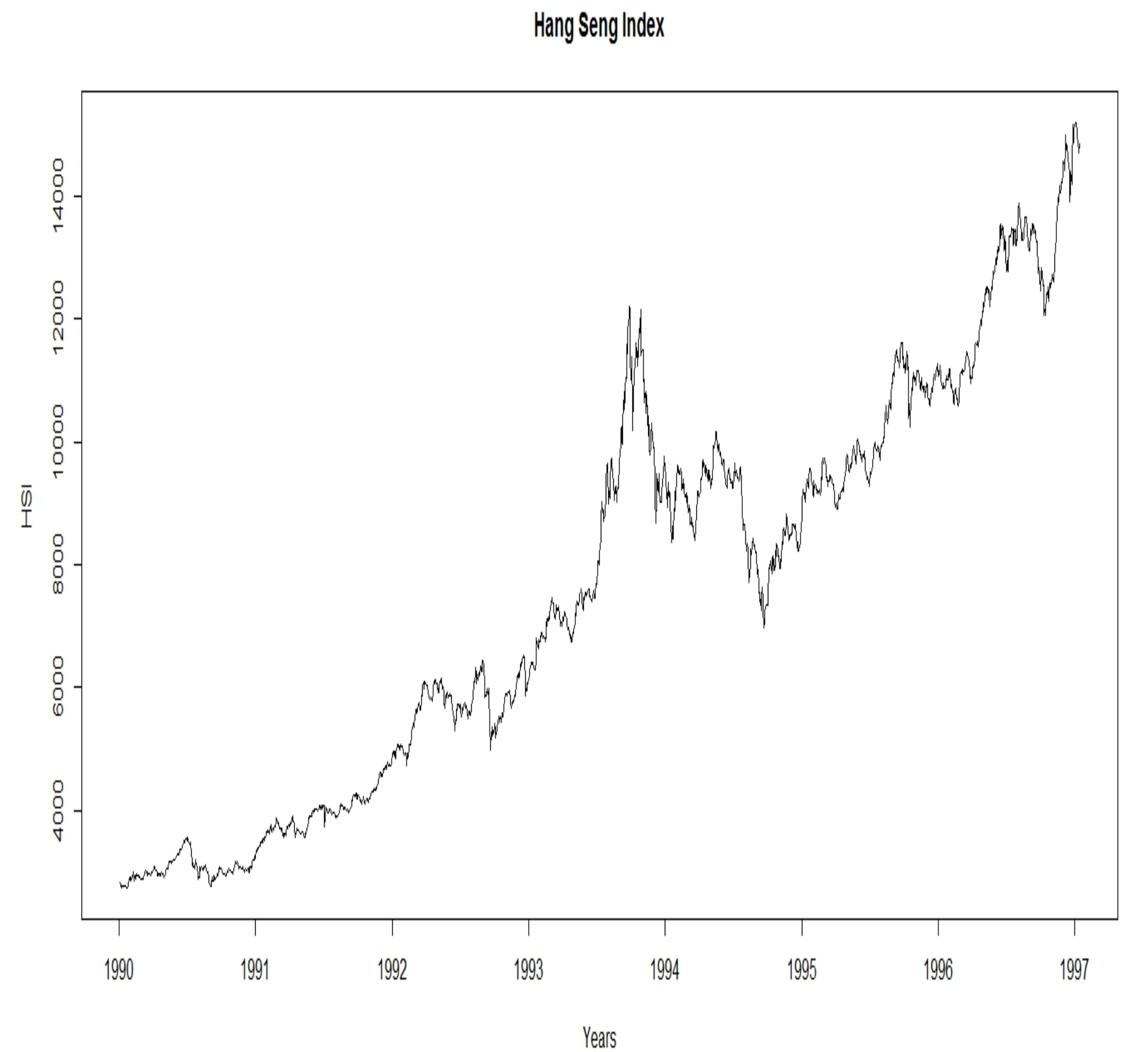


FIGURE 2. Hang Seng Index(HSI) in years dating from March 1,1990 to November 17,1997

From the plot, a general upward trend is observed. We know that if a time series contains a trend, then it is a non-stationary series. Hence, in modeling the data we would need to perform some operations to stationalize the data. It can also be observed that there was a period of steep increase followed by a decrease towards the end of 1993, in this period it can be seen that there was a record high climbing above 10,000. It is possible that this may have been the result of positive economic growth due to increased investments from investors.

2. Model building

From the preceeding section, it was concluded that the data observed was non-stationary. In order to model the data accurately, we would have to stationalize the data since the trend will affect the value of the time series at different times. Since we are working with financial data, a price to returns function was used to stationalize the data in R.

2.1. **Choosing the model.** For the stationary data, the ACF/PACF were examined to test whether an AR(p) or MA(q) model is appropriate, this can be shown in Fig3. The PACF shown in Fig.3 is suggestive of an AR(1) model. The

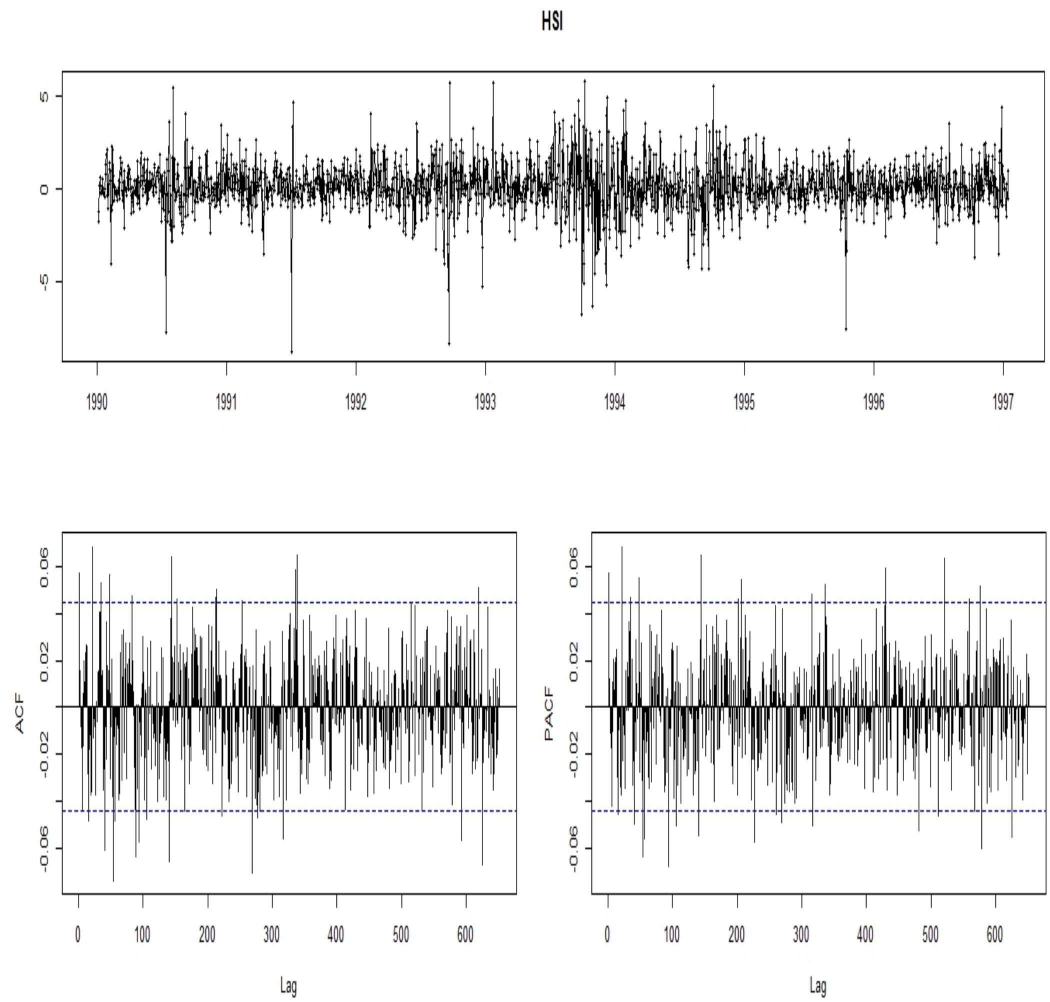


Figure 3. Time plot and ACF and PACF plots of the stationary data for HSI

Arima(p,d,q)[2] function in R was used to aid in the model selection process. Since an AR(1) model is possible based on the PACF, an initial candidate model may be Arima(1,0,0). We will fit this model along with other variations resulting in the following set of models creating a family of models.

- Model1: Arima(x,order=c(1,0,1))
- Model2: Arima(x,order=c(1,0,0))
- Model3: Arima(x,order=c(0,0,1))
- Model4: Arima(x,order=c(2,0,1))
- Model5: Arima(x,order=c(2,0,0))
- Model6: Arima(x,order=c(1,0,2))
- Model7: Arima(x,order=c(0,0,2))
- Model8: Arima(x,order=c(2,0,2))

Of these, Model2 resulted in the best fitted model having the smallest AICc. Fig4 shows the R output of Model 2.

```
Series: x
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1  intercept
      0.0569     0.0847
s.e.  0.0226     0.0318

sigma^2 estimated as 1.76:  log likelihood=-3325.17
AIC=6656.34   AICc=6656.35   BIC=6673.07
```

FIGURE 4. R output of the summary of Arima(1,0,0)

[2]for the Arima() function, the parameters p refers to the Autoregressive model, d refers to the amount of differencing required to stationalize the data and q refers to the Moving average model.

3. Model diagnostics

The ACF plot of the residuals from the ARIMA(1,0,0) model shows all correlations within the threshold limits indicating that the residuals are behaving like white noise. The Time plot also suggest that the residuals are stationary over time. A diagnostic Box-Ljung test returns a large p-value suggesting the residuals are white noise. These are shown in Fig.5 and Fig.6 respectively.

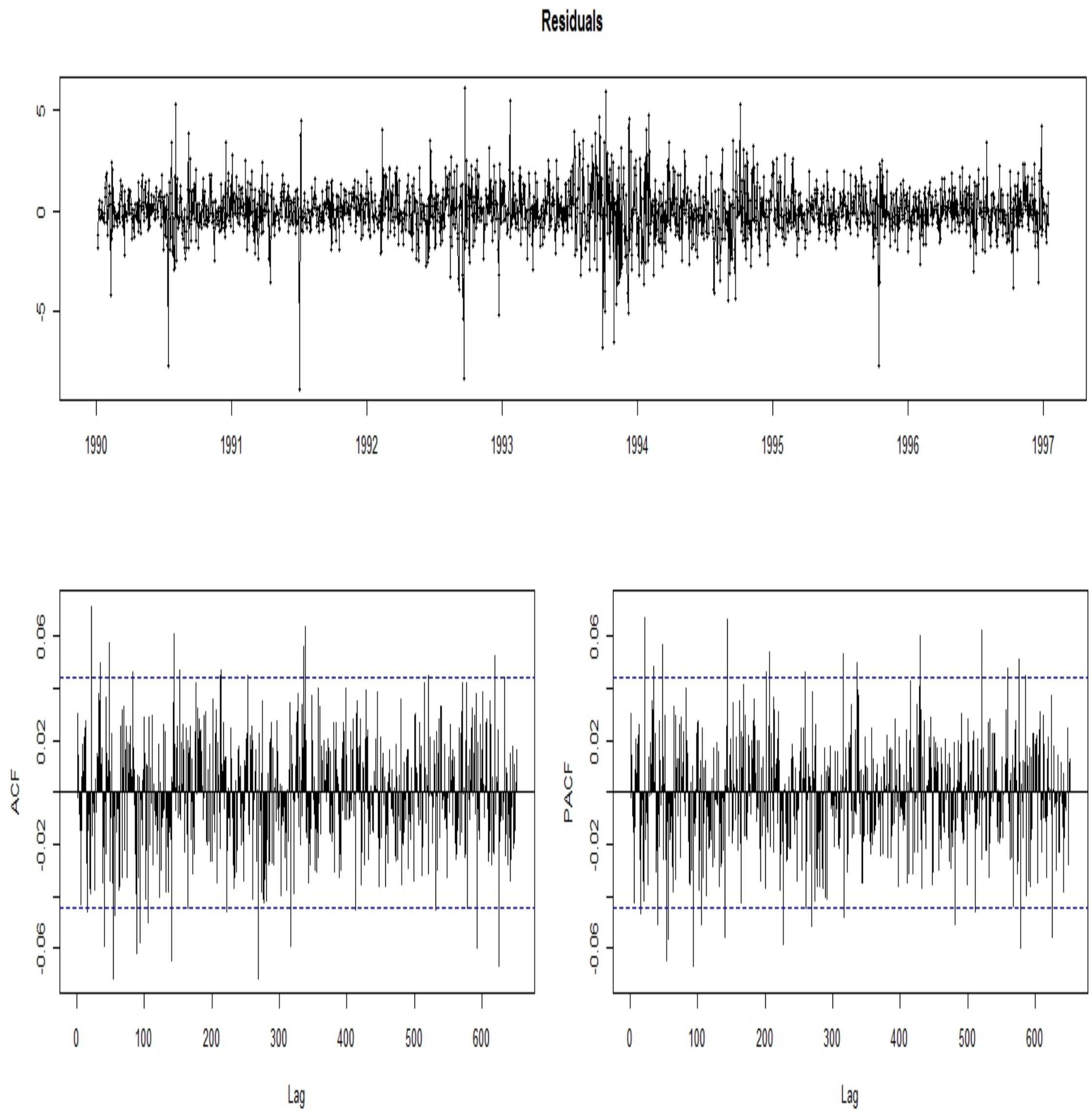
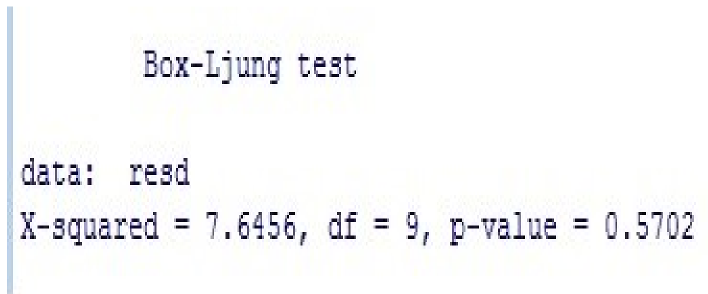


Figure 5. Time plot and Acf and Pacf plot of residuals



```
        Box-Ljung test

data:  resd
X-squared = 7.6456, df = 9, p-value = 0.5702
```

Figure 6. Box-Ljung test

These are positive results as a good forecasting method will yield residuals that are uncorrelated with zero mean. If the mean differs from zero, then the forecast will be biased. In addition, it may be useful for the residuals to be normally distributed. To test for normality of the residuals, a Q-Q plot as well as an Histogram was generated to show the results. Fig7 shows little difference between the standardised residuals and the residuals themselves. The plots suggested that normality was observed.

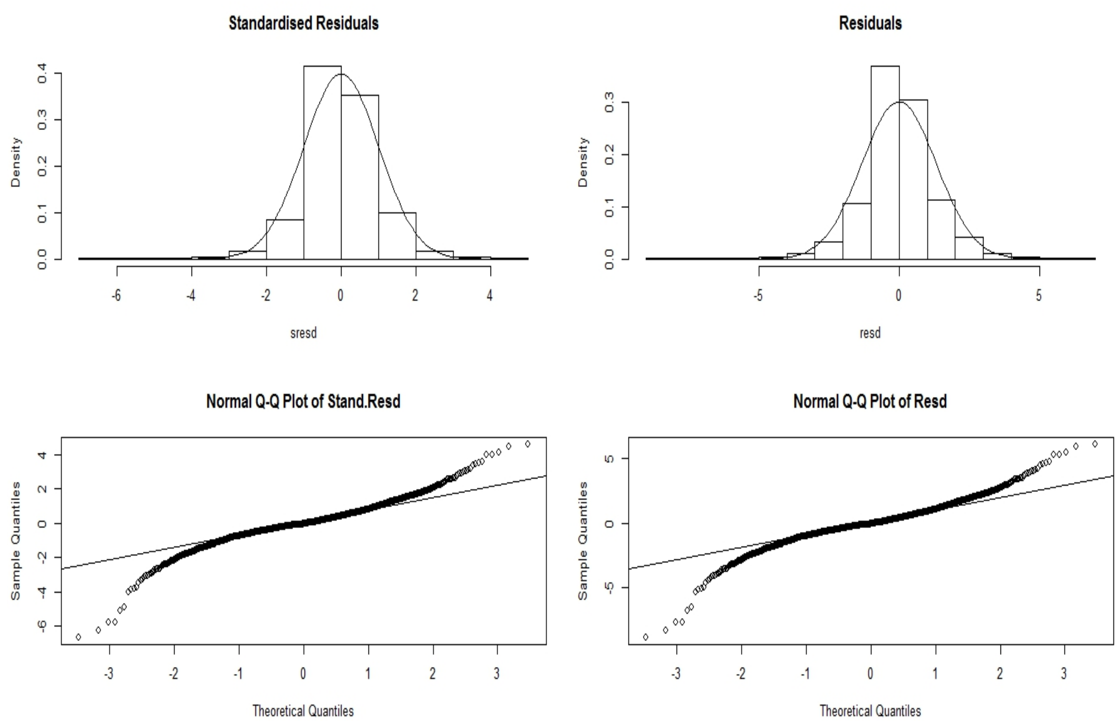


Figure 7. Normality of the Residuals and Standardised residuals

4. Mathematical equation of the chosen model

Based on Fig.4, our mathematical equation of the chosen, Arima(1,0,0) model, also could be written as AR(1), can be represented as the following:
Generally, for an AR(p) model, we have,

$$Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + Z_t \tag{1}$$

Where Y_t is the series at time t and $\mu, \alpha_1, \ldots, \alpha_p$ are constants and Z_t i.i.d with $N(0, {\sigma^2}_z)$ and not correlated to Y_t
In our case, we wish to prove that our AR(1) may be modeled as,

$$Y_t = a_0 + \alpha Y_{t-1} + Z_t, \qquad a_0 \in \mathbb{R}$$

Lemma 1. *It is known that from Fig.4,*
$\mu = 0.0847$,
$\alpha = 0.0569$,
now,

Proof. Subtracting the mean, μ, from each process in (1), when $p = 1$,we get

$$Y_t - \mu = \alpha(Y_{t-1} - \mu) + Z_t,$$
$$\Leftrightarrow Y_t = \mu(1 - \alpha) + \alpha Y_{t-1} + Z_t,$$
$$\Leftrightarrow Y_t = 0.07988 + 0.0569 Y_{t-1} + Z_t.$$

□

Therefore, our ARIMA(1,0,0) or AR(1) model, is represented as;

$$Y_t = 0.07988 + 0.0569 Y_{t-1} + Z_t$$

5. FORECAST

The 3 step ahead forecast for the stationary data was modeled in R, generating Fig.8 and Fig.9.

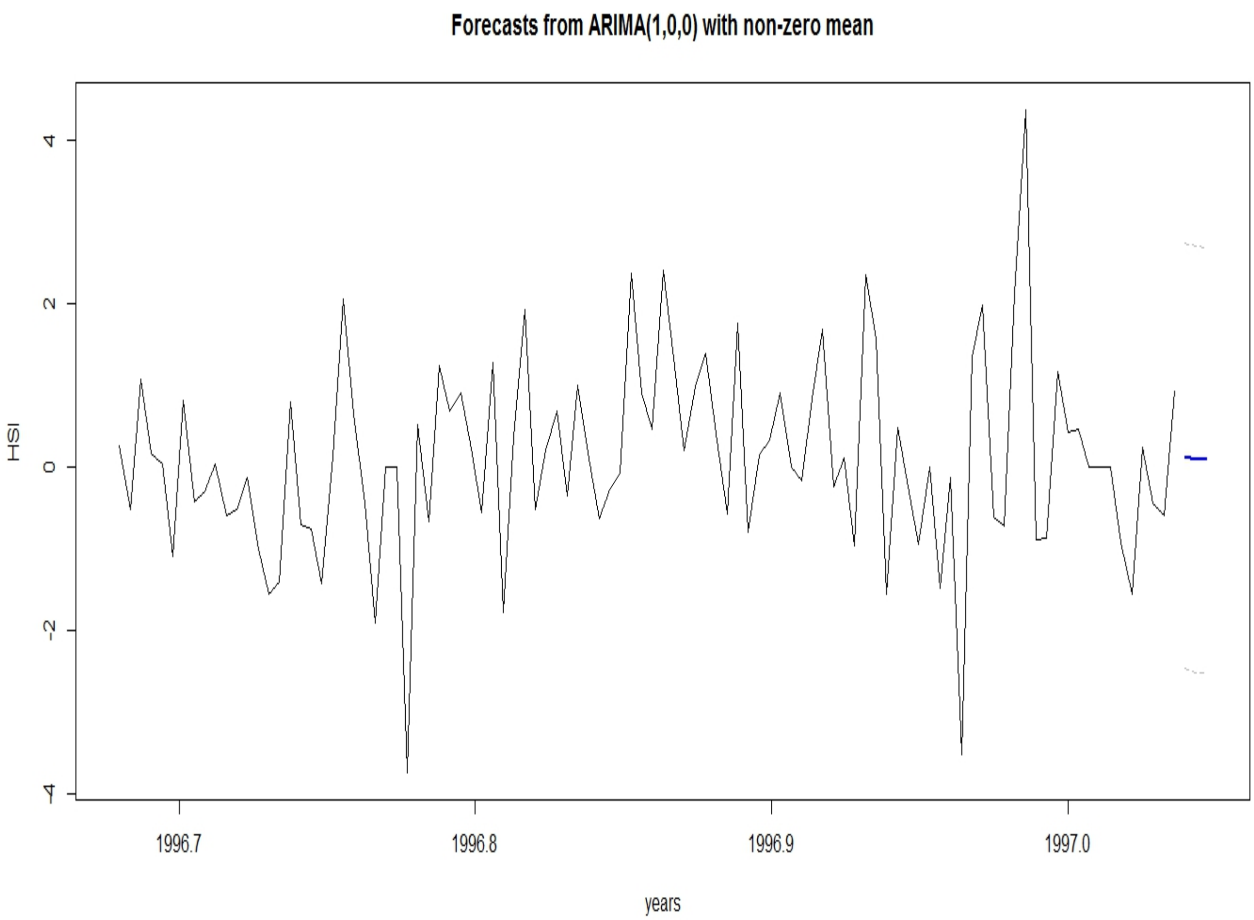


FIGURE 8. 3 step ahead forecast with predicted interval

```
Forecast method: ARIMA(1,0,0) with non-zero mean

Model Information:
Series: x
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1  intercept
      0.0569     0.0847
s.e.  0.0226     0.0318

sigma^2 estimated as 1.76:  log likelihood=-3325.17
AIC=6656.34   AICc=6656.35   BIC=6673.07

Error measures:
                       ME     RMSE       MAE MPE MAPE      MASE         ACF1
Training set 5.698238e-05 1.326819 0.9230588 NaN  Inf 0.6476457 -0.001813539

Forecasts:
         Point Forecast     Lo 95    Hi 95
1997.040     0.13202570 -2.468492 2.732544
1997.043     0.08735430 -2.517368 2.692077
1997.047     0.08481315 -2.519923 2.689549
```

Figure 9. R summary of the 3 step ahead forecast