ON EQUAL PREDICTIVE ABILITY AND PARALLELISM OF SELF-EXCITING THRESHOLD AUTOREGRESSIVE MODEL

by

Uzochukwu

Emeka Calistus

PG/MSc/12/61631

Being a project submitted to the Department of Statistics in partial fulfillment of the requirement for the award of a degree of Master of Science (MSc.) in Statistics at the University of Nigeria, Nsukka.

August 24, 2015

Certification

This is to certify that this project was carried out by Uzochukwu, Emeka Calistus, with registration number PG/MSc/12/61631 of the Department of Statistics, University of Nigeria, Nsukka.

Supervisor
Prof. F.I Ugwuowo
Department of Statistics,
University of Nigeria,
Nsukka.

Head of Department
Prof. F.C. Okafor
Department of Statistics,
University of Nigeria,
Nsukka.

External Examiner

Dedication

To God Almighty for His sustained mercy and blessings despite my persistent short-comings and to my family for their ever present support.

Acknowledgements

Special thanks goes to my supervisor, Professor F.I. Ugwuowo, who has been a pleasure to work with, and always a prolific source of ideas and support. My gratitude goes to my senior colleagues and colleagues in the Department of Statistics, University of Nigeria: Prof. F.C. Okafor, Associate Professor W.I.E Chukwu, Professor P.E. Chigbu, Dr. E.O. Ossai, E.I. Mba, A.U. Udom, V.A. Oladugba, T.E. Ugah, U.C. Nduka, E.C. Ukaegbu, M.S. Madukaife, S.S. Sanni, C.E. Ogbonnaya and J.O. Agada . I also wish to thank Dr. Y. Shehu in the Department of Mathematics, University of Nigeria and Obodo, C.S in the Department of Physics and Astronomy, University of Nigeria for their cooperation and support. Finally, I would like to thank Uzochukwu's family for their support.

Uzochukwu Emeka Calistus

Abstract

Several authors have developed statistical procedures for testing whether two models are similar. In this work, we not only present the notion of equivalence but also extend this to a measure of predictive ability of a time series following a stationary self-exciting threshold autoregressive (SETAR) process. A proposition and a lemma were used to join the structure of the predictability measure to the coefficients and sample autocorrelation of the SETAR process. Illustrative examples are given to show how to conduct the test which can help practitioners avoid mistakes in decision making.

Table of Contents

Chapter 1

INTRODUCTION

1.1 Introduction

Popularisation and extensive research for linear time series modelling began in 1927 with Yule's Autoregressive models, used in studying sunspot numbers. In the decades that followed, these models have been successfully applied in different fields, this is because as far as one-step ahead prediction is concerned, linear time series models are often adequate. However, this is not always so as can be seen from the Sunspot numbers (which will be discussed later). The causes of this are mentioned later herein.

Nonlinear time series analysis gained attention in the 1970's. The interest grew due to the need to model nonlinear changes in everyday time series data exhibiting nonlinearity. Autoregressive Integrated Moving Average (ARIMA) models cannot describe adequately limit cycles, time-irreversibility, amplitude-frequency dependency and jump phenomena (the Sunspot numbers mentioned earlier is a good example). As a result, Tong (1978) came up with a procedure for modelling nonlinear changes in time series data in which different Autoregressive (AR) processes are functioning, and the switch between these AR models depends on the delay parameter and threshold value(s), which are certain time lag values from the given time series. Tong and Lim (1980) and Tong (1983) followed up the work with an extensive description of the procedure. Tsay (1989) proposed a much simpler procedure. Tsay (1989) noted that the Tong's (1983) procedure is not statistically adequate for formally determining if a given data can be described using a threshold model (see Tsay 1989).

Several nonlinear time series models (Nonlinear Autoregressive model (AR) and Closed-loop Threshold Autoregressive model (TARSC)) have been proposed over the years and the Threshold Autoregressive (TAR) models, which is the piece-wise linearization of non-linear models over the state space by the introduction of the thresholds $\{r_o, ..., r_i\}$, has been of significant interest because of its ability to model nonlinear data adequately. Common notion were employed by Priestly (1965), and Ozaki and Tong (1975), in the analysis of non-stationary time series and time dependent systems, in which local stationarity was the counterpart of our present local linearity. The overall process is nonlinear when there are at least two regimes with different parameters and/or process order. Tong and Lim (1980) proposed the following requirements for the modelling of nonlinear time series, in order of preference:

- statistical identification of an appropriate model should not entail excessive compu tation;
- they should be general enough to capture some of the nonlinear phenomena men tioned previously;

• one-step-ahead prediction should be easily obtained from the fitted model and, if the adopted model is nonlinear, its overall prediction performance should be an improve ment upon the model;

• the fitted model should preferably reflect, to some extent the structure of the mecha nism generating the data based on theories outside statistics;

• and they should preferably possess some degree of generality and be capable of gen eralization to the multivariate case, not just in theory but also in practice.

Predictive ability in time series informs on the degree to which the past can be used in ascertaining the future. Predictive ability is fundamental in time series analysis. Assessing whether there is predictability among macroeconomic variables has always been a central issue for applied researchers. For example, much effort has been devoted to analyzing whether money has predictive content for output. This question has been addressed by using both simple linear Granger Causality (GC) tests (e.g. Stock and Watson (1989)) as well as tests that allow for non-linear predictive relationships (e.g. Amato and Swanson (2001) and Stock and Watson (1999), among others). Several authors have studied predictive ability and used it in several fields; for instance tourism, finance etc. However, not much has been done to investigate whether more than one series have equal predictive ability (Otranto and Traccia (2007)). Testing whether the models provide similar forecast performance represents a test of equal predictive ability. Testing equal predictive ability is essential in risk management; where, it could be interesting to establish if time series which have the same variables (economic, climate, etc), recorded in different spatial areas or calculated with different methodologies, have equal predictive abilities.

This work presents a test of equal predictive ability in relation to parallelsim of the Self-Exciting Threshold Autoregressive (SETAR) model. We use the Wald test used by Steece and Wood (1985) and Otranto and Triacca (2007) to investigate the similarity of SETAR processes.

1.2 Statement of Problem

Previous research works on parallelism and equal predictive ability centered on Autoregressive Integrated Moving Average models (ARIMA) and the Generalised Autoregressive Conditional Heteroscedastic (GARCH) models. Here we consider parallelism and equal predictive ability for Self-Exciting Threshold Autoregressive model. We link a measure of equal predictive ability and the structure of the model using the autocorrelation and coefficients of the model. It will be necessary to also consider whether transformations are parallel to the original data this is because in building time series models, Box and Jenkins (1970) have devised an iterative strategy of model identification, estimation and diagnostic checking. The identification stage of their model building cycle relies on the recognition of typical patterns of behaviour or structure in the sample autocorrelation function and the partial autocorrelation. We investigate these in this work.

1.3 Research Objectives

This work deals with the predictive ability in time series exhibiting nonlinearity. The study aims to achieve the following objectives:

1. to apply a test of equal predictive ability to suit nonlinear time series,
2. to establish the condition necessary for parallelism and equal predictive ability of a nonlinear time series,
3. to validate the test with real life data.

1.4 Significance of the Study

When testing for equal predictive ability, the question that is of interest is whether one forecast model is better than another. This question can be addressed by testing the null hypothesis that the two series have the same structure. This testing problem is important for applied analysts, because several ideas and specifications are often used before a model is selected. This test can be narrowed down to testing if the different series are parallel which is a way of checking similarities in the structure of different series. Instead of testing for predictive equality we can test for similarity in the structure of the series (parallelism). There are several instances where it is important to check if two or more time series are equivalent. For instance, the task of predicting the demand for common items in different markets may be possible if it can be shown that the models characterizing demand are equivalent in various markets. If the hypothesis of parallelism between two time series is accepted, one can obtain better estimates of the model parameters by pooling the data sets, also by using series with more similar structure one can forecast the volatility of one series from the other(s) and it can be used to choose among several procedures of seasonal adjustment.

1.5 Scope of the Study

We consider Self-Exciting Threshold Autoregressive models in relation to parallelism and equal predictive ability. Since the R^2 index can be used to test for the predictive ability we show that it can be expressed as a function of the parameters of the time series model and autocorrelation of the given time series. These helps in describing the structure of the series. We use this index to test equal predictive ability and parallelism between different models. We test the hypothesis by considering a test proposed by Steece and Wood (1985) where they presented a simple method for assessing the equivalence of k time series, we then relate this to the predictive ability of different time series.

Chapter 2

LITERATURE REVIEW

2.1 Review of Literatures

Given the obvious desirability of a formal statistical procedure for forecast comparison, different authors (Dhrymes et al. (1972), Howrey et al. (1974), Vuong (1989)) have proposed different test procedures but the widely applicable test for comparing two forecast was first proposed by Deibold and Mariano (1995) (DM). Their approach is capable of assessing the significance of divergences between models and data. The test can handle forecast errors that are non-Gaussian, nonzero mean and serially correlated. They considered the following: Given two forecasts $\{\hat{Z}_{it}\}_{t=1}^{T}$ and $\{\hat{Z}_{jt}\}_{t=1}^{T}$ of a time series $\{Z_t\}_{t=1}^{T}$, with respective error forecasts as $\{\epsilon_{it}\}_{t=1}^{T}$ and $\{\epsilon_{it}\}_{t=1}^{T}$, where $g(Z_t, \hat{Z}_{it}) = g(\epsilon_{it})$ and $g(Z_t, \hat{Z}_{jt}) = g(\epsilon_{jt})$ are arbitrary loss functions e.g mean square prediction error (MSPE) given by

$$g(\epsilon_t) = T^{-1/2} \sum_{t=1}^{T} \hat{\epsilon}_t^2$$

The null hypothesis of equal forecast accuracy for two forecasts is $E[g(\epsilon_{it})] = E[g(\epsilon_{jt})]$ or $E[d_t] = 0$, where $d_t \equiv [g(\epsilon_{it}) - g(\epsilon_{jt})]$ is the loss differential. $\sqrt{T}(\overline{d} - \mu)$ converges in distribution to $N(0, 2\pi f_d(0))$, where $\overline{d} = \frac{1}{T}\sum_{t=1}^{T} |g(\epsilon_{it}) - g(\epsilon_{jt})|$ is the sample loss differential, $f_d(0) = \frac{1}{2\pi}\sum_{t=-\infty}^{\infty} \gamma_d(\tau)$ is the spectral density of the loss differential at frequency 0, $\gamma_d(\tau) = E[(d_t - \mu)(d_{t-\tau} - \mu)]$ is the autocovariance of the loss differential. The large sample test statistics is given by,

$$S_1 = \frac{\overline{d}}{\sqrt{\frac{2\pi \widehat{f_d(0)}}{T}}} \quad \sim \quad N(0, 2\pi f_d(0)) \tag{2.1}$$

Deibold and Mariano (1995) show that the procedure works very well for both large and small samples (i.e for n>15). Deibold and Mariano (1995), however, did not discuss procedures for performing inferences about predictions when predictions rely on model estimates.

West (1996), asserted that economic predictions rely on estimates from the model. Model based predictions have been used by authors, one of the earliest being Nelson (1976). Interest is in $(l \times 1)$ vectors of moments Ef_t, where f_t depends on $(k \times 1)$ unknown parameter vector β^* and t observations. If $\widehat{\beta}_t$ represents the estimate of β^*, then the estimate of Ef_t is given by, $\overline{f} = \frac{\sum_{t=R}^{T} f_{t+\tau}(\widehat{\beta}_t)}{P}$, where $f_{t+\tau}(\widehat{\beta}_t)$ can be Mean Square prediction Error (MSPE) or Mean Prediction Error (MPE), P the number of predictions, M is the sample size used in the first prediction, $M + 1$ is the sample size used in the second prediction and $M + P - 1$ is the sample size used in the last prediction. In this case, $\sqrt{P}(\overline{f} - Ef_t) \sim N(0, \Omega)$, where $\Omega = \gamma_j = E(f_t - Ef_t)(f_{t-j} - Ef_t)$. In which case the $\Omega^{-2}(\overline{f} - Ef_t)^2 \sim \chi_1^2$ and the hypothesis to be tested is the $Eu_{1t}^2 = Eu_{2t}^2$, where

$u_t^2 = (\overline{f} - Ef_t)^2$. Using the procedure, the vector of sample prediction errors was used to test zero mean, variance equal to the underlying population variance and zero serial correlation. They showed that the standard Deibold and Mariano (DM) test can be used in the cases where either the loss function is quadratic or the number of the prediction, P, is less than the sample size, M. West (1996) did not consider the case where there could be errors in the parameters estimated, (i.e in the absence of parameter uncertainty).

Corradi et al (2001), addressed the conditions under which the Deibold and Mariano (1995) test for comparative predictive ability can be extended to the case of two forecasting models, each of which may include cointegrating relation, when allowing for parameter errors. This is done by examining the effect of parameter error in cointegrating framework, using the methods proposed by West (1996). Monte Carlo experiments were conducted in other to examine the impact of cointegrating vector parameter estimation error on Deibold and Mariano (DM) test. Corradi et al (2001) suggest that the test performs poorly when parameters are unknown. The null hypothesis of equal predictive ability they tested is given by;

$$H_0: \; E(f(\hat{\upsilon}_{0,t+h}) - f(\hat{\upsilon}_{k,t+h})) \; = \; 0 \tag{2.2}$$

$\hat{\upsilon}_{i,t+h} = f(\upsilon_{i,t+h}, \upsilon_{i,t+h})$, $i = 0, 1, ..., l$ represents h step ahead the estimated forecasting errors from models and f is an arbitrary the loss function which can be

$$MPE : f = P^{-1/2} \sum_{t=1}^{P} \hat{\upsilon}_{t+h}$$

They noted that

$$P^{-1/2}(\sum_{t=R}^{T-1}(f(\hat{\upsilon}_{k,t+h}) - E(f(\hat{\upsilon}_{k,t+h}))))$$

does not converge to normal distribution but the result from West (1996) gives an asymptotic normality result when the process contains a trend component. The literature contains other procedures for testing equal predictive ability. While some authors addressed the issue of equal predictive ability using the forecasts errors, others addressed the same issue in the framework of parallelism, (by considering similarity in the structure of the process).

Steece and Wood (1985) introduced a simple method for assessing the similarity of k time series (exhibit parallelism and concurrence) when the series are not necessarily independent. There are practical situations in which it would be of value to know whether more than one time series are equivalent. For instance, the task of forecasting the demand for common items in different markets may be facilitated if it can be shown that the models characterising demand are equivalent in the various markets. An extension of this application on a broader scale is that equivalence of economic series in separate geographic areas could suggest that the economic conditions are same in the studied regions, therefore simplifying comparative analysis, Steece and Wood (1985). The hypothesis for testing parallelism or equivalence can be expressed as $H_0 : \underline{A\beta} = 0$ where $\underline{A}$ is a known $(q \times m)$ matrix and $\underline{\beta}$ is an m-dimensional vector of parameters. The Wald test statistic for the null hypothesis is expressed as

$$W = (\underline{A\widehat{\beta}})'(\underline{A\widehat{\Lambda}A'})^{-1}(\underline{A\widehat{\beta}}) \tag{2.3}$$

$\underline{\widehat{\beta}}$ are the maximum likelihood estimator of $\underline{\beta}$ and $\underline{\widehat{\Lambda}}$ is the maximum likelihood estimator of the covariance matrix of $\underline{\widehat{\beta}}$. W is asymptotically distributed as a chi-square random variable with q degrees of freedom. There are other methods for testing parallelism.

Triacca (2004) analyzed the relationship between a measure of dissimilarity (previously used for choosing a representative element from a large collection of time series and for clustering time series data proposed by Piccolo (1990)) between ARIMA models and a notion of parallelism of two ARIMA processes, noting that if two time series could be considered as coming from a common ARMA model (i.e if the hypothesis of parallelism is accepted), one can obtain better estimates of the model parameters by pooling the data sets. The measure of dissimilarity is given by

$$v(X_t, Y_t) = [\sum_{i=1}^{\infty}(\pi_{xi} - \pi_{yi})^2]^{1/2} \tag{2.4}$$

which is a distance measure, where X_t and Y_t are different processes, $\pi's$ are constants obtained from $\phi(z)/\theta(z)$. The two ARIMA processes are parallel if and only if the distance between them is null. To test the null hypothesis $H_0 : \phi_x(z) = \phi_y(z)$, $\theta_x(z) = \theta_y(z)$ they used the following test statistic

$$\hat{v}^2 = \sum_{i=1}^{\infty}(\hat{\pi_{xi}} - \hat{\pi_{yi}})^2 \tag{2.5}$$

which follows chi-square with p degrees of freedom. The finite sample properties of the test was analyzed, and suggests that a sample size of 50 is not long enough for the test to display its asymptotic behaviour but a sample size of 100 has a good performance. The notion of parallelism is useful not only in ARIMA models but also in other models as well.

Financial time series are in general subject to similar volatility structures, due to the strong influence among financial markets. Otranto (2004) then extended well known results developed for the ARIMA models to GARCH models by considering financial time series characterized by similar volatility structures. The argument being that the selection of series having similar behaviour could be important for the analysis of the transmission mechanisms of volatility and to forecast the time series. Using the series with more similar structure movements in a given time series could help to forecast the movements of a similar time series. In practice, the basic idea is that the estimation of GARCH models provides the statistical structure of the financial time series. So that the comparison of the models underlying the data generating process is equivalent to comparing the volatility structures of each series. To achieve this the ARMA representation of the GARCH process was used and interest being the GARCH(1,1) model because it is the most adopted for financial time series. While investigating the distance measure and parallelism on GARCH processes, they noted that the distance (and hence the hypothesis of parallelism) increased when the difference in model parameters estimates increase. They, however, did not consider its relationship to equal predictive ability which was addressed in the following work, which was considered by Otranto and Triacca (2007).

Otranto and Triacca (2007) used a measure of predictive ability in the framework of Granger and Newbold (1976) where they proposed a natural measure of one-step ahead predictability for time series based on R^2 and considered a h-step ahead predictability measure for Autoregressive Moving Average (ARMA) processes. They developed necessary and sufficient conditions of equal predictability in terms of the notion of parallelism proposed by Steece and Wood (1985) for the case of two or more independent time series by working in a univariate frame. This result allows the use of simple tests to verify the condition of equal predictability. The null hypothesis can be expressed in terms of ψ-weights which represents the dependence structure of the series.

$H_0 : \psi^2_{1,k} = \psi^2_{2,k} = \cdots = \psi^2_{n,k} \quad \forall k$. The wald test statistic is expressed as;

$$W = (A\hat{\psi}^2)'(AG\hat{\Lambda}G'A')^{-1}(A\hat{\psi}^2) \tag{2.6}$$

$\hat{\psi}^2$ is the maximum likelihood estimator of ψ^2, and $G\hat{\Lambda}G'$ is the covariance matrix, where G is the $[nm \times np]$ matrix whose kth row is the vector of derivatives of the kth function in ψ^2 with respect to the vector of parameters of the model, and $\underline{\widehat{\Lambda}}$ is the maximum likelihood estimator of the covariance matrix of the vector of parameters of the model. W is asymptotically distributed as a chi-square random variable with $m(n-1)$ degrees of freedom. The conditions allows help them to choose among several procedures of seasonal adjustment. However, to the best of my knowledge, no further work has been done to check predictive ability and parallelism.

This work is centered on equal predictive ability and its relationship with parallelism among Self-Exciting Threshold Autoregressive SETAR models.

Chapter 3

METHODOLOGY

The objective of this chapter is to introduce the tool which will form the basis for the test we wish to conduct. We work in a univariate frame bearing in mind that our model of interest is the SETAR model. We thus proceed with our method.

3.1 Method

The R^2 statistic which measures the proportion of the total variation in the data that is explained by the time series model can be stated in the context of the self-exciting threshold autoregressive (SETAR) time series models as one minus the ratio of the residual variance to the total variance of the time series. R^2 makes available a measure of the relative predictive ability of a time series given its past history. The understanding of R^2 in the time series context gains much from the similarity to linear regression, but the special character of time-series models offers the opportunity of additional insight which can be of considerable value in actual data analysis. R^2 can be related to the underlying parameters of a time series and to its autocorrelation structure, which shall be presented shortly.

We proceed with some basic definitions.

3.2 Definition of Basic Concepts

In what follows, we introduce the concepts and definitions that are relevant to the work.

Self-Exciting Threshold Autoregressive (SETAR) model: A time series Z_t is a Self-Exciting Threshold Autoregressive process if it is of the form:

$$Z_t = \phi_o^j + \sum_{i=1}^{p} \phi_i^j Z_{t-i} + \epsilon_t^j \ \ if \ \ r_{j-1} \leq \ Z_{t-d} < r_j \tag{3.1}$$

where $j = 1, ..., k$; k is the number of regimes separated by $(k-1)$ threshold values and $d \in \ N^+$ is the delay parameter ($d \leq \ p$). The threshold values are -$\infty < r_o < r_1 < ... < r_j < \infty$ for each $j = 1, ..., k$; $\{\epsilon_t^j\}$ is a sequence of independent normal variables with zero mean and variance $\sigma_{\epsilon_j}^2$; $\{\phi_o^j, \phi_i^j\}$, $j = 1, ..., k : 1 = 1, ..., p$ are the model parameters for regime j. This process divides its self into k regimes with each regime being an autoregressive process. The process becomes nonlinear if there exist at least two different autoregressive models, each having different coefficients and possibly different orders. This model is capable of explaining limits cycles, amplitudes and jumps which cannot be explained by linear models, Tsay (1989).

Limit Cycle: A limit cycle is a state or behaviour which a dynamic system tends to evolve, represented as a point or orbit in the system phase space. In dynamic system, a limit cycle consists of two or more points between which periodic motions occurs. It is the stability boundary for linear and non-linear systems, Tong and Lim (1980).

Amplitude-frequency dependence: This is the dependence of the amplitude of a sinusoidal oscillation on its frequency at the output. The amplitude-frequency is often called simply the frequency. For ease of visualisation it is plotted in the form of a curve; the amplitude along the vertical axis and frequency along the horizontal axis, Tong and Lim (1980).

Jump Resonance: Jump resonance is a phenomenon where smooth changes in the output of a nonlinear filter leads to abrupt changes in the output. This term is used in the case of a sudden jump of amplitude and/or frequency of a periodic output of a nonlinear system, Tong and Lim (1980).

Time Irreversibility: Time irreversibility is an attribute of some stochastic and deterministic processes. If a stochastic process is time irreversible, then it is possible to determine, given the states at a number of points in time after running the stochastic process, which state came first and which state arrived later. A process Z_t is said to be time reversible if the finite dimensional distribution functions $Z_{t_1}, Z_{t_2}, ..., Z_{t_n}$ and $Z_{t_n}, Z_{t_{n-1}}, ..., Z_{t_1}$ are the same for any n, otherwise it is not time reversible, Tong and Lim (1980).

Predictive Ability: Predictability is the degree to which a correct forecast of a system can be made either qualitatively or quantitatively. This measures the accuracy with which the system can define the location of the next state of the system. It indicates to what extent the past can be used to determine the future Otranto and Traccia

(2007).

Parallelism: This is the independent emergence of a similar pattern in different but closely related data.

3.3 R^2 Defined for SETAR Time Series Models

We are interested in time series $\{Z_t\}$ which have a self-exciting autoregressive (SETAR) representation of the form

$$Z_t = \phi_o^j + \sum_{i=1}^{p} \phi_i^j Z_{t-i} + \epsilon_t^j \qquad if\;\; r_{j-1} \leq \; Z_{t-d} < r_j$$

where $\phi_i^j \;\; i = 1, ..., p \;\; and \;\; j = 1, ..k$ are parameters for the different regimes satisfying conditions for stationarity and ϵ_i^j are sequences of uncorrelated random variables with common variance $\sigma_\epsilon^{2(j)}$. The fraction of the variation in Z, $V(Z)$ which cannot be predicted is therefore

$$1 - R^2 = \frac{\sigma_\epsilon^{2(j)}}{V(Z)^{(j)}}$$

and the fraction that can be predicted is defined as

$$R^2 = 1 - \frac{\sigma_\epsilon^{2(j)}}{V(Z)^{(j)}}, \tag{3.2}$$

For a series which cannot be forecasted, such as white noise, this ratio is equal to zero. If the series can be forecasted without error, the ratio is 1. The interpretation of R^2 gives extra meaning when it can be associated with the parameters of the time series and to its autocorrelation function. This tool can be very useful at the identification stage when it is the desire of an analyst to assess the predictive ability of a time series. Theoretically, R^2 can be associated with the coefficients and autocorrelation coefficients in the following way; denoting the autocovariances of the process by γ_i^j, the variance of Z_t is given by

$$V(Z_t)^{(j)} = E(Z_t^2)^{(j)} = \phi_1^j \gamma_1^j + ... + \phi_p^j \gamma_p^j + \sigma_\epsilon^{2(j)}, \tag{3.3}$$

, we then have $R^{2(j)}$ to be

$$\begin{aligned} 1 - R_p^{2(j)} &= \frac{\sigma_\epsilon^{2(j)}}{V(Z_t)^{(j)}} \\ &= 1 - \phi_1^j \rho_1^j - .... - \phi_p^j \rho_p^j \\ R_p^{2(j)} &= \phi_1^j \rho_1^j + ... + \phi_p^j \rho_p^j \end{aligned}$$

$$R_p^{2(j)} = \underline{\rho}\prime_p^{(j)} \underline{\phi}_p^{(j)} \tag{3.4}$$

$$\underline{\phi}_p^{(j)} = R_p^{2(j)} \left(\underline{\rho}\prime_p^{(j)}\right)^{-1} \tag{3.5}$$

where $\underline{\rho}\prime_p^{(j)}$ denotes the column vector $(\rho_1^j, ..., \rho_p^j)$ and $\underline{\phi}_p^{(j)}$ is $(\phi_1^j, ..., \phi_p^j)$.
Yule-Walker equations is given by,

$$\underline{\rho}_p^{(j)} = \Gamma_p^{(j)} \underline{\phi}_p^{(j)} \tag{3.6}$$

where $\Gamma_p^{(j)}$ is the correlation matrix for (Z_{t-1}, Z_{t-p}). Using equation (3.5)

$$\underline{\phi}_p^{(j)} \ = \ \Gamma^{(j)-1} \underline{\rho}_p^{(j)}$$

we can then obtain $R_p^{2(j)}$ as

$$R_p^{2(j)} = \underline{\rho}\prime_p^{(j)} \Gamma^{(j)-1} \underline{\rho}_p^{(j)} \tag{3.7}$$

Proposition 3.1. *Let* $\{Y_t; t = 0, \pm 1, ...\}$ *and* $\{Z_t; t = 0, \pm 1, ...\}$ *be two SETAR processes. The processes* $\{Y_t; t = 0, \pm 1, ...\}$ *and* $\{Z_t; t = 0, \pm 1, ...\}$ *are equally predictable if*

$$\underline{\rho}\prime_p^{(j)} \underline{\phi}_{p_Y}^{(j)} \ - \ \underline{\rho}\prime_q^{(j)} \underline{\phi}_{q_Z}^{(j)} = 0$$

Proof. If $\{Y_t; t = 0, \pm 1, ...\}$ and $\{Z_t; t = 0, \pm 1, ...\}$ have equal predictive ability, then

$$\begin{aligned}
R_{p_Y}^{2(j)} \ &= \ R_{q_Z}^{2(j)} \\
\underline{\rho}\prime_p^{(j)} \underline{\phi}_{p_Y}^{(j)} \ &= \ \underline{\rho}\prime_q^{(j)} \underline{\phi}_{q_Z}^{(j)} = 0 \\
\rho_1^j \phi_{1Y}^j \ + \ \rho_2^j \phi_{2Y}^j \ + \ \cdots \ + \ \rho_p^j \phi_{pY}^j \ &= \ \rho_1^j \phi_{1Z}^j \ + \ \rho_2^j \phi_{2Z}^j \ + \ \cdots \ + \ \rho_q^j \phi_{qZ}^j \\
\rho_1^j \phi_{1Y}^j \ + \ \rho_2^j \phi_{2Y}^j \ + \ \cdots \ + \ \rho_p^j \phi_{pY}^j \ &- \ \rho_1^j \phi_{1Z}^j \ - \ \rho_2^j \phi_{2Z}^j \ - \ \cdots \ - \ \rho_q^j \phi_{qZ}^j
\end{aligned}$$

The measures are the same if

$$\rho_i^j \phi_{i_Y}^j \ - \ \rho_i^j \phi_{i_Z}^j \ = \ 0, \quad i = 1, ..., \ p, q \ \ for \ \ each \ \ i$$

Hence,

$$\rho_1^j \phi_{1_Y}^j \ + \ \rho_2^j \phi_{2_Y}^j \ + \ \cdots \ + \ \rho_{p_y}^j \phi_{p_Y}^j \ - \ \rho_1^j \phi_{1_Z}^j \ - \ \rho_2^j \phi_{2_Z}^j \ - \ \cdots \ - \ \rho_{q_z}^j \phi_{q_Z}^j \ = \ 0$$
$$\rho\prime_p^{(j)} \phi_{p_Y}^j \ \ - \ \ \rho\prime_{q_Z}^{(j)} \phi_{q_Z}^j = 0$$

$\square$

3.3.1 Parallelism and Equal Predictive Ability

Let $\{Y_t; t = 0, \pm 1, ...\}$ and $\{Z_t; t = 0, \pm 1, ...\}$ be two SETAR processes

$$\begin{aligned} \phi_Y^j(B) Y_t &= \epsilon_t^j, \quad \epsilon_{Y,t}^j \sim WN(0, \sigma_Y^{j2}) \\ \phi_Z^j(B) Z_t &= \epsilon_t^j, \quad \epsilon_{Z,t}^j \sim WN(0, \sigma_Z^{j2}) \end{aligned}$$

where $\phi_Y^j(L)$ and $\phi_Z^j(L)$ are finte polynomials in L of degrees p_Y and p_Z respectively; If $\{Y_t; t = 0, \pm 1, ...\}$ and $\{Z_t; t = 0, \pm 1, ...\}$ are parallel then $\phi_Y^j(L)$ = $\phi_Z^j(L)$. This definition is given in Steece and Wood (1985), in Guo (1999) and in Otranto and Triacca (2007) for the case of AR models.

Lemma 3.1

Let $\{Y_t; t = 0, \pm 1, ...\}$ *and* $\{Z_t; t = 0, \pm 1, ...\}$ *be two SETAR processes. If* $\{Y_t; t = 0, \pm 1, ...\}$ *and* $\{Z_t; t = 0, \pm 1, ...\}$ *are parallel, then they have equal predictive ability.*

Proof. If $\{Y_t; t = 0, \pm 1, ...\}$ and $\{Z_t; t = 0, \pm 1, ...\}$ are parallel, then

$$Y_t \phi_Y^j(L) \ = \epsilon_t^j \ = \epsilon_t^j \ = \ \phi_Z^j(L) Z_t$$
$$Y_t \ = \ \frac{\epsilon_t^j}{\phi_Y^j(L)} = \frac{\epsilon_t^j}{\phi_Z^j(L)} = Z_t$$

with $\phi_Y^j(L) \ \neq 0$ and $\phi_Z^j(L) \ \neq \ 0$ for $\mid L \mid < 1$

$$\frac{\epsilon_t^j}{\phi_Y^j(L)} \ = \ (1 \ + \ \psi_{1y}^j L \ + \ \psi_{2y}^j L^2 \ + \cdots) \epsilon_t^j$$
$$= (1 \ + \ \psi_{1z}^j L \ + \ \psi_{2z}^j L^2 \ + \cdots) \epsilon_t^j \ = \frac{\epsilon_t^j}{\phi_Z^j(L)}$$
$$\phi_Y^j(L)(1 \ + \ \psi_{1y}^j L \ + \ \psi_{2y}^j L^2 \ + \cdots) \ = \ \phi_Z^j(L)(1 \ + \ \psi_{1z}^j L \ + \ \psi_{2z}^j L^2 \ + \cdots)$$

$$(1 - \phi_{1y}^j L - \phi_{2y}^j L^2 - \cdots)(1 + \psi_{1y}^j L + \psi_{2y}^j L^2 + \cdots) = (1 - \phi_{1z}^j L - \phi_{2z}^j L^2 - \cdots)(1 + \psi_{1z}^j L + \psi_{2z}^j L^2 + \cdots) \tag{3.8}$$

From the LHS we obtain

$$1 + \psi_{1y}^j L + \psi_{2y}^j L^2 + \cdots - \phi_{1y}^j L - \psi_{1y}^j \phi_{1y}^j L^2 - \psi_{2y}^j \phi_{1y}^j L^3 - \cdots - \phi_{2y}^j L - \psi_{1y}^j \phi_{2y}^j L^3 - \cdots = 1 \tag{3.9}$$

and the RHS

$$1 + \psi_{1z}^j L + \psi_{2z}^j L^2 + \cdots - \phi_{1z}^j L - \psi_{1z}^j \phi_{1z}^j L^2 - \psi_{2z}^j \phi_{1z}^j L^3 - \cdots - \phi_{2z}^j L - \psi_{1z}^j \phi_{2z}^j L^3 - \cdots = 1 \tag{3.10}$$

$$L^1 : \psi^j_{1y} - \phi^j_{1y} = 0$$

$$\Longrightarrow \phi^j_{1y} = \psi^j_{1y}$$

$$L^2 : \psi^j_{2y} - \psi^j_{1y}\phi^j_{1y} - \phi^j_{2y} = 0$$

$$\Longrightarrow \phi^j_{2y} = \psi^j_{2y} - \psi^j_{1y}\phi^j_{1y}$$

$$L^3 : \psi^j_{3y} - \psi^j_{2y}\phi^j_{1y} - \psi^j_{1y}\phi^j_{2y} - \phi^j_{3y} = 0$$

$$\Longrightarrow \phi^j_{3y} = \psi^j_{3y} - \psi^j_{2y}\phi_{1y} - \psi^j_{1y}\phi^j_{2y}$$

$$\vdots$$

$$L^p : \ \psi^j_{p,y} \ - \phi^{\ j}_{\ 1,y}\psi^j_{p-1,y} \ + \ \phi^j_{2,y}\psi^j_{p-2,y} \ + \ \cdots \ + \ \phi^j_{p-1,y}\psi^j_{1,y} \ = \ \phi^j_{p,y}$$

$$\vdots$$

For the model of order p, we obtain,

$$\phi^j_{p,y} \ = \ \psi^j_{p,y} \ - \ \phi^{\ j}_{\ 1,y}\psi^j_{p-1,y} \ - \ \phi^j_{2,y}\psi^j_{p-2,y} \ - \ \cdots \ - \ \phi^j_{p-1,y}\psi^j_{1,y} \tag{3.11}$$

$$\phi^j_{p,y} \ = \psi^j_{py} - \sum_{i=1}^{p-1} \phi^j_{iy}\psi^j_{p-i,y} \tag{3.12}$$

$$\psi^j_{ly} = \sum_{i=1}^{p} \phi^j_{iy}\psi^j_{l-i,y} \quad for \quad l > p \tag{3.13}$$

The same applies for z process.

$$\phi^j_{p,y} \ = \psi^j_{py} - \sum_{i=1}^{p-1} \phi^j_{iy}\psi^j_{p-i,y} = \psi^j_{pz} - \sum_{i=1}^{p-1} \phi^j_{iz}\psi^j_{p-i,z} = \ \phi^j_{p,z} \tag{3.14}$$

$$\phi_{k,y}^{j} \;=\; R_y^2(\rho\prime_{k,y}^{(j)})\prime \;=\; R_z^2(\rho\prime_{k,z}^{(j)})\prime \;=\; \phi_{k,z}^{j}$$

$$R_y^2 \;=\; \rho\prime_p^{(j)}\phi_{pY}^{j} \;\;=\;\; \rho\prime_q^{(j)}\phi_{qZ}^{j} \;=\; R_z^2$$

$$R_y^2 \;-\; R_z^2 \;=\; 0$$

$\square$

3.3.2 Testing for Equal Predictive Ability

The theoretical results of the previous section suggest the possibility of testing the equal predictive ability of two SETAR processes by checking if the structure of the series are similar. Verifying the hypothesis of equal predictive ability is an important task in several fields. For example, it could help to choose among several procedures of seasonal adjustment; it could help in investment decisions, establishing the different degrees of predictive ability among several returns Otranto and Triacca (2007).

Lemma 3.1 provides the way to confirm the similarity of the equal predictive ability and parallelism between two SETAR processes. The null hypothesis can be expressed in terms of the parameters of the time series.

$$H_o: \;\; \phi_{k,y}^{j} \;=\; \phi_{k,z}^{j} \;\; vs \;\; H_1: \;\; \phi_{k,y}^{j} \;\neq\; \phi_{k,z}^{j} \tag{3.15}$$

The null hypothesis can be expressed as a set of linear restrictions:

$$\mathbf{A\Theta} \;=\; \mathbf{0}$$

where **Θ** = $(\phi^j_{1,y}, \phi^j_{2,y}, ..., \phi^j_{p,y}, \phi^j_{1,z}, \phi^j_{2,z}, ..., \phi^j_{p,z})\prime$ and **A** is a $m \times m$ matrix of the form:

$$A = \begin{pmatrix} I_m & -I_m & 0_m & \cdots & 0_m & 0_m \\ 0_m & I_m & -I_m & \cdots & 0_m & 0_m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_m & 0_m & 0_m & \cdots & I_m & -I_m \end{pmatrix}$$

with I_m representing an $m \times m$ identity matrix and 0_m an $m \times m$ matrix with all the elements equal to zero, $m = p$. The Wald test statistic for the null hypothesis is of the form:

$$W = (A\hat{\Theta})'(A\hat{\Lambda}A')^{-1}(A\hat{\Theta})$$

where $\hat{\Theta}$ is the maximum likelihood estimator of Θ, whereas $\hat{\Lambda}$ is the maximum likelihood estimator of the covariance matrix of Θ. W is asymptotically distributed as a central chi-square random variable with m degrees of freedom.

Chapter 4

SOME APPLICATIONS

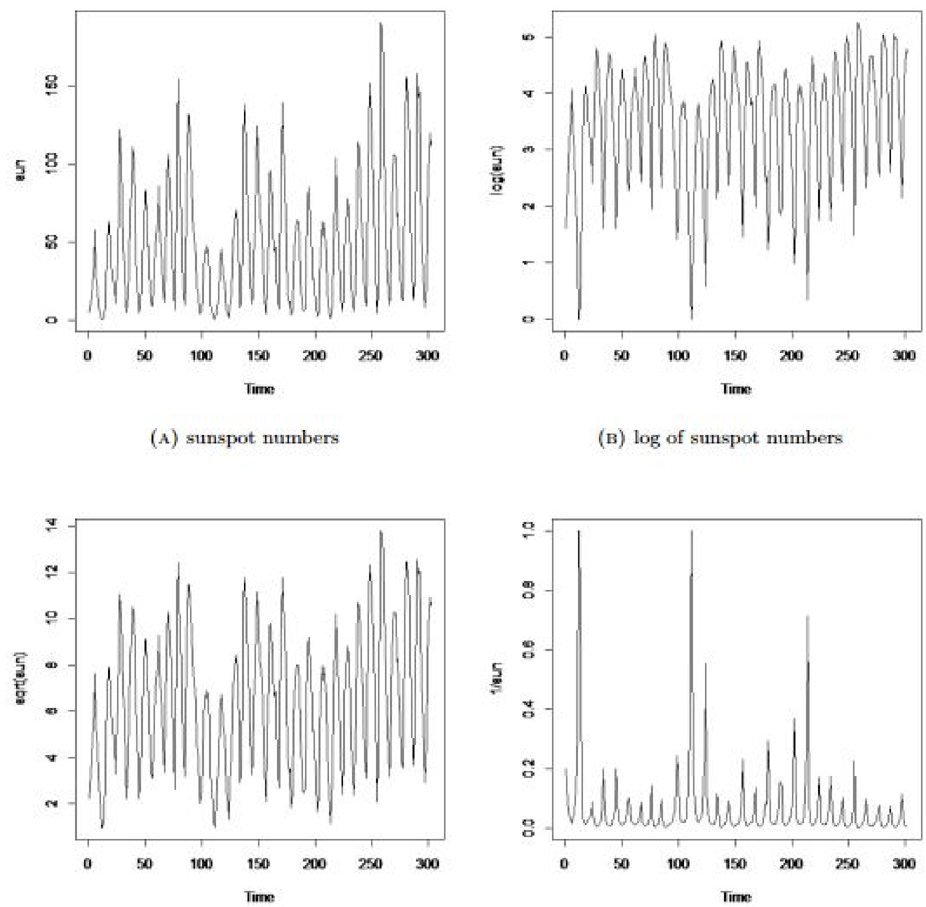
The purpose of the numerical examples presented here is to evaluate equal predictive ability between different Self-Exciting Threshold Autoregressive process by using the parameters of the models.

4.1 Numerical examples

4.1.1 Example 1: Consider different transformations of the Wolf Yearly Sunspots numbers.

In this application we use the Wolf Sunspots collected from 1700 to 2001, the series consists of 302 observations. The sunspots number is a weighted average of solar activities measured from a network of observatories. Historically, the daily sunspot number was computed as some weighted sum of the count of visible, distinct spots and that of clusters of spots on the solar surface. The sunspot numbers reflects the intensity of solar activity.

The sunspots activity is cyclical and reaches its maximum cycle around 9 to 11 years (source Wei (2006)). From Figure 1 we see that the plot of the sunspot (A) differs from (B) the log transform and (D) the reciprocal, but considerably similar to (B) the square root transform.



(A) sunspot numbers

(B) log of sunspot numbers

(C) square root of sunspot numbers

(D) reciprocal of sunspot numbers

FIGURE 1. Different transformations of the sunspot numbers

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	13.358	1.729	7.723	0
lag 1	1.591	0.053	29.888	0
lag 2	-0.82	0.047	-17.578	0
RSE	15.68			
n	260			
R^2	0.914			
Upper regime				
Intercept	-9.985	15.86	-0.63	0.533
lag 1	1.247	0.125	9.938	0
lag 2	-0.356	0.081	-4.370	0
RSE	16.958			
n	40			
R^2	0.979			

Table 4.1: Estimates of the parameters for Wolf's sunspot numbers. SETAR(2;2,2), d=1, Estimated threshold=100.1

In the time series, data falling in the lower regime consists of those lag 1 values which are less less than 100.1, whereas data falling in the upper regime are those lag 1 values which are greater than 100.1. The nonlinearity is most observed at lag 1 and the point of change is observed at 100.1. The model has high predictive ability since R^2 are 0.914 and 0.979 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are significantly different from zero except for the intercept of the upper regime. The model is given below;

$$z_t = \begin{cases} 13.358 + 1.591 z_{t-1} - 0.82 z_{t-2}, & z_{t-1} \leq 100.1 \\ -9.985 + 1.247 z_{t-1} - 0.356 z_{t-2}, & z_{t-1} > 100.1 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	1.098	0.319	3.446	0
lag 1	1.089	0.187	5.837	0
lag 2	-0.371	0.138	-2.692	0
RSE	0.851			
n	66			
R^2	0.891			
Upper regime				
Intercept	0.132	0.137	0.963	0.057
lag 1	1.617	0.047	33.26	0
lag 2	-0.7	0.036	-19.262	0
RSE	0.333			
n	234			
R^2	0.993			

Table 4.2: Estimates of the parameters for Log_e Wolf's sunspot numbers. SETAR(2;2,2), d=1, Estimated threshold=2.632

In the time series, data falling in the lower regime consists of those lag 1 values which are less less than 2.632, whereas data falling in the upper regime are those lag 1 values which are greater than 2.632. The nonlinearity is most observed at lag 1 and the point of change is observed at 2.632. The model has high predictive ability since R^2 are 0.891 and 0.993 for the respective regimes. Also RSE is different in both regimes signifying conditional heteroscedasticity. The parameters are significantly different from zero except for the intercept of the upper regime.

The model is given below;

$$z_t = \begin{cases} 1.098 + 1.089 z_{t-1} - 0.371 z_{t-2}, & z_{t-1} \leq 2.632 \\ 0.132 + 1.617 z_{t-1} - 0.7 z_{t-2}, & z_{t-1} > 2.632 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	0.809	0.709	1,141	0.25
lag 1	1.439	0.326	4.414	0
lag 2	-0.469	0.176	-2.665	0.04
RSE	1.359			
n	52			
R^2	0.877			
Upper regime				
Intercept	1.599	0.223	7.164	0
lag 1	1.471	0.044	33.613	0
lag 2	-0.732	0.04	-18.124	0
RSE	1.099			
n	248			
R^2	0.979			

Table 4.3: Estimates of the parameters for Square root Wolf's sunspot numbers. SETAR(2;2,2), d=1, Estimated threshold=3.317

In the time series, data falling in the lower regime consists of those lag 1 values which are less less than 3.317, whereas data falling in the upper regime are those lag 1 values which are greater than 3.317. The nonlinearity is most observed at lag 1 and the point of change is observed at 3.317. The model has high predictive ability since R^2 are 0.877 and 0.979 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are significantly different from zero except for the intercept of the lower regime .

The model is given below;

$$z_t = \begin{cases} 0.809 + 1.439 z_{t-1} - 0.469 z_{t-2}, & z_{t-1} \leq 3.317 \\ 1.599 + 1.471 z_{t-1} - 0.732 z_{t-2}, & z_{t-1} > 3.317 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	0.0023	0.004	0.608	0.544
lag 1	1.536	0.1	15.31	0
lag 2	-0.252	0.039	-6.483	0
RSE	0.0016			
n	267			
R^2	0.665			
Upper regime				
Intercept	0.083	0.071	1.179	0.205
lag 1	0.843	0.208	4.045	0
lag 2	-0.425	0.197	-2.155	0.03
RSE	0.249			
n	33			
R^2	0.635			

Table 4.4: Estimates of the parameters for reciprocal Wolf's sunspot numbers. SETAR(2;2,2), d=1, Estimated threshold=0.125

In the time series, data falling in the lower regime consists of those lag 1 values which are less less than 0.125, whereas data falling in the upper regime are those lag 1 values which are greater than 0.125. The nonlinearity is most observed at lag 1 and the point of change is observed at 0.125. The model has good predictive ability since R^2 are 0.665 and 0.635 for the respective regimes. Also RSE is different in both regimes signifying conditional heteroscedasticity. The parameters are significantly different from zero except for the intercepts of the two regimes.

The model is given below;

$$z_t \;=\begin{cases} 0.0023 \;+\; 1.536 z_{t-1} \;-\; 0.252 z_{t-2}, \;\; z_{t-1} \leq 0.125 \\ \\ 0.083 \;+\; 0.843 z_{t-1} \;-\; 0.425 z_{t-2}, \;\; z_{t-1} > 0.125 \end{cases}$$

Using the Wald's test statistic given by;

$$W \;=\; (A\hat{\Theta})'(A\hat{\Lambda}A')^{-1}(A\hat{\Theta})$$

where,

$$A \;=\; \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

$$\hat{\Lambda}^{j} \;=\; \frac{1}{n}\begin{pmatrix} 1-\phi_2^2 & -(\phi_1 \;+\; \phi_1\phi_2) \\ -(\phi_1 \;+\; \phi_1\phi_2) & 1-\phi_2^2 \end{pmatrix}^{j}$$

with j signifying the regime of interest.

$$\hat{\Theta} \;=\; (\phi^{j}_{1z_1}, \phi^{j}_{2z_1}, \phi^{j}_{1z_2}, \phi^{j}_{2z_2})$$

The covariance matrices for the different models are as follows;

For the sunspots data

$$\hat{\Lambda}^{1} \;=\; \begin{pmatrix} 0.0013 & -0.0011 \\ -0.0011 & 0.0013 \end{pmatrix}^{1}$$

$$\hat{\Lambda}^{2} \;=\; \begin{pmatrix} 0.022 & -0.022 \\ -0.02 & 0.02 \end{pmatrix}^{2}$$

For the square root of Wolf sunspot numbers,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.015 & -0.014 \\ -0.014 & 0.015 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.0019 & -0.0015 \\ -0.0015 & 0.0019 \end{pmatrix}^2$$

For log_e of Wolf sunspot numbers,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.013 & -0.01 \\ -0.01 & 0.013 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.0022 & -0.0021 \\ -0.0021 & 0.0022 \end{pmatrix}^2$$

For the reciprocal of Wolf sunspot numbers,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.004 & -0.004 \\ -0.004 & 0.004 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.025 & -0.015 \\ -0.015 & 0.025 \end{pmatrix}^2$$

Test for parallelism and equal predictive ability for Sunspot numbers

Hypothesis

$$H_o: \ \phi^j_{k,z_1} \ = \ \phi^j_{k,z_2} \quad vs \quad H_1: \ \phi^j_{k,z_1} \ \neq \ \phi^j_{k,z_2} \ , \quad i = 1,2 \tag{4.1}$$

$\alpha = 0.05,$

Test statistics:

$$W \ = \ (A\hat{\Theta})'(A\hat{\Lambda}A')^{-1}(A\hat{\Theta})$$

Computation:

For comparison of the sunspots and log_e transforms, we have the following. Others are computed in the same manner.

$\mathbf{\Theta} \ = \ (\mathbf{1.591}, \mathbf{-0.82}, \mathbf{1.089}, \mathbf{-0.371})$

$$\hat{\Lambda} \ = \begin{pmatrix} 0.0013 & -0.0011 & 0 & 0 \\ -0.0011 & 0.0013 & 0 & 0 \\ 0 & 0 & 0.013 & -0.01 \\ 0 & 0 & -0.01 & 0.013 \end{pmatrix}$$

	S	RS	SS
RS	676 (0)		
	10.006 (0.004)		
SS	20.529 (0.0001)	49.49 (0)	
	366.835 (0)	15.026 (0.0006)	
LS	18.556 (0.0001)	55.128 (0)	12.012 (0.0025)
	5.666 (0.058)	24.747 (0.00001)	32.528 (0)

Table 4.5: Values of the test statistic and $p - values$ in parenthesis for Sunspot numbers. Where SS is the square root of the sunspot numbers, LS is its log_e, and RS is its reciprocal.

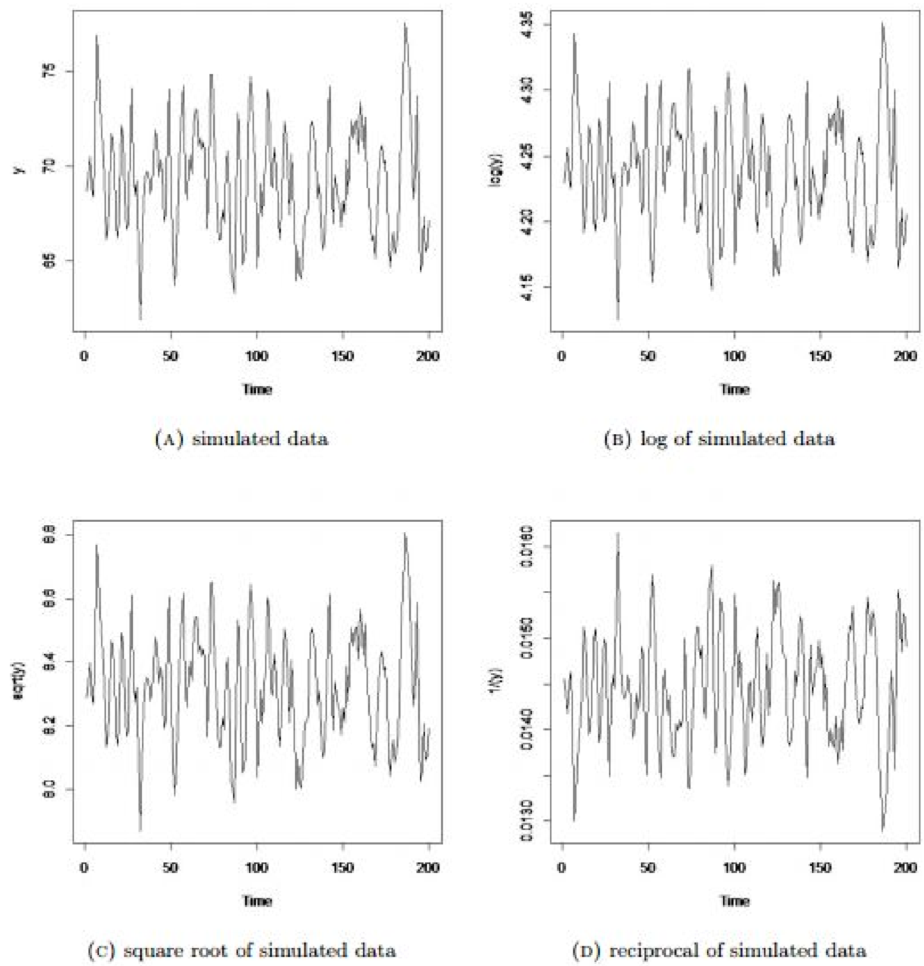
Conclusion: We conclude as follows; the sunspots numbers is not parallel to any of these transformations and neither is any of the transformations to one another. Even though the different models arrived at by using these transformations possess similar predictive abilities, for this reason of no parallelism, using these transformations may be inappropriate for modelling and inference.

4.1.2 Example 2: Consider different transformations of a simulated data.

The series consists of 200 observations. The series is simulated from the model give below;

$$z_t = \begin{cases} \phi_0^1 + \phi_1^1 z_{t-1} + \phi_2^1 z_{t-2} + \epsilon_t^1, & z_{t-2} \leq r \\ \phi_0^2 + \phi_1^2 z_{t-1} + \phi_2^2 z_{t-2} + \epsilon_t^2, & z_{t-2} > r \end{cases}$$

For $\phi_0^1 = 22$, $\phi_1^1 = 1.39$, $\phi_2^1 = -0.24$, $\phi_0^2 = 29$, $\phi_1^2 = 1.081$, $\phi_2^2 = -0.5$ and $r = 4$ The simulated data has cycles of approximately 8. In Figure 2 the simulated data has similar pattern except for plot (D) the peaks are not at the same time as the peaks of (A), (B) and(C). The cycles are evident, time irreversibility, etc.



(A) simulated data

(B) log of simulated data

(C) square root of simulated data

(D) reciprocal of simulated data

FIGURE 2. Different transformations of the simulated data

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	29.993	4.3	6.951	0
lag 1	1.222	0.067	18.354	0
lag 2	-0.656	0.079	-8.269	0
RSE	2.016			
n	157			
R^2	0.999			
Upper regime				
Intercept	10.699	11.59	0.923	0.362
lag 1	0.992	0.111	8.944	0
lag 2	-0.169	0.178	-0.952	0.347
RSE	1.811			
n	42			
R^2	0.999			

Table 4.6: Estimates of the parameters for Simulated data. SETAR(2;2,2), d=2, Estimated threshold=71.88

In the time series, data falling in the lower regime consists of those lag 2 values which are less less than 71.88, whereas data falling in the upper regime are those lag 2 values which are greater than 71.88. The nonlinearity is most observed at lag 2 and the point of change is observed at 71.88. The model has high predictive ability since R^2 is 0.999 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are largely significantly different from zero except for the intercept and lag 2 of the upper regime.

The model is given below;

$$z_t = \begin{cases} 29.993 + 1.222 z_{t-1} - 0.656 z_{t-2}, & z_{t-2} \leq 71.88 \\ 10.699 + 0.992 z_{t-1} - 0.169 z_{t-2}, & z_{t-2} > 71.88 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	0.005	0.0014	3.467	0
lag 1	1.254	0.0898	13.959	0
lag 2	-0.595	1.116	-5.143	0
RSE	0.0004			
n	102			
R^2	0.999			
Upper regime				
Intercept	0.0087	0.0013	6.447	0
lag 1	1.0722	0.0751	14.286	0
lag 2	-0.6626	0.0915	-7.243	0
RSE	0.0004			
n	96			
R^2	0.999			

Table 4.7: Estimates of the parameters for the reciprocal of Simulated data. SETAR(2;2,2), d=2, Estimated threshold=0.0145

In the time series, data falling in the lower regime consists of those lag 2 values which are less less than 0.0145, whereas data falling in the upper regime are those lag 2 values which are greater than 0.0145. The nonlinearity is most observed at lag 2 and the point of change is observed at 0.0145. The model has high predictive ability since R^2 is 0.999 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are significantly different from zero.

The model is given below;

$$z_t \ = \begin{cases} 0.005 \ + \ 1.254 z_{t-1} \ - \ 0.595 z_{t-2}, \ \ z_{t-2} \leq 0.0145 \\ \\ 0.0087 \ + \ 1.0722 z_{t-1} \ - \ 0.413 z_{t-2}, \ \ z_{t-2} > 0.0145 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	1.974	0.251	7.871	0
lag 1	1.209	0.064	18.888	0
lag 2	-0.676	0.074	-9.151	0
RSE	0.0292			
n	163			
R^2	1			
Upper regime				
Intercept	1.179	0.821	1.436	0.161
lag 1	0.965	0.129	7.489	0
lag 2	-0.246	0.216	-1.139	0.263
RSE	0.0256			
n	35			
R^2	1			

Table 4.8: Estimates of the parameters for the log_e of Simulated data. SETAR(2;2,2), d=2, Estimated threshold=4.281

In the time series, data falling in the lower regime consists of those lag 2 values which are less less than 4.281, whereas data falling in the upper regime are those lag 2 values which are greater than 4.281. The nonlinearity is most observed at lag 2 and the point of change is observed at 4.281. The model has high predictive ability since R^2 is 1 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are largely significantly different from zero except for the intercept and lag 2 of the upper regime..

The model is given below;

$$z_t = \begin{cases} 1.974 + 1.209z_{t-1} - 0.676z_{t-2}, & z_{t-2} \leq 4.281 \\ 1.179 + 0.965z_{t-1} - 0.246z_{t-2}, & z_{t-2} > 4.281 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	3.849	0.495	7.785	0
lag 1	1.216	0.064	18.958	0
lag 2	-0.680	0.075	-9.122	0
RSE	0.1216			
n	163			
R^2	0.999			
Upper regime				
Intercept	2.345	1.603	1.462	0.153
lag 1	0.951	0.127	7.481	0
lag 2	-0.239	0.210	-1.138	0.264
RSE	0.108			
n	35			
R^2	0.999			

Table 4.9: Estimates of the parameters for square root of Simulated data. SETAR(2;2,2), d=2, Estimated threshold=8.506

In the time series, data falling in the lower regime consists of those lag 2 values which are less less than 8.506, whereas data falling in the upper regime are those lag 2 values which are greater than 8.506. The nonlinearity is most observed at lag 2 and the point of change is observed at 8.506. The model has high predictive ability since R^2 is 0.999 for the respective regimes. Also RSE is different in both regimes signifying conditional heteroscedasticity. The parameters are largely significantly different from zero except for the intercept and lag 2 of the upper regime..

The model is given below;

$$z_t = \begin{cases} 3.849 + 1.216 z_{t-1} - 0.680 z_{t-2}, & z_{t-2} \leq 8.506 \\ 2.345 + 0.951 z_{t-1} - 0.239 z_{t-2}, & z_{t-2} > 8.506 \end{cases}$$

The covariance matrices for the different models are as follows; For the simulated data,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.004 & -0.003 \\ -0.003 & 0.004 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.023 & -0.019 \\ -0.019 & 0.023 \end{pmatrix}^2$$

For the reciprocal of simulated data,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.006 & -0.005 \\ -0.005 & 0.006 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.006 & -0.004 \\ -0.004 & 0.006 \end{pmatrix}^2$$

For the log_e of simulated data,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.003 & -0.002 \\ -0.002 & 0.003 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.027 & -0.021 \\ -0.021 & 0.027 \end{pmatrix}^2$$

For the square root of simulated data,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.003 & -0.003 \\ -0.003 & 0.003 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.027 & -0.021 \\ -0.021 & 0.027 \end{pmatrix}^2$$

Test for parallelism and equal predictive ability for simulated data

Hypothesis

$$H_o: \ \phi^j_{k,z_1} = \phi^j_{k,z_2} \quad vs \quad H_1: \ \phi^j_{k,z_1} \neq \phi^j_{k,z_2}\ , \quad i = 1, 2 \tag{4.2}$$

$\alpha = 0.05$,

Test statistics:

$$W = (A\hat{\Theta})'(A\hat{\Lambda}A')^{-1}(A\hat{\Theta})$$

Computation:

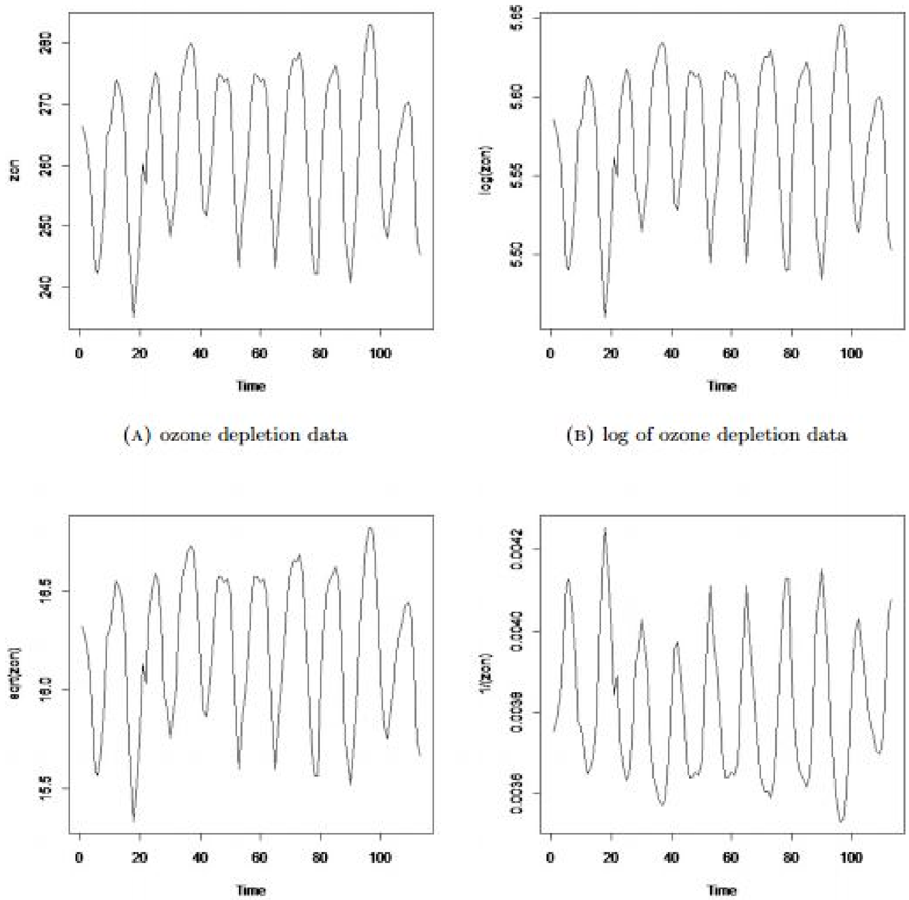
	S		RS		SS	
RS	2.186	(0.335)				
	17.408	(0.0002)				
SS	0.463	(0.793)	3.866	(0.145)		
	0.703	(0.704)	8.274	(0.016)		
LS	0.274	(0.874)	4.01	(0.135)	0.0098	(0.995)
	0.634	(0.728)	8.343	(0.015)	0.0044	(0.998)

Table 4.10: Values of the test statistic and $p-values$ in parenthesis for the Simulated data. Where S represents the simulation, SS is the square root of simulation, LS is its log_e, and RS is its reciprocal.

Conclusion: We conclude as follows; the reciprocal of the simulated data is not parallel to the originally simulated data, the square root transform and the log_e, but they all have equal predictive ability. But for the purpose of modelling and inference it will be appropriate to use either the square root or log_e transforms of the data, since they possess similar structure with the actual simulated data.

4.1.3 Example 3: Consider different transformations of Abuja Ozone Depletion data.

Ozone layer prevents most harmful ultraviolent (UV) wavelengths of ultraviolent light from passing through to the Earth's atmosphere. Ozone depletion is measured by reduction in the total ”column ozone” above a point on the Earth's surface. Ozone column is observed using instrument such as the total ozone mapping spectrometer (TOMS). The series consists of 114 observations (source: Department of Physics and Astronomy, University of Nigeria, Nsukka). Figure 3 displays Abuja Ozone depletion data showing the signs of time irreversibility, it cyclical nature etc. The plot shows that (A), (B) and (C) are similar and different from (D).



(A) ozone depletion data

(B) log of ozone depletion data

(C) square root of ozone depletion data

(D) reciprocal of ozone depletion data

FIGURE 3. Different transformations of the ozone depletion data

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	76.143	21.136	3.603	0
lag 1	0.974	0.109	8.934	0
lag 2	-0.063	0.165	-0.384	0.703
lag 3	-0.191	0.115	-1.662	0.104
RSE	3.213			
n	47			
R^2	0.999			
Upper regime				
Intercept	59.359	24.001	2.348	0.02
lag 1	1.772	0.140	12.628	0
lag 2	-1.053	0.251	-4.188	0
lag 3	0.065	0.170	0.385	0.701
RSE	3.377			
n	63			
R^2	0.999			

Table 4.11: Estimates of the parameters for Ozone Depletion data. SETAR(2;3,3), d=3, Estimated threshold=260.7

In the time series, data falling in the lower regime consists of those lag 3 values which are less less than 260.7, whereas data falling in the upper regime are those lag 3 values which are greater than 260.7. The nonlinearity is most observed at lag 3 and the point of change is observed at 260.7. The model has high predictive ability since R^2 is 0.999 for the respective regimes. Also RSE is slightly different in both regimes signifying slightly conditional heteroscedasticity. The parameters are significantly different from zero except for the lag 2 and 3 in the lower regime and lag 3 of the upper regime.

The model is given below;

$$z_t \;=\; \begin{cases} 76.143 \;+\; 0.974 z_{t-1} \;-\; 0.063 z_{t-2} \;-\; 0.191 z_{t-3}, & z_{t-3} \leq 260.7 \\ 56.359 \;+\; 1.772 z_{t-1} \;-\; 1.053 z_{t-2} \;+\; 0.065 z_{t-3}, & z_{t-3} > 260.7 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	4.674	1.312	3.562	0
lag 1	0.968	0.108	8.976	0
lag 2	-0.067	0.162	-0.413	0.682
lag 3	-0.185	0.113	-1.634	0.110
RSE	0.0998			
n	47			
R^2	1			
Upper regime				
Intercept	3.477	1.483	2.344	0.023
lag 1	1.779	0.141	12.608	0
lag 2	-1.059	0.254	-4.169	0
lag 3	0.065	0.172	0.379	0.706
RSE	0.105			
n	63			
R^2	1			

Table 4.11: Estimates of the parameters for square root of Ozone Depletion data. SETAR(2;3,3), d=3, Estimated threshold=16.14.

In the time series, data falling in the lower regime consists of those lag 3 values which are less less than 16.14, whereas data falling in the upper regime are those lag 3 values which are greater than 16.14. The nonlinearity is most observed at lag 3 and the point of change is observed at 16.14. The model has high predictive ability since R^2 is 1 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are significantly different from zero except for the lag 2 and 3 in the lower regime and lag 3 of the upper regime.

The model is given below

$$z_t = \begin{cases} 4.674 + 0.968z_{t-1} - 0.067z_{t-2} - 0.185z_{t-3}, & z_{t-3} \leq 16.14 \\ 3.477 + 1.779z_{t-1} - 1.059z_{t-2} + 0.065z_{t-3}, & z_{t-3} > 16.14 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	1.609	0.451	3.567	0
lag 1	0.963	0.107	9.019	0
lag 2	-0.071	0.160	-0.443	0.66
lag 3	-0.179	0.111	-1.606	0.118
RSE	0.012			
n	47			
R^2	1			
Upper regime				
Intercept	1.194	0.514	2.325	0.024
lag 1	1.786	0.142	12.583	0
lag 2	-1.065	0.257	-4.147	0
lag 3	0.065	0.175	0.372	0.711
RSE	0.013			
n	63			
R^2	1			

Table 4.12: Estimates of the parameters for Log_e of Ozone Depletion data. SETAR(2;3,3), d=3, Estimated threshold=5.563

In the time series, data falling in the lower regime consists of those lag 3 values which are less less than 5.563, whereas data falling in the upper regime are those lag 3 values which are greater than 5.563. The nonlinearity is most observed at lag 3 and the point of change is observed at 5.563.The model has high predictive ability since R^2 is 1 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are largely significantly different from zero except for the lag 2 and 3 in the lower regime and lag 3 of the upper regime. The model is given below;

$$z_t = \begin{cases} 1.609 + 0.963z_{t-1} - 0.071z_{t-2} - 0.179z_{t-3}, & z_{t-3} \leq 5.563 \\ 1.194 + 1.786z_{t-1} - 1.065z_{t-2} + 0.065z_{t-3}, & z_{t-3} > 5.563 \end{cases}$$

Lower regime	Estimates	Std Err	t-value	$Pr(>\|t\|)$
Intercept	0.0008	0.0004	2.328	0.023
lag 1	1.799	0.144	12.523	0
lag 2	-1.076	0.263	-4.097	0
lag 3	0.064	0.180	0.355	0.724
RSE	0.0004			
n	63			
R^2	0.999			
Upper regime				
Intercept	0.001	0.0003	3.433	0
lag 1	0.951	0.104	9.106	0
lag 2	-0.079	0.155	-0.507	0.615
lag 3	-0.167	0.108	-1.547	0.129
RSE	0.0001			
n	47			
R^2	0.999			

Table 4.13: Estimates of the parameters for Reciprocal of Ozone Depletion data. SETAR(2;3,3), d=3, Estimated threshold=0.004

In the time series, data falling in the lower regime consists of those lag 3 values which are less less than 0.004, whereas data falling in the upper regime are those lag 3 values which are greater than 0.004. The nonlinearity is most observed at lag 3 and the point of change is observed at 0.004. The model has high predictive ability since R^2 is 0.999 for the respective regimes. Also RSE is not so different in both regimes signifying slight conditional heteroscedasticity. The parameters are largely significantly different from zero except for the lag 2 and 3 in the lower regime and lag 3 of the upper regime. The model is given below;

$$z_t \;=\; \begin{cases} 0.0008 \;+\; 1.799z_{t-1} \;-\; 1.076z_{t-2} \;+\; 0.064z_{t-3}, \;\; z_{t-3} \leq 0.004 \\ \\ 0.001 \;+\; 0.951z_{t-1} \;-\; 0.079z_{t-2} \;-\; 0.167z_{t-3}, \;\; z_{t-3} > 0.004 \end{cases}$$

The covariance matrices for the different models are as follows;

For the Ozone Depletion data,

$$\hat{\Lambda}^1 \;=\; \begin{pmatrix} 0.021 & -0.021 & 0.005 \\ -0.021 & 0.041 & -0.021 \\ 0.005 & -0.021 & 0.021 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 \;=\; \begin{pmatrix} 0.016 & -0.027 & 0.015 \\ -0.027 & 0.048 & -0.027 \\ 0.015 & -0.027 & 0.016 \end{pmatrix}^2$$

For the square root of Ozone Depletion data,

$$\hat{\Lambda}^1 \;=\; \begin{pmatrix} 0.021 & -0.021 & 0.005 \\ -0.021 & 0.040 & -0.021 \\ 0.005 & -0.021 & 0.021 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.016 & -0.027 & 0.015 \\ -0.027 & 0.048 & -0.027 \\ 0.015 & -0.027 & 0.016 \end{pmatrix}^2$$

For the log_e of Ozone Depletion data,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.021 & -0.021 & 0.005 \\ -0.021 & 0.040 & -0.021 \\ 0.005 & -0.021 & 0.021 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.016 & -0.027 & 0.015 \\ -0.027 & 0.048 & -0.027 \\ 0.015 & -0.027 & 0.016 \end{pmatrix}^2$$

For the reciprocal of Ozone Depletion data,

$$\hat{\Lambda}^1 = \begin{pmatrix} 0.016 & -0.028 & 0.015 \\ -0.028 & 0.049 & -0.028 \\ 0.015 & -0.028 & 0.016 \end{pmatrix}^1$$

$$\hat{\Lambda}^2 = \begin{pmatrix} 0.021 & -0.021 & 0.005 \\ -0.021 & 0.040 & -0.021 \\ 0.005 & -0.021 & 0.021 \end{pmatrix}^2$$

Test for parallelism and equal predictive ability for Abuja Ozone Depletion data

Hypothesis

$$H_o: \ \phi^{j}_{k,z_1} \ = \ \phi^{j}_{k,z_2} \quad vs \quad H_1: \ \phi^{j}_{k,z_1} \ \neq \ \phi^{j}_{k,z_2}\ , \quad i = 1, 2 \tag{4.3}$$

$\alpha = 0.05,$

Test statistics:

$$W \ = \ (A\hat{\Theta})'(A\hat{\Lambda}A')^{-1}(A\hat{\Theta})$$

Computation:

	Z	RZ	SZ
RZ	21.03 (0.0001)		
	20.994 (0.0001)		
SZ	0.0035 (0.9999)	21.56 (0.0001)	
	0.0127 (0.9996)	21.37 (0.0001)	
LZ	0.0122 (0.9996)	21.816 (0.0001)	0.0027 (0.9999)
	0.0506 (0.9970)	21.75 (0.0001)	0.0127 (0.9996)

Table 4.14: Values of the test statistic and $p - values$ in parenthesis for the Abuja Ozone Depletion data. Where Z represents the ozone depletion data, SZ is the square root of ozone depletion data, LZ is its log_e, and RZ is its reciprocal.

Conclusion: We conclude as follows; the reciprocal of the ozone depletion data is not parallel to the ozone depletion data, the square root transform and the log_e, but they all have equal predictive ability. But for the purpose of modelling and inference it will be appropriate to use either the square root or log_e transforms of the data since they possess similar structure with the actual data.

Chapter 5

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 SUMMARY

In this work, we considered the parallelism and equal predictive ability between different Self-Exciting Threshold Autoregressive models in which the comparison was between different transformations of the given time series data. We established conditions for the relationship between equal predictive ability and parallelism. We applied the testing procedure proposed by Otranto and Traccia (2007) to suit nonlinear time series. The test was performed using the parameters of the models. We checked the test using different data sets including a simulated one. The transformations considered include; square root transform, log_e transform and reciprocal transform. These transformations were tested against one another and the non-transformed data.

5.2 CONCLUSION

From the work we see that the equal predictive ability of different models can be tested for by using the parameters of the model. We also see that when the models are parallel (judging from the plots of the data set and the test) they also have equal predictive ability, for this reason we can simply check whether the data sets are parallel, instead of computing the predictive ability of each data set. While in other cases where the estimated R^2 shows equal predictive ability for all, the test will still indicate which model(s) (or data transformation(s)) are more appropriate for forecasting by considering the parallelism between the series (which indicates similarities in the structure of the time series).

5.3 RECOMMENDATION

We recommend that when data need transforming, it will be necessary to also consider whether transformations are parallel to the original data this is because in building time series models, Box and Jenkins (1970) have devised an iterative strategy of model identification, estimation and diagnostic checking. The identification stage of their model building cycle relies on the recognition of typical patterns of behaviour in the sample autocorrelation function (its transform and the partial autocorrelation). Our results indicate that if the series to be fitted is subjected to transformation, it can be the case that the autocorrelation function of the transformed series exhibits markedly different behaviour patterns from that of the original series which can be easily spotted by checking if there are parallel. When different data sets (rather than transformations) which may be affected by similar innovations or economics variables are of interest, it will be much easier to investigate if the data sets are parallel, and if they are, it will lessen the task of analysing and forecasting for each data set. This can be done by collecting the data into different parallel groups and since they are parallel, we select one of them and use it to make inferences of the others.

References

Amato, J. and Swanson, N.R. (2001) The real time predictive content of money for output. *Journal of Monetary Economics*.

Box, G.E.P. and Jenkins, G.M.(1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Corradi, V. Swanson, N. R. and Olivetti, C, (2001) Predictive Ability with Cointegrated Variables. *Journal of Econometrics*, 104, 315-358.

Cryer, J.D. and Chan,K. (2008). *Time Series Analysis with Applications in R*. Springer Science.

Diebold, F.X. and Mariano, R.S. (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253263.

Dhrymes, P. J., Howery, E. P., Hymans, S. H., Kmenta, J., Learner, E. E., Quandt, R. E., Ramsey, J. B., Shapiro, H. T., and Zarnowitz, V. (1972), Criterion for Evaluation of Econometric Models. *Annals of Economic and Social Measurement*, 1, 291-324.

Hamilton, J.D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.

Howery, E. P, Klein, L. R and McCarthy, M. D. (1974), Notes on Testing the Predictive Performance of Econometric Models, *International Economic Review*, 15, 366-383.

Granger, C. W, J and Newbold, P. (1976) Forecasting Transformed Series. *J. R Statistical Society*, 189-203.

Guo, J. H (1999) A Nonparametric Test for Parallelism of Two First-Order Autoregressive Processes. *The Australian and Newzealand Journal of Statistics*, 41, 59-65.

Nelson C. (1976) The Interpretation of R^2 in Autoregressive-Moving Average Time Series Models. *The American Statistician*, 30, 175-180.

Otranto, E. (2004) Classifying the Market Volatility with ARMA Distance Measures. *Quaderni di Statistica*, 7, 1-19

Otranto, E. and Triacca, U. (2007) Testing for Equal Predictability of Stationary ARMA Proceses, *Journal of Applied Statistics*,34 , 1091-1108.

Ozaki, T and Tong, H (1975) On Fitting Non-stationary Autoregressive Models in Time Series Analysis. *In Proc. 8th Hawaii Int, Conf. on System Sciences*, pp. 225-226.North Hollywood: Western Periodicals.

Piccolo D. (1990) A Distance Measure for Classifying ARIMA Models. Journal of Time Series Analysis, 11,153-164.

Priestley, M. B (1965) Evolutionary Spectra and Non-stationary Processes. *J. R. Statistical Society*,B, 27.

Steece, B. and Wood, T. (1985) A Test For The Equivalence of k ARMA Models, *Empirical Economics*, 10, 1-11.

Stock, J. H and Watson, M. W (1989) Interpreting the evidence on moneyincome causality *Journal of Econometrics* 40, 161181.

Stock, J. H and Watson, M. W (1999) Forecasting Inflation. *Journal of Monetary Economics*, 44, 293-335.

Tong, H (1978) On a Threshold Model in pattern Recognition and Signal Processing. *In Pattern Recognition and Signal Processing* (ed. C. H. Chen). Amsterdam:

Sijithoff and Noordhoff.

Tong, H and Lim K. S (1980) Threshold Autoregressive, Limit Cycles and Cyclical Data. *J. R. Statistical Society*, B, 42. 245-292.

Tong, H (1983) Threshold Models in Nonlinear Time Series Analysis. *Lecture Notes Statistics*.

Tong, H.(1990) *Nonlinear Time Series: A Dynamic System Approach.* Oxford Uni versity Press, New York.

Tsay, R.S. (1989) Testing and modelling threshold autoregressive processes, *Journal of American Statistical Association*, 84, 231-240.

Triacca, U. (2004) A Note on Distance and Parallelism Between Two ARIMA Pro cesses, *Quaderni di Statistica* 6, 1-9.

Vuong, Q. H (1989) Likelihood Ratio Tests for Model Selection and Non-nested Hypothesis, *Econometrica*, 57, 307-334

Wei, W.W.S. (2006) *Time Series Analysis: Univariate and Multivariate Methods.* Boston: Pearson Education,Inc.

West, K.D.(1996) Asymptotic inference about predictive ability. *Econometrica.* 64, 1067-1084.