# IMPUTATION OF MISSING VALUES FOR BILINEAR TIME SERIES MODELS

**OWILI POTI ABAJA**

**A Thesis Submitted to the Institute of Postgraduate Studies In Partial Fulfillment of the Requirement for the Award of Degree Doctor of Philosophy in Statistics**

**KABARAK UNIVERSITY**

**OCTOBER, 2016**

# COPYRIGHT

## DECLARATION

This research thesis is my original work and has not been presented for a degree, diploma or any other award in this or any other university.

Sign:________________________________ Date__________________________

Poti Owili Abaja

GDS/M/1022/9/10

Kabarak University, Private Bag_20157, Kabarak, Kenya

## RECOMMENDATIONS

To The Institute of Post Graduate Studies and Research:

The thesis entitled "Imputation of Missing Values for Bilinear Time Series Models" written by Poti Owili Abaja is presented to the Institute of Post Graduate Studies and Research of Kabarak University. We have reviewed the thesis and recommend it to be accepted in partial fulfillment of the requirements for the Degree of **Doctor of Philosophy in Statistics**.

Sign…………………………….. Date………………………..

Prof. Dankit K. Nassiuma

Africa International University

P.O. Box 24686-00502

**Nairobi**

Sign……………………………. Date……………………………………

Dr. Luke Orawo

Senior Lecturer

Egerton University

Private Bag

**Egerton**

## DEDICATION

I dedicate this work to my late dad Owili Abaja Abuor whose inspiration in academics was phenomenal; my late brother Stephen Opiyo and late sister Milcah Atieno.

## AKNOWLEDGEMENT

First and foremost I am grateful to my Lord Jesus Christ. His grace proved more than sufficient in the many hours of need.

Secondly, I want to really appreciate the contribution of my supervisors, Prof. Dankit Nassiuma, Dr. Luke Orawo and the late Dr. Jesse Mwangi for their guidance. It was splendid. They inculcated a sense of academic thinking that spurred me in navigating the uncertainties of academic wilderness. Prof. Nassiuma conceived this idea many years ago but I only realized its workability much later. This idea is what I have now belatedly presented here. I am highly indebted to him. Without the programming acumen of Dr. Orawo, it would have been a miracle obtaining the data that I have used in this research. I also pay tribute to the late Dr. Jesse Mwangi for creating in me self-belief, the vital impetus to accomplish the work. 'Your last words to me hours before you were unceremoniously nipped are a monument in my heart.'

I also wish to thank Kabarak University for the opportunity it gave me to study in the institution. I would also like to express my sincere gratitude to the Management of Laikipia University, especially the Vice Chancellor, Prof Francis Lelo, for the support and opportunity to study. The presence of my friends and colleagues in Laikipia, Egerton and Kabarak Universities was a constant reminder that I should not disappoint in this work.

Finally, I wish to acknowledge the contribution of my mother Monicah Owili, my children; Nicole, Stephen and Emmanuel. The encouragement and patience shown by my wife Susan was humbling and exceptional.

## ABSTRACT

Missing observations is a common occurrence in data collection. To solve this problem, researchers have developed missing value imputation techniques for some linear and nonlinear time series models with normal and stable innovations using estimating function criterion. This criterion does not take into consideration the distribution of the innovation sequence of the time series model. Therefore the aim of this study was to develop explicit optimal linear estimators of missing values for several classes of bilinear models whose innovation sequences are governed by the normal, student-t and generalized autoregressive heteroscedasticity using the minimum dispersion error criterion. For comparison purposes, estimates based on artificial neural networks and exponential smoothing were also obtained. Data was generated using the R statistical software. 100 samples of size 500 each were simulated for different bilinear time series models. In each sample, artificial missing observations were created randomly at points 48, 293 and 496 and estimated. The mean squared error was used to measure the efficiency of the estimates. The study found that the efficiency of the estimates was correlated with the probability distribution of the innovation sequence. Optimal linear estimates were the most efficient estimates when the models had normal and student-t innovations. However, for bilinear models with generalized autoregressive heteroscedasticity innovations, the artificial neural network estimates were the most efficient. The study recommends the use of optimal linear estimates for bilinear models with either normal or student-t errors. When the data is bilinear with generalized autoregressive heteroscedasticity errors, artificial neural network estimates are preferred. These findings can be used by econometricians in developing more accurate models.

**Key Words: Neural Networks, Exponential Smoothing, Optimal Linear Estimates**

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| ANN | Artificial Neural Networks |
| AR | Autoregressive |
| ANN-AR | Artificial Neural Networks and Autoregressive |
| ARCH | Autoregressive Conditional Heteroscedasticity |
| ARIMA | Autoregressive Integrated Moving Average |
| ARMA | Autoregressive Moving Average |
| BL | Bilinear |
| EM | Expectation Maximization |
| EM-MCMC | Expectation Maximization Monte Carlo Markov Chain |
| EOF | Empirical Orthogonal Functions |
| EXP | Exponential Smoothing |
| FFBP | Feed Forward Back Propagation |
| GARCH | Generalized Autoregressive Conditional Heteroscedasticity |
| HIV | Human Immunodeficiency Virus |
| IID | Independent Identically Distributed |
| K-NN | K- Nearest Neighbors |
| LES | Linear Exponential Smoothing |
| LLSimpute | Local Least Squares Imputation |
| MAD | Mean Absolute Deviation |
| MAE | Mean Absolute Error |
| MBE | Mean Bias Error |
| MINLP | Mixed Integer Mathematical Linear Programming |

| | |
|---|---|
| MLP | A Multi-Layer Perceptron |
| MSE | Mean Squared Error |
| NPRA | Nearest Parametric Regression Approach |
| NR | Normal Ratio |
| OLE | Optimal Linear Estimates |
| RHS | Right Hand Side |
| SES | Simple Exponential Smoothing |
| SMA | Simple Moving Average |
| SOM | Self Organizing Feature Maps |
| SSA | Singular Spectrum Analysis |
| Student-t | Student Distribution |
| TSCS | Time Series Cross-Sectional |

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background to the study

A time series is defined as data recorded sequentially over a specified time period. There are cases where all the data within the specified period are obtained resulting in a complete data set. This data is collected at equally spaced time steps and can be analyzed easily since techniques developed for complete and regular series are available (Musial, 2011). Further, inferences can be made that can preserve the statistical facts of the system (Campozan, et. al., 2014). However, data analysts are frequently faced with situations where one or several time series observations are missing at certain points within the data set collected for modeling (Pigott, 2001). This leads to missing values at such points.

Missing values may occur for various reasons which may include poor record keeping, lost records, technical errors, non-responses at the time of data collection, deletion of suspected outliers that were collected by mistake and also because the time series data was originally acquired at irregular times (Pigot, 2000; Fung, 2006). In addition, a peculiar case can arise when one might be interested in determining the likely value of a variable of interest at a time that may not coincide with a particular measurement or observation (Musial, 2011). Being unable to account for missing value(s) may lead to a severe misrepresentation of the phenomenon under study. In addition, the results of the analysis may be characterized by poor estimates and forecasts of time series (Abraham and Thavaneswaran, 1991). Ng and Panu (2010) also states that an incomplete data set may lead to complications and uncertainty in the analysis of the data. Ferreiro (1987) observed that the occurrence of missing observations is quite common in time series and in many cases it is necessary to estimate them. Hence it is imperative to find a solution to the missing value problem. Different suggestions have been made for dealing with missing values for different types of data.

## 1.2 Missing value imputation

Gupta (1996) suggests that one of the possible ways of dealing with missing values in multivariate data is to delete the incomplete cases from the dataset. This approach may lead to loss of valuable information. The other approach is to compute the missing value(s) using the rest of information in the dataset (Gupta and Lam, 1996). This approach is referred to as imputation. Imputation is defined as a procedure that is used to fill in missing values by using substitutes. It can also be defined as a statistical technique that is used to estimate missing values in an irregular time series (Fung, 2006; Owili, Nassiuma and Orawo, 2015a).

According to Abrahantes, et. al. (2011), imputation broadly comprises several techniques that have been developed to compute missing values. These techniques may employ basic and simple strategies such as mean substitution and neural networks approach (Denk and Weber, 2011). It may also involve the use of appropriate statistical prediction or forecasting models such as regression or time series models. More advanced modeling methods such as those based on Markov chain and Monte Carlo methods can also be used.

Imputation may also require the analysis of a similar and comparable record to the dataset with missing value(s) or use of skilled knowledge or experience (Sa¨rndal and Lundstro¨m, 2005). This is because for any incomplete dataset, the collected data values are deemed to provide indirect evidence about the expected values of the unobserved ones. This evidence can be combined with certain assumption to imply a predictive probability distribution for the missing value (Schafer and Olsen, 1998). Generally, the aim of missing data imputation approach is to compute a reasonable substitute for a missing observation and use the new complete dataset to carry out the desired modeling or analysis (McKnight, et. al., 2007).

Several factors should be considered in identifying an appropriate imputation method for a given data. Kaiser (2012) suggests the consideration of the structure of the data. He states that the commonly used method for missing values imputation in non continuous data is to substitute missing values of each attribute by its arithmetic average. For time series data and especially nonlinear time series models, advanced statistical methods may be required.

Complications do arise in the imputation of missing values due to various factors. These may include the number of missing patterns or observations and the nature of the data. That is, if categorical and continuous random variables are involved (Horton and Ken, 2007). However, routines, procedures, or packages capable of generating imputations for incomplete data in databases are now widely available. For databases one can use regression, correlation analysis and other non-parametric methods in computing the missing value. This does not apply to time series data especially when one takes into account innovation sequence.

Researchers have varied reasons for computing missing values. There are cases where they impute missing values so that they can use them to evaluate the accuracy of the estimates of the parameters of the model fitted after filling in the missing observations. However, in some situations, they could be interested in determining the quality of the imputed values at the level of the individual. In this case, no further analysis is done with the data after imputing the missing value. This is an issue that has not received much attention (Cortiñas, et. al., 2011). This study was concerned with the finding the accuracy of imputed value(s) at the specific points where they occurred in contrast to finding parameter estimates of the resulting model after the infilling of the missing value(s).

Several criteria may be used in the derivation of missing value(s) in nonlinear time series models. Abraham and Thavaneswaran (1991) used estimating function criterion. They developed an estimator for missing value for only a particular order of the simple bilinear time series model, BL (1, 0, 2, 0). This is a unique type of bilinear time series model where the lagged errors of the bilinear term do not include the innovation sequence. The other possible criteria that may be employed are the least squares method and maximum likelihood function. This study used a different criterion from estimating functions. Estimates were derived by minimizing the dispersion error. The estimates obtained are referred to as optimal linear estimates. This criterion has not been used before for estimating missing values for bilinear time series models.

### 1.3 Statement of the problem

Missing observations is a common occurrence facing data analysts and researchers involved in statistical modeling in diverse fields. To solve this problem, missing value imputation techniques have been developed for several linear and nonlinear time series models. Unfortunately these techniques are only appropriate for the autoregressive moving average (ARMA) models whose innovation sequences follow either the normal or infinite variance stable distributions. A bilinear time series model is a class of nonlinear time series which has ARMA models as its special case. As far as bilinear time series models is concerned, an estimator for missing values was developed for only a particular order of the simple bilinear time series, BL (1 ,0 ,2 ,0). Thus for several classes of bilinear time series models, there is no explicit method for estimating missing values. Further, the estimation of the missing value for BL (1, 0, 2, 0) was based on the estimating functions criterion which does not consider the distribution of the innovation sequence of models. Therefore, this study sought to fill these gaps by developing explicit methods for estimating missing values for different classes of bilinear time series models whose innovations follow the normal, student-t and generalized autoregressive heteroscedasticity probability distributions by minimizing the dispersion error. Pure bilinear and the general bilinear time series models were the main classes of bilinear time series considered. For comparison purposes, estimates of missing values for bilinear time series were also obtained using two commonly used nonparametric methods of artificial neural networks (ANN) and exponential smoothing (EXP).

## 1.4 Objectives of the study

The general and specific objectives of the study are given below.

### 1.4.1 General objective

The purpose of the study was to develop estimators of missing observations of bilinear time series models with different innovation sequence by minimizing the dispersion error.

### 1.4.2 Specific objectives

The specific objectives of the study were to:

a) Derive estimators for missing observations for bilinear time series models using linear interpolation technique when the innovations are identically and independently distributed normal sequences by minimizing dispersion error.

b) Derive estimators for missing observations for bilinear time series models when the innovations have independent and identical student-t distribution using linear interpolation techniques by minimizing dispersion error.

c) Derive estimators for missing values for bilinear time series models with GARCH errors using linear interpolation techniques by minimizing dispersion error.

d) Estimate missing values for bilinear time series models using non-parametric methods of artificial neural networks (ANN) and exponential smoothing (EXP) techniques.

e) Compare the efficiency of the estimates obtained and determine how they vary with the position of the missing data point.

**1.5 Significance of the study**

Time series models, among them bilinear time series models, are widely used in decision making especially in economics, environment and finance for prediction and forecasting purposes. These models play a key role in budgeting, forecasting and enhancing the understanding of the mechanisms generating data. For accurate and reliable results, the models constructed must be based on all the data that is supposed to be collected, be it sample or census data. This study has shown that efficient estimates of missing values can be obtained using optimal linear interpolation technique for bilinear time series models with normally and t-distributed innovation. The artificial neural networks can be used to estimate missing values for bilinear time series models with GARCH innovations. It is also evident that the estimation of missing values depends on the distribution of innovation sequence of the data. These findings are of benefit to researchers, university lecturers, data analysts and planners in the Government who are involved in modeling financial, economic and seismology data that can be modeled using bilinear time series models.

### 1.6 Limitations of the study

The study used the Time Series Model (TSM) software that is usually used for the analysis of time series data. However, it cannot model higher order pure diagonal bilinear models and hence only simple pure diagonal models were studied.

### 1.7 Scope of the study

The study focused on estimating missing values for bilinear time series only. While several methods for estimating missing values exist, the study used only two other methods of missing values imputations namely, artificial neural networks and exponential smoothing.

### 1.8 Assumptions of the study

The models used were assumed to be stationary bilinear time series and that the higher order moments were deemed insignificant. The innovation sequence was assumed to be independent and identically distributed (i.i.d) random variables when the models had the normal or the student-t distribution.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 Introduction

In this chapter, we first review the literature on nonlinear time series models, then review one on imputation of missing values for different types of time series data, namely: cross-sectional time series, micro-array time series and spatial-temporal time series. We also examined imputation methods used for linear and nonlinear time series models. Imputation methods used for nonparametric methods are also discussed.

## 2.2 Nonlinear time series models

Most of the real-life time series encountered in practice are adequately described by nonlinear models. Nonlinear models are appropriate for data that exhibits time irreversibility, outlying points and cyclicity. According to Nassiuma (1994), nonlinearity can be approached in two different ways; in the first case, nonlinearity is introduced in the structure of model but it is assumed that the distribution of the innovation sequence is Gaussian. Bilinear time series models by Subba Rao and Silva (1992), threshold autoregressive models by Tong (1983), exponential models by Haggan and Ozaki (1981), random coefficient autoregressive (RCA) models by Nicolls and Quinn (1982), state dependent models by Priestly (1980) and several of their modifications are good examples of models that fall in this category.

The second case is to assume that the process is still linear but the innovation sequence is non-Gaussian. In this category, we have finite and infinite variance time series models. Finite variance non-Gaussian processes include the gamma and exponential autoregressive processes (Gaver and Lewis, 1980; Jacobs and Lewis, 1977 and Lawrence and Lewis, 1980). A more complicated case involves models that consist of a blend of non-Gaussian and nonlinearity. A classic example is the bilinear models with infinite variance innovations (Liu, 1989). The important nonlinear time series models used in this study are described below.

### 2.2.1 ARCH models

A process $\{\varepsilon_t\}$ is an autoregressive heteroscedastic ARCH (q) model if the conditional distribution of $\{\varepsilon_t\}$ given the available information $\psi_{t-1}$ is expressed as

$$\varepsilon_t / \psi_{t-1} \sim N(0, h_t)$$

where

$$\varepsilon_t = \eta_t h_t^{1/2},\ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$$

The parameters of this model satisfy the conditions: $\alpha_i \geq 0$ for all i=0 1,2,3..., $\sum_{i=1}^{q} \alpha_i < 1$ and $\eta_t$ is a sequence of independent and identically distributed (i.i.d) random variables with mean zero and unit variance (Engle, 1982). An important property of these models is that they can describe the time varying stochastic conditional volatility (Islam, 2013). This can be used to improve the reliability of forecasts and to help in understanding the process. It is important to realize that the series $\{\varepsilon_t\}$ is a martingale difference and hence cannot be predicted. However, the squared series $\varepsilon_t^2$ can be forecasted with the best forecast given as

$$E(\varepsilon^2 / \psi_{t-1}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 .$$

These models include lagged variances in the prediction of future variances as indicated in Engle (2004), and thus  can be used in the measurement and forecasting of the time varying volatility of returns and financial assets observed at high sampling frequencies such as daily returns (Andersen, Bollerslev, Diebold and Labys, 2003). Further, they specifically take the dependence of the conditional second moments when modeling into consideration. This accommodates the increasingly important demand to explain and to model risk and uncertainty in financial time series (Degiannakis and Xekalaki, 2004; Engle, 2004; Fan and Yao, 2003).

Despite their importance in modeling financial data, the ARCH models have a relatively long lag length in the variance equation (Wagalla, et al., 2012). This implies that they contain many parameters that have to be estimated and hence it is not a parsimonious model. Bollerslev (1986) developed a more parsimonious model called the Generalized ARCH (GARCH) model. It uses a few number of parameters than ARCH model for modeling a given time series data. For the GARCH (p, q) models, the conditional variance is specified as

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \alpha_q \varepsilon_{t-q} + \beta_1 h_{t-1} + ... + \beta_p h_{t-p}$$

with the inequality conditions $\alpha_0 > 0$, $\alpha_i \geq 0$ for i=1,…,q, $\beta_i \geq 0$, for i=1,…,p. This ensures that the conditional variance is strictly positive.

### 2.2.2 ARMA models with ARCH errors

The ARMA model can be combined with an ARCH model to obtain an ARMA (k, l) process whose innovations $\{\varepsilon_t\}$ are ARCH (q). The ARMA (k, l) model with ARCH error is given by

$$x_t = \sum_{i=1}^{k} a_j x_{t-j} + \sum_{i=1}^{l} b_j \varepsilon_{t-j} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is ARCH (q) (Weiss,1984). The above model is basically an ARMA model and much of the theory of Box-Jenkins identification and estimation approach can be applied to it.

### 2.2.3 Bilinear time series models

A discrete time series process $X_t$ is said to be a bilinear time series model BL (p, q, m, k) if it satisfies the difference equation

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + \sum_{i=1}^{m} \sum_{j=1}^{k} b_{ij} X_{t-i} e_{t-j} + e_t,$$

where θ, $\phi$ and $B_{ij}$ are constants while $e_t$ is a purely random process and $\theta_o$ =1 (Granger and Andersen, 1978a; Subba Rao,1981). For example, the bilinear model BL (1, 1, 1, 1), is expressed as

$$x_t = \phi x_{t-1} + \theta e_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t \,.$$

Subba Rao (1984) showed that with a large bilinear coefficient $b_{ij}$, a bilinear model can have sudden large amplitude burst that can be suitable for some kind of seismological data such as earthquake and underground nuclear explosion data. A bilinear process is also time dependent. This feature enables bilinear processes to be used in modeling financial data (Maravall, 1983). Bilinear model is a member of the general class of nonlinear time series models referred to as 'State dependent models' formed by adding the bilinear term to the conventional autoregressive moving average (ARMA) model (Priestly, 1980).

Bilinear time series models and its variants have been used successfully for forecast improvement. Wagalla, et al. (2014) modeled different time series stock data at Nairobi Securities Exchange (NSE) and found that bilinear models with GARCH innovations gave more efficient estimates. Earlier, De Gooijer (1989) reported a forecast improvement with bilinear models in forecasting stock prices. In a much earlier study, Maravall (1983) used a bilinear time series model to forecast Spanish monetary data and reported a near 10% improvement in one step-ahead mean square forecast error over several ARMA alternatives.

The statistical properties of such models have been analyzed in detail by Granger and Andersen (1978a), Subba Rao and Gabr (1984), Hannan (1982), Liu and Brockwell (1988) etc., while an economic application is presented in Howitt (1988).

### 2.2.4 Bilinear time series model with ARCH innovations

According to Weiss (1984), the combined bilinear model with ARCH errors is given by

$$\Phi(B)(X_t - u) = \Theta(B)\varepsilon_t + \sum_{i=1}^{p}\sum_{j=1}^{Q} B_{ij} X_{t-i}\varepsilon_{t-j}$$

and

$$\varepsilon_t = \eta_t h_t^{1/2},$$

where E( $\varepsilon_t / \psi_{t-1}$ )=0 and $h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \delta_o (\hat{X}_t - u)^2 + \sum_{i=1}^{s} \delta_i (X_{t-i} - u)^2$ .

$\Theta(B)$ and $\Phi(B)$ are the characteristic polynomials, $\mu$ is the mean of the time series observations and $\{e_t\}$ is the innovation sequence. A stationary bilinear model can be expressed in a kind of moving average with infinite order according to Wold (1954) decomposition theorem. This enhances its application in making inferences.

**2.3 Identification of bilinear time series models**

Given a time series data, the first step in the identification process of bilinear time series model is to test whether the data can be modeled either as a linear time series or belongs to the broader class of nonlinear time series models. This involves testing a null hypothesis that the data is linear. This can be done using one of the statistical tests of linearity (Keenan, 1985; Tsay, 1986). If the null hypothesis is rejected then the data can be appropriately modeled by a nonlinear series model and a bilinear model is one of the candidate models that may be considered (Subba Rao, 1981; Subba Rao and Gabr, 1984). If the data is nonlinear then the second step follows.

The second step in the identification process is to determine the class of the nonlinear models to which the data belongs. This involves the use of moments and cumulants. It has been noted that BL (p, 0, p, 1) and ARMA (p, 1) models have similar second order moments and hence these moments cannot be conclusively used in identification of the bilinear time series models (Subba Rao, 1991). Consequently, it is imperative to use higher order moments of the data in the identification process. The higher moments are known to satisfy the Yule-Walker type difference equations (Subba Rao, 1988, 1991). Thus, the Yule-Walker type difference equations could be used for model identification of the bilinear time series models. The difference between bilinear time series and other nonlinear time series models is that the higher order moments of a bilinear time series (including the fourth moments) decay slowly as the lag tends to infinity. However, the fourth moments of the other nonlinear time series models do not behave in this manner.

After determining that the data is bilinear, then the order of the model is determined using canonical correlation analysis carried between the linear combinations of the observations and linear combinations of higher powers of the observations. The technique of identification of a given nonlinear model can be extended to more general bilinear models provided there are difference equations for higher order moments and cumulants (Subba, Rao and da Silva 1992).

For some super diagonal and diagonal bilinear time series, the third order moments are not equal to zero. This pattern of nonzero moments can be used to discriminate between white noise and the bilinear models and also between different bilinear models (Kumar, 1986). Using the patterns presented in a table of third order moments, one can easily distinguish bilinear models from pure ARMA or mixed ARMA models. Third order moments may also be useful in detecting non-normality in the distribution of the innovation sequence (Poti, Nassiuma and Orawo, 2015b).

## 2.4 Estimation of parameters of the bilinear time series models

Several estimation techniques have been proposed in the literature. Some of them deal with particular classes of the bilinear time series models. Subba Rao (1981) first proposed two methods for the estimation of the model parameters of a bilinear time series models, namely the use of Newton Raphson technique and the Marquart Algorithm. He applied both methods to the estimation of the parameters of a bilinear time series model identified for sunspost and seismology data. Secondly, he proposed estimation of the parameters of bilinear models using maximum likelihood method. More recently Shiqing, Liang, and Fukang (2015) proposed a generalized autoregressive conditional heteroskedasticity-type maximum likelihood estimator for estimating the unknown parameters for a special bilinear model. They showed that their proposed estimator was consistent and asymptotically normal under only finite fourth moment of errors. Mathews and Moon (1991) proposed the use of covariance estimates based on the least squares method on the parameters of the bilinear model BL (p , 0, p, 1). Won, Kim, Billard and Basawa (1990) estimated the parameter of the simple diagonal bilinear model BL (0, 0, 1, 1) using the least squares method.

## 2.5 Missing value imputation for time series cross-sectional data

Numerous imputation techniques have been proposed in the literature (Rubin, 1996; Särndal and Lundström, 2005) for the imputation of missing values in time series cross-sectional data (TSCS). These techniques are classified according to the type of dataset used; whether a non-parametric model is used or not, and if randomization is used or not for selection of imputed value. Among the first imputation methods used in TSCS was kernel density estimation in combination with nonparametric bootstrap (Titterington and Sedbranks, 1989). The other methods used for imputing missing values include: Expected Maximum Algorithm (EM), Single Imputation (SI), Multiple Imputation (MI) and Artificial Neural Networks (Bishop, 1995).

In Single Imputation (SI) approach each missing value is replaced by single imputed value using, say, interpolation approach or regression analysis. The replaced value is then treated as if it were an actual data value. This approach enables analysis with procedures designed for complete datasets. This method is simple and can be applied to any dataset. Its main disadvantage is that it does not account for the uncertainty about the predictions of the imputed value. Therefore the estimated variances of the parameters are biased towards zero, leading to statistically invalid inferences (Rubin, 1987). This can be overcome using multiple imputation method.

Multiple imputations (MI) is a methodology for estimating missing observations using a set of M reasonable estimates that represent the uncertainty about the right values. It has received considerable attention in the literature (Schafer, 1997). It maintains the flexibility and relative ease of application of single imputation while taking into account the variability due to the imputation of the missing values.

The application of MI has focused mainly on cross-sectional models for survey data. However, it has also been used on panel and time series data (King, et al, 2001). Although from a theoretical point of view there is no reason why MI cannot be used for time series data, its application has been difficult in practice. With cross sectional data, discarding records with data missing completely at random (MCAR) has the effect of only reducing the available sample. However, in a time series each record is unique and the data is also correlated; dropping it would leave the series with gaps, unusable for many purposes.

These imputation techniques have also been extended to sample surveys, where the object is to generalize estimates obtained in the survey to a larger population. For surveys based on registers, random imputation for qualitative variables has been suggested (Wallgren and Wallgren, 2007). Also Fiedler and Schodl (2008) applied random imputation for person's occupation and education in a test of a register based census. Multiple imputations in registers have been used by Abowd et al. (2006).

Honaker and King (2010) developed an approach to analyzing data with missing values that is suitable for large numbers of variables. This characteristic is common in multivariate data used in comparative politics and international relations; or when qualitative knowledge exists about specific missing cell values. Their method greatly increased the information researchers are able to extract from a given data. This study was neither interested in cross-sectional time series data nor in survey data. The study sought to obtain single estimates for missing values for data which is not cross-sectional.

### 2.6 Estimation of missing values for micro-array time series data

Dempster, et al. (1977) formalized the EM algorithm; a computational method for efficient estimation from incomplete data. Cao, et al. (2008) proposed a new method for estimating missing value in a micro-array data based on non-parametric regression combined with nearest neighbor approach, referred to as NPRA, which can capture both linear and non-linear relations between genes and arrays. They performed a comparative study of the imputation methods using different public datasets. The NPRA method produced more superior results than the other methods for different cases of missing value points and sizes of missing values. This study was also not based on micro-array time series nor did it extend the methods used in micro-array to bilinear time series.

### 2.7 Estimation of missing values for spatio-temporal time series

Multiple time series data that correspond to different spatial locations are referred to as the Spatio-temporal time series. One approach to analyzing spatial data with missing values was outlined by (Gomez, et al., 1995). It uses the bootstrap method to input missing natural resource inventory data. Yozgatligil, et al. (2012) compared several imputation methods used to compute the missing values of spatio–temporal meteorological time series. They assessed six imputation methods with respect to various statistical properties of the estimators such as accuracy, robustness, precision and efficiency for artificially created missing data in monthly total precipitation and mean temperature series obtained from the Turkish State Meteorological Service. These methods were classified as either simple or computational intensive. Simple arithmetic average, normal ratio (NR), and NR weighted with correlations comprised the simple ones. Multilayer perceptron type neural network and multiple imputation strategy adopted by Monte Carlo Markov Chain based on expectation–maximization (EM-MCMC) were classified as computationally intensive. They concluded that despite the computational inefficiency of EM-MCMC methods, they seem good for the imputation of meteorological time series which has several cases of missing values. Further, they concluded that using the EM-MCMC algorithm for imputing missing values before conducting any statistical analyses of meteorological data definitely decreases the amount of uncertainty and give more robust results.

A different approach from EM-MCMC that uses nonlinear and mixed integer mathematical programming (MINLP) models with binary variables for estimating missing values in precipitation data was developed and evaluated by Teegavarapu (2012). It overcomes the limitations associated with spatial interpolation methods relevant to the arbitrary selection of weighting parameters, namely the number of control points within a neighborhood and its size. Daily precipitation data obtained from 15 rain gauging stations were used to test and derive conclusions about the efficiency of these methods. The developed methods were compared with some other approaches namely, multiple linear regression, nonlinear least-square optimization, kriging, global and local trend surface and thin-plate spline functions. The new method of mathematical programming formulation gave superior estimates than to those obtained from all the other spatial interpolation methods.

Abdalla and Marwalla (2005) compared two algorithms for imputing missing values, namely the Expectation Maximization (EM) Algorithm and the Auto-Associative Neural Networks and Genetic Algorithms combination, using three datasets obtained from an industrial power plant, an industrial winding process and Human Immunodeficiency Virus (HIV) survey. Their results showed that Expectation Maximization Algorithm is appropriate and performs better in cases where the input variables are largely independent, whereas the auto-associative neural network and genetic algorithm combination is suitable when the variables in the model have an intrinsic nonlinear relationships.

In a different study, Toth, et al. (2000) compared the accuracy of the short-term rainfall forecast using three techniques: nearest neighbors, artificial neural networks, and auto-regressive moving average models. The performance of the nearest neighbor technique was investigated through a subjective trial-and-error process by varying the number of neighbors in the range [5, 100]. They observed that the performance of the forecast improved when the number of neighbors was increased; however, the improvement was insignificant when the numbers of neighbors were more than 20. They also found out that the results obtained by the nearest neighbors' method were better than those obtained from autoregressive moving average models. Based on quality of the performance of the approaches, the artificial neural networks gave the best results followed by k-nearest neighbor's method while autoregressive moving average models gave the worst results. Among the nonparametric methods used, ANN performed the best. This is one of the reasons that motivated the use of ANN in this study to determine how it performs with bilinear time series models.

**2.8 Missing value imputation for linear time series models with finite variance**

Damsleth (1979) developed a method for imputing missing values in a time series which can be represented as an ARIMA time series model based on computing the optimal linear combination of the forward and back forecasts. Another approach based on forecasting was developed by Abraham (1981) who used forecasting techniques to estimate missing observations in time series. He used the minimum mean squared error estimate to measure efficiency of the estimates. Missing values for linear processes with finite variance were also obtained by Miller and Ferreiro (1984). This was later extended by Ferreiro (1987) who discussed different alternatives methods for the estimation of missing observation in stationary ARIMA time series models. His article offered a series of alternatives techniques for estimating missing observations.

Smoothing methods based on state space formulation have also been used to estimate missing values and these are described in general terms (Anderson and Moore, 1979). Algorithms for computing the likelihood function when there are missing data in scalar case have been provided by Jones (1980), and Ljung (1982, 1993) and Harvey and Pierse (1984) for stationary models. They showed how to predict and interpolate missing observations and obtain the mean squared error of the estimate. Beveridge (1992) also extended the concept of using minimum mean squared error linear interpolator for missing values in time series to handle any pattern of non-consecutive observations. He applied the method to simple ARMA models to discuss the usefulness of either the non-parametric or the parametric form of the least squares interpolator.

State space representation has also been used for estimation of missing values in ARIMA models (Jones, 1985; Harvey and Pierse, 984). Harvey and Pierse (1984) further discussed maximum likelihood estimation of the parameters in an autoregressive integrated moving average (ARIMA) model when some of the observations are subject to temporal aggregation. They pointed that imputation problem can be solved by setting up the model in state space form and then applying the Kalman filter. Nieto and Martinez (1996) demonstrated a linear recursive technique that does not use the Kalman filter to estimate missing observations in an invertible ARIMA model. This procedure is based on the restricted forecasting approach, and the recursive linear estimators are obtained when the minimum mean-squared error is least.

The Kalman filter is a set of mathematical equations that recursively provides an efficient computational means to estimate the state of a process in a way that minimizes the mean squared error (Welch and Bishop, 2011). The filter is very powerful in several aspects: it supports estimation of past, present, and even future states, and can also do so even when the precise nature of the modeled system is unknown. These are concerned with finding the best linear estimates of the state vector $X_t$ in terms of the observations $Y_1, Y_2$ and a random vector $Y_0$. Recursive equations update the mean and covariance matrix and hence the distribution of the state vector, $Y_{t+1}$, after the new observation , $Z_{t+1}$, has become available. The update estimates $\hat{Y}_{t+1}$ of the state is the sum of projected estimates using observation at time t, and the one-step-ahead forecast error.

Thus kalman filtering is a recursive updating procedure that consists of forming a preliminary estimate of the state and then revising the estimate by adding a correction to this preliminary estimate. The ease of implementation of Kalman filter algorithm has now made it become widely used in many applications (Kohn and Ansley, 1983).

In practice, the kalman filter equations are more easily adapted to cope with missing values. When a missing observation is encountered at time t, the prediction equations are processed at the point based on the previous values. That is at every point in the time series, a prediction is made of the next value based on a few of the most recent estimates (Vijayakumar and Plale, 2007).

Pascal (2005) investigated influence of missing values on the prediction of a stationary time series process by applying Kaman filter fixed point smoothing algorithm. He developed simple bounds for the prediction error variance and asymptotic behavior for short and long memory process. Fung (2006) derived recursive smoothing methods associated with Kalman filter to estimates missing values and their mean squared error in ARMA models.

Pena and Tiao (1991) demonstrated that missing values in time series can be treated as unknown parameters and estimated by maximum likelihood method or as random variables and predicted by expected values. They provided examples to illustrate the difference between these two procedures. It is argued that the expected value is generally more suitable for estimating missing values in time series.

Norazián, et al. (2008) used interpolation and mean imputation techniques for simulated missing values from annual hourly air pollution data. They found out that the most appropriate imputation method was to replace each missing value with the mean of the two data points adjacent to the missing value. This approach is referred to as the mean-before-after method.

### 2.9 Imputing missing values for linear time series with infinite variance

Pourahmadi (1984) developed alternative techniques suitable for a limited set of stable cases with characteristic index $\alpha \in (1, 2]$. This was later extended to the ARMA stable process with characteristic index $\alpha \in (0,2]$ (Nassiuma, 1994). He developed an algorithm applicable to general linear and nonlinear processes by using the state space formulation and applied it to the estimation of missing values.

## 2.10 Missing value imputation for nonlinear time series models

Thavaneswaran and Abraham (1987) derived a recursive estimation procedure for estimating model parameters based on optimal estimating function. They applied this procedure to the estimation of missing observations. A more general method for estimation of missing values was developed by Abraham and Thavaneswaran (1991). They developed a general nonlinear time series model which included several standard nonlinear models such as GARCH and bilinear time series. They offered two methods for estimating missing observations based on prediction algorithm which included; the fixed point smoothing algorithm and estimating functions equations. It was used to recursively estimate missing observations in an autoregressive conditional heteroscedasticity (ARCH) model and the estimation of missing observations in a linear time series model. Bilinear model was considered as a special case. However, they only considered a particular model order, BL (1, 0, 2, 0) using estimating function approach. No simulation was done to assess its accuracy.

On vector time series, Luceno (1997) estimated missing values in possibly partially non-stationary vector time series. He extended Ljung (1989) method for estimating missing values and evaluating the corresponding function in scalar time series. The series is assumed to be generated by a possibly partially non-stationary and non-invertible vector autoregressive moving average process. He assumed no pattern of missing values. Future and past values were taken as special cases of missing data that can be estimated in the same way.

## 2.11 Nonparametric methods for estimating missing values

Nonparametric methods have also been proposed for estimating values. Titterington and Mill (1983) considered kernel estimation of a multivariate density for data with incomplete observations. When the parameter of interest is the mean of a response variable which is subject to missing values, Cheng (1994) proposed using the kernel conditional mean estimator. Hirano, et. al. (2003) studied the estimation of average treatment effects using non-parametrically estimated propensity scores. In survey statistics, Kim and Fuller (2004) proposed the fractional hot deck imputation method, in which multiple values are drawn from the same imputation cell as the missing observation, and a weight is assigned to each imputed value.

Many other approaches have been developed to deal with missing values, such as k-nearest neighbor (Troyanskaya, et al., 2001), Bayesian PCA (BPCA) (Oba. et al., 2003), least square imputation (LSimpute) (Hellem, 2004), local least squares imputation (LLSimpute) (Kim, et al., 2005) and least absolute deviation imputation (LADimpute) (Cao and Poh, 2006).

## 2.12 Estimating missing values using singular spectrum analysis

The principal component methods for multivariate data can be generalized to analyze time series data using a non-parametric approach called the Singular Spectrum Analysis (SSA). There are different SSA-based methods for filling in missing values in datasets (Schoellhamer, 2001; Kondrashov, et al., 2005; Golyandina and Osipov, 2006; Kondrashov, 2006). The motivation to use SSA is because it works well with arbitrary any statistical processes; whether linear or nonlinear, stationary or non-stationary, Gaussian or non-

Gaussian (Hassani, 2007). Musial, et al. (2011) compared the performance of some of the currently used approaches to fill gaps and smooth time series such as Smoothing Splines and Singular Spectrum Analysis in terms of either reconstructing the original record or in minimizing model selection criteria such as the Mean Absolute Error (MAE), Mean Bias Error (MBE) and chi-squared test statistics. They concluded that each method showed strengths and weaknesses, and that the choice of an approach largely depends on the properties of the underlying time series and the goal of the research.

SSA approach may be integrated with other methods in estimating missing value. Rodrigues, et al. (2001) proposed an imputation method to be used with singular spectrum-based techniques which is based on a weighted combination of the forecasts and hind-casts yield by the recurrent forecast method. They used it to estimate missing data in the total volume of passengers in a group of international airlines data (Box, et al., 2008). They observed that the method was easy to implement and the results obtained suggested an overall good performance. This method incorporates elements from a wide range of mathematical fields including classical time series analysis, multivariate statistics and geometry, dynamical systems, as well as signal processing (Golyandina, et al., 2001). It aims at describing the structure of the time series as a sum of trend, seasonal variations and noise. The workflow of the SSA gap-filling and smoothing algorithm proceeds in four phases.

The first phase in SSA iterative gap filling algorithm includes centering the original time series on zero by subtracting the mean value of all its elements and zeroing the missing data values (Musial, 2011, page 7905-7923). The inner loop of the SSA procedure which comprises decomposition, grouping and reconstructing is then performed first on this transformed time series. The missing values are replaced by computed values of the leading Empirical Orthogonal Functions (EOF) and on this basis the first estimate of the first constructed component is generated.

Missing values replaced by the first estimate are now replaced by the second estimate of the first leading component. The SSA gap filling algorithm is suitable for reconstructing time series with a highly harmonic oscillation shapes (Vautard, et al., 1992) or nonlinear trends (Ghil, et al., 2002). It is economical in the sense that a small number of SSA eigenmodes may be enough in the reconstruction of the original time series (Musial, 2011). This is an advantage over traditional spectral methods which require many trigonometric functions with different phases and amplitudes to provide a reliable estimate. On the other hand, the many steps in the computational requirements of the SSA gap-filling algorithm implementation are taken as weaknesses in estimating missing values involving a large number of time series. In addition, it has been noted that this method may not give good estimates when there are many missing values in time series (Kondrashov and Ghil, 2006). The SSA gap filling method can be extended to spatial-temporal data or to regenerate missing values in multivariate time series.

## 2.13 Artificial neural networks

Artificial Neural Networks (ANN) provides a rich, powerful and robust non-parametric modeling framework with proven and potential applications in many fields of the sciences (Popova, et al., 2014). Indeed, the network model is largely determined by the characteristics of the data. The advantage of neural networks is that they can flexibly model nonlinear relationships without any prior assumptions about the underlying data generation process (Qi, et al., 2001). These characteristics of ANNs have encouraged many researchers to use the neural network models in broad spectrum of real-world applications. Sometimes, the ANNs provide better alternatives than the other techniques for solving a variety of problems (Wenzel and Schröter, 2010; Pashova and Popova, 2011). Artificial neural network are in general, flexible nonlinear tools capable of approximating any sort of arbitrary function (Hornick et. al., 1989).

Modeling univariate time series using ANN is generally carried out using a certain number of lagged terms in the series as input and the forecasts as the output. Masters (1993) established that if there is a known seasonality in the data, then the number of seasons in that data can be used to identify the lags in the ANN model for forecasting.

There are several types of ANNs that are used in modeling. One of them is the multilayer perceptron which is widely used for modeling of nonlinear dependences (Rumelhart and Clelland, 1986). A multi-layer perceptron (MLP) model is made up of a layer of N input neurons, a layer of M output neurons and one or more hidden layers, although it has been shown that for most problems it would be enough to have only one layer of hidden neurons ( Hornick, et al, 1998). In this type of framework, the connections between neurons are always feed-forward, that is, the connections feed from the neurons in a certain layer towards the neuron in the next layer. According to Moreno (2011), the mathematical representation of the function applied by the hidden neurons in order to obtain an output $b_{pj}$ when faced with the representation of an input vector or pattern $X_{pi}$ : $X_{p1}$, $X_{p2}$, $X_{pi}$, $X_{pN}$ is given by

$$b_{pj} = f_L(\theta_j + \sum_{i=1}^{L} w_{ij} \bullet X_{pi})$$

where i=1,2,..,p, $f_L$ is the activation function of hidden layers L, $\theta_j$ is the threshold of hidden neuron j, $w_{ij}$ is the weight of the connection between input neuron i and hidden neuron j and $X_{pi}$ is the input signal received by input neuron i for pattern p. The output signal $\hat{y}_{pk}$ provided by output neuron k for pattern p, is given by

$$\hat{y}_{pk} = f_m(\theta_k + \sum_{j=1}^{L} v_{jk} b_{pj}),$$

where $f_m$ the activation function of output neuron m, $\theta_k$ is the threshold of hidden neuron k, $v_{jk}$ is the input signal received by input neuron j and output neuron k. In a general way, sigmoid function is used in the hidden layer neurons in order to give the neural network the capacity of learning the possible nonlinear function. MLP network training is carried using the application of gradient descent algorithm (Rumelhart, et al, 1986). This is shown figure 1 below.

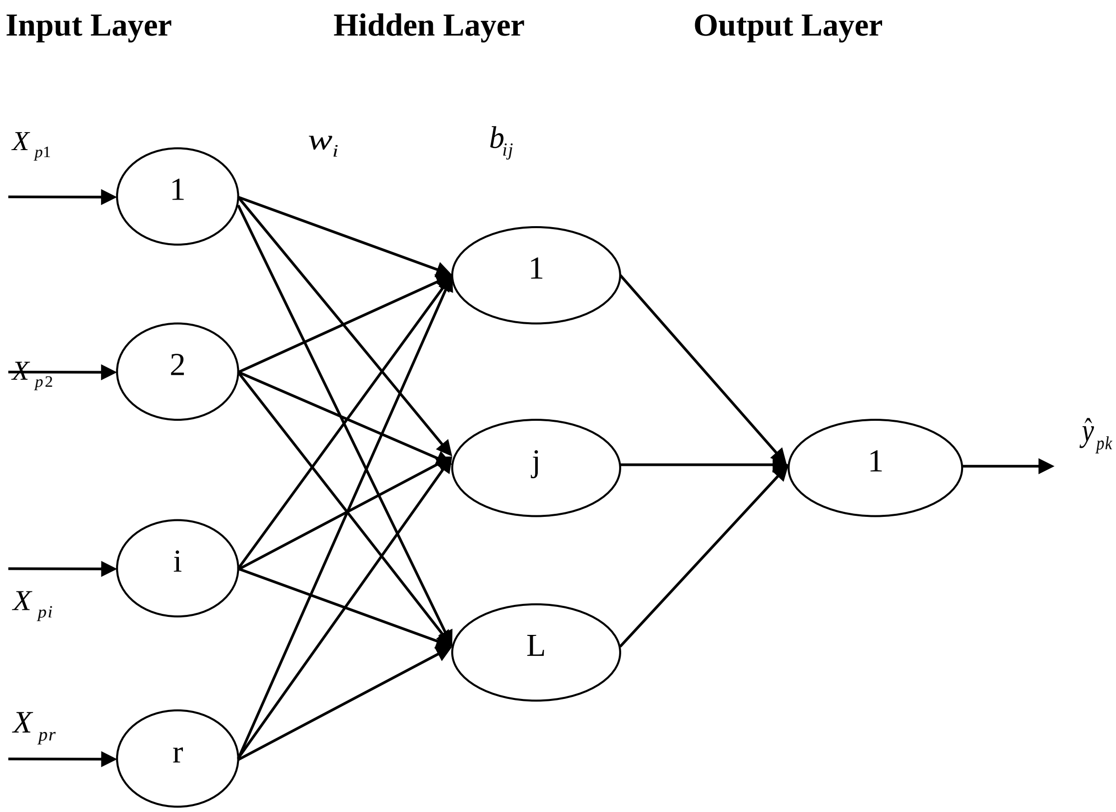



**Figure 1: Layers of Artificial Neural Network**

$\hat{y}_{pk}$ is used as an estimate of the missing value when we input the vector of the r-lagged values $(X_{p1}, X_{p2}, \ldots, X_{pr})$ in the neural network model developed.

The multilayer perceptron (MLP) model can be considered as a semi-parametric nonlinear function which relates the input data to the output data. It has been widely used to model complex relationships in data (Haykin, 1999). In time series modeling and atmospheric research, MLP is extensively used to capture the unknown relationships in data. It is also used in time series imputation researches due to the reported benefits (Junninen, et al., 2004).

### 2.13.1 Model fitting using ANN

Studies indicate that consideration of statistical principles in an ANN model building process may improve the model performance (Cheng and Titterington, 1994; Sarle, 1994). Consequently, it is important to adopt a procedure in the development of ANN models; taking into account issues such as data pre-processing, the determination of adequate model inputs and a suitable network architecture, parameter estimation (optimization) and model validation (Maier and Dandy, 1999b). In addition, careful selection of a number of internal model parameters is vital. The general function of these networks is given as

$$f(X,w) = \beta_0 + \sum_{h=1}^{H} \beta_h g\left( \gamma_0 + \sum_{i=1}^{I} \gamma_{hi} x_i \right)$$

where $X = [x_0, x_1, x_2, ..., x_n]$ is the vector of the lagged observations or inputs of the time series, and $w = (\beta, \gamma)$ are the network weights. I and H are the number of input and hidden units in the network and g(.) is a non-linear transfer function (Anders, et. al, 1998). How to select the input vector of a MLP and the number of hidden units in the hidden layer remains unresolved in research (Hornick, 1999).

### 2.13.2 Optimal architecture of ANN

Single hidden layer feed forward network is the most widely used model for time series modeling and forecasting (Zhang, et al., 1998). The model has three layers of simple processing units connected by acyclic links. These layers include input, hidden and the output layers. A MLP is trained using different number of hidden layers. It has been shown that ANNs with one hidden layer can approximate any function, given that sufficient degrees of freedom (i.e., connection weights) are provided (Hornik, et al., 1989).

However, in practice many functions are difficult to approximate with one hidden layer, requiring a large number of hidden layer nodes (Cheng and Titterington, 1994; Flood and Kartam, 1994). The uses of more than one hidden layer provide greater flexibility and enables approximation of complex functions with fewer connection weights in many situations (Flood and Kartam, 1994; Sarle, 1994; Tamura and Tateishi, 1997). Flood and Kartam (1994) suggest using two hidden layers as a starting point. However, it must be stressed that optimal network geometry is highly data dependent. The number of nodes in the input layer is fixed by the number of model inputs, whereas the number of nodes in the output layer equals the number of model outputs. The choice of number of hidden nodes, q, is subjective (Mehdi and Mehdi, 2010).

Another essential task of ANN modeling of time series is the selection of the number of lagged observations, denoted by p, the dimension of the input vector (Zhang, 2012). This is perhaps the most important parameter to be estimated in an ANN model because it plays a major role in determining the (nonlinear) autocorrelation structure of the time series.

In practice, simple network structure that has a small number of hidden nodes often works well in out-of-sample forecasting. This may be due to the over fitting effect typically found in the neural network modeling process. An over-fitted model has a good fit to the sample used for model building but has poor generalizability to out of the sample data (Demuth and Beale, 2004).

Although many different approaches exist in for finding the optimal architecture of an ANN, these methods are usually quite complex in nature and are not easy to implement (Zhang et al., 1998). Furthermore, none of these methods can guarantee the optimal solution for all real prediction problems. To date, there is no simple clear-cut method for the determination of these parameters and the usual procedure is to test numerous networks with varying numbers of input (p) and hidden (q). For each network, estimate generalization error. The network with the lowest generalization error is selected (Hosseini, et al., 2006).

### 2.13.3 Data pre-processing in ANN modeling

ANN models are no exception to the pre-processing of data (Kaastra and Boyd, 1995). Data pre-processing can have a significant effect on model performance. The available data should be divided into their respective subsets which include training, testing and validation prior to any data pre-processing (Burden, et al., 1997). Generally, different variables cover different ranges. In order to ensure that all variables receive equal weight during the training process, they should be standardized. In addition, the variables must be scaled in such a way as to be proportional to limits of the activation functions used in the output layer (Mills and Hall, 1996). For example, since the outputs of the logistic transfer function lie between 0 and 1, the data have to be generally scaled in the range (0.1–0.9) or (0.2–0.8). It should be noted that when the transfer functions in the output layer are not bounded (e.g. linear), scaling is not strictly required (Karunanithi, et al., 1994). However, scaling to uniform ranges is still recommended (Masters, 1993).

Another important issue to consider is stationarity of the data. Until recently, this has received very little attention in the development of ANN models. However, there are good reasons why the removal of deterministic components in the data (i.e. trends, variance, seasonal and cyclic components) should be considered (Masters, 1993). Two methods used in transforming a non-stationary model to a stationary model include differencing and logarithmic transformation techniques. Differencing has already been applied to neural network modeling of non-stationary time series (Chng, et al., 1996). However, use of the classical decomposition model may be preferable, as differenced time series can possess infinite variance (Irvine and Eberhardt, 1992).

### 2.13.4 Training in ANN modeling

The data presented to the neural networks are scaled in the range [0, 1]. All neural networks have a single output with an identity function. Gradient descent back-propagation is used for the training. The model parameters, learning rate, a cooling factor per epoch and momentum, are set. The momentum term may be helpful in preventing the learning process from being trapped into poor local minima, and is usually chosen in the interval [0:1]. Once a network structure (p, q) is specified, the network is ready for training; this is a process of parameter estimation that ensures the minimization of the mean square error on the test data. The mean squared error is evaluated in every epoch and the training proceeds until early stopping criterion is satisfied (Kourentzes and Crone, 2008). Finally, the estimated model is evaluated using a separate hold-out sample that has not been used in the training process. The network performance on the test set is a good indicator of its ability to generalize and handle data on which it has not been trained. If the performance on the test is poor, the network configuration or learning parameters can be changed. The network is then retrained until its performance is satisfactory. The test and train procedures involve training the network on most of the input data (around 70 % or more) and testing on the remaining data.

A major problem in training an ANN is deciding when to stop training. Since the ability to generalize is fundamental for these networks to predict future values, overtraining is a serious issue (Lawrence, 1997). Overtraining occurs when the system memorize patterns and thus loses the ability to generalize. Overtraining can occur by as a result of using too many hidden nodes or training for too many time periods (epochs). However, overtraining can be prevented by performing test and train procedures or cross-validation. Most studies suggest that the number of iterations during training should lie between 85 and 5000 iterations (Doboeck, 1994). He further claims that training is affected by many varied parameters and so it is difficult to determine a general value for the number maximum number of epochs. Most neural network software program provide default values for learning rate that typically work well (Önder, E., Bayır, F., Hepsen, A.,2013).

### 2.13.5 Performance measures

The predictive capabilities of the optimal linear estimates are compared with estimates obtained from artificial neural network (ANN) and exponential smoothing methods. The Mean Absolute Deviation (MAD) and Mean Squared Error (MSE) are computed and employed as performance indicators. These measures are given by

$$\text{MAD}=\frac{\sum_{t=1}^{n}|e_t|}{n} \tag{1}$$

and

$$\text{MSE}=\frac{\sum_{t=1}^{n}\left|e_t\right|}{n} \tag{2}$$

### 2.13.6 Validation in ANN modeling

Once the training (optimization) phase has been completed, the performance of the trained network has to be validated on an independent data set using the criteria chosen. It is important to ensure that the validation data should not have been used as part of the training process for whatever reasons. If the difference in the error obtained using the validation set is remarkably different from that obtained using the training data, it is likely that the two data sets are not representative of the same population or that the model has been over-fitted (Masters, 1993).

### 2.13.7 Empirical studies on missing values using artificial neural networks

Most of the studies done on missing values on artificial neural networks have been based on hydrological and meteorological time series data. Shukur and Lee (2015) claimed that wind-speed time series data is generally prone to missing values. They further noted that when the data is nonlinear, other methods such as K-nearest neigbour, kalman filter and linear interpolation may not be appropriate for estimating missing values. Therefore they proposed a hybrid of ANN and AR methods denoted by ANN-AR. Their tests showed that ANN-AR estimates give more accurate results for hydrological data. Abdalla and Marwala (2005) have used neural networks and genetic algorithms to approximate missing data in a database.

Other hybrid models that incorporate various artificial neural networks have also been used to specifically estimate missing stream-flow data (Elshorbagy, et al., 2002). Current advances in estimation techniques for predicting missing stream-flow data continues to incorporate basic ANN concepts (Ng, et al., 2009). Malek, et al. (2008) developed a data infilling model that utilizes the basic principles of artificial neural network (ANN) combined with the nearest neighbors imputation technique. Their results showed that the method proposed was robust enough to cope with vagaries due to varying sample sizes and extreme data insufficiency. Further, the study showed that ANN is superior in filling in missing values for hydrological data. Pachepsky and Yakov (2010) developed a model that incorporated artificial neural network for infilling missing values in time series meteorological data.

### 2.13.8 Estimation of missing values using exponential smoothing

Gupta and Srinivasan (2011) used exponential smoothing (EXP) method in estimating missing values for time series data on water flow. They reported that they obtained good results. Since time series data are noisy, the ARIMA models may not provide better estimates than those obtained from the nearest neighbors and cold deck methods. It has been found that the exponential smoothing with a constant $\alpha$=0.2 may produce better forecasts than those based on ARIMA models (Background Facts on Economic Statistics, 2013). Time series smoothers estimate the level of a time series at a given time as its conditional expectation given present, past and future observations, with the smoothed value depending on the estimated time series model (Ledolter, 2008).

Nassiuma and Thavaneswaran (1992) derived a recursive form of the exponentially smoothed estimates for a nonlinear model with irregularly observed data and discussed its asymptotic properties.

It is evident from the literature that imputation of missing values is a topic of interest to many researchers. This is reflected by the many studies done in this area. Further, different imputation methods have been developed for different types of time series models. Most of these are linear time series models. It is also evident that these methods have been based on criteria such as maximum likelihood method, estimating function and simple linear interpolation techniques. Other simple methods like mean of adjacent values have been suggested. However, what is lacking in the literature is an explicit method for estimating missing values for a class of nonlinear time series called bilinear time series models. The only case recorded so far is the estimation of missing values for a simple order BL (1,0 , 2 ,0). This method was based on estimating function. No estimates of missing values have been derived using the dispersion error. Further, no simulation study has been done on the performance of estimates obtained. These are important gaps in the literature that the study set to fill.

# CHAPTER THREE

# OPTIMAL LINEAR ESTIMATORS OF MISSING VALUES

## 3.1 Introduction

Optimal linear estimates of missing values for several simple, pure and general bilinear time series models whose innovations follow Gaussian, Student-t and GARCH distributions are derived by using minimum dispersion error criterion. Two assumptions were made in the process of the derivations; the first one is that the time series models used are stationary and thus their roots lie within the unit circle. Secondly, the higher powers (of orders greater than two or products of coefficients of orders greater than two) of the coefficients are approximately negligible. This is a consequence of the result of the first assumption. Further the innovations consisted of sequence of independent and identically distributed (i.i.d) random variables for the normal and t-distribution innovations.

## 3.2 The steps in deriving the optimal linear estimators

The steps followed in the derivation of the optimal estimates are as follows:

- The first step is to express the stationary bilinear time series model as a linear combination of the innovation sequence of the series.
- Then find the h-steps ahead forecast for the time series obtained in the first step.
- Obtain the h-steps-ahead forecast error.
- Square the forecast error and take its expectation. This is the dispersion error.
- Differentiate the dispersion error with respect to the coefficient $a_k$ to obtain the value of the coefficient that result in minimum dispersion error.

The optimal linear estimate $x_m^*$ for estimating the missing observation $x_m$ is by given

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) \tag{3}$$

where $\hat{x}_m$ is the estimate obtained from the model based on the previous observations of the data before the point m. The coefficients $a_k$ (k=1, 2,.., k-m) are to be estimated by minimizing the dispersion error (disp $x_m$) given by equation (3) (Nassiuma,1994).

### 3.3 Estimating missing values for bilinear models with normally distributed innovations

Pure bilinear time series models are models described by the bilinear parameter term only (Owili, Nassiuma and Orawo 2015c). The coefficients of autoregressive and moving average components are zero. We look at both the simple and the general pure bilinear time series models.

#### 3.3.1 Simple pure bilinear time series models with normally distributed innovations

The simplest pure bilinear time series model of order one, BL (0, 0, 1, 1) is of the form

$$x_t = b_{11} x_{t-1} e_{t-1} + e_t, \ \ where \ \ e_t \sim N(0,1) \tag{4}$$

The missing value estimate for this model is based on the following theorem 3.1.

**Theorem 3.1**

The optimal linear estimate for BL (0, 0, 1, 1) with normally distributed innovation is given by

$$
\begin{aligned}
x_m^* &= \hat{x}_m \\
&= E(x_t / x_{m-1}, x_{m-2}, .....x_1) \\
&= \hat{b}_{11} x_{m-1} \hat{e}_{m-1}
\end{aligned}
$$

**Proof**

Through recursive substitution of equation (4), the stationary BL (0, 0, 1, 1) is obtained

$$a\ x_t = \sum_{i=1}^{\infty} \left\{ \prod_{j=1}^{i} b_{11} e_{t-j} \right\} e_{t-i} + e_t$$

The h-steps ahead forecast is

$$x_{t+h} = \sum_{i=1}^{\infty} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{t+h-i} + e_{t+h}$$

Therefore the forecast error is

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-j} \right\} e_{t+h-i} + e_{t+h}\,. \tag{5}$$

Equation (5) can be expressed as

$$x_k - \hat{x}_k = \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-j} \right\} e_{k-i} + e_k\,. \tag{6}$$

Substituting equation (6) in equation (3), we have

$$
\begin{aligned}
disp\ \ x_m &= E(x_m - \hat{x}_m)^2 - 2E(x_m - \hat{x}_m) \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) + E\left\{ \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) \right\}^2 \\
&= E(\hat{e}_m{}^2) \quad - 2E(\hat{e}_m) \sum_{k=m+1}^{n} a_k \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{k-i} + e_k
\end{aligned}
$$

$$+E\left\{\sum_{k=m+1}^{n}a_k\sum_{k=m+1}^{n}a_k\sum_{i=1}^{h-1}\left\{\prod_{j=1}^{i}b_{11}e_{t+h-i}\right\}e_{k-i}+e_k\right\}^2 \tag{7}$$

Simplifying each of the terms of equation (7) separately, we obtain

$$\textit{first term on the right} \qquad E(x_m-\hat{x}_m)^2=\sigma^2$$

$$2E\hat{e}_m \bullet \sum_{k=m+1}^{n}a_k\sum_{i=0}^{k-m}\left(\prod_{j}^{i}b_{11}e_{k-j}\right)e_{k-i}$$

$$=2E\hat{e}_m \bullet \left[\begin{array}{l} a_{m+1}b_{11}e_m{}^2+a_{m+2}\left(b_{11}e_{m+1}{}^2+b_{11}{}^2e_{m+1}e_m{}^2\right) \\ +a_{m+3}\left(b_{11}e_{m+2}{}^2+b_{11}{}^2e_{m+2}e_{m+1}{}^2+b_{11}{}^3e_{m+2}e_{m+1}e_m{}^2\right) \\ +\ldots \end{array}\right]$$

$$=0$$

$$\textit{Second term} \quad E\left[2\hat{e}_m \bullet \sum_{k=m+1}^{n}a_k\sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i}b_{11}\hat{e}_{k-j}\right\}\hat{e}_{k-i}\right]$$

$$=2Ee_m \bullet a_{k-m}(b_{11}{}^{k-m}\hat{e}_{k-1}.\hat{e}_{k-2}\ldots\hat{e}_{k-m+1})$$

$$=2.Ee_m^2a_{k-m}b_{11}{}^{k-m}E\prod_{i=1}^{m-1}e_{k-i}$$

$$=0$$

$$\textit{Third term:} \quad E\left[\sum_{k=m+1}^{n}a_k\sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i}b_{11}\hat{e}_{t+h-j}\right\}\hat{e}_{t+h-i}+\hat{e}_{t+h}\right]^2$$

$$=E\left[\begin{array}{l} a_{m+1}(b_{11}\hat{e}_m{}^2+\hat{e}_{m+1}{}^2)+a_{m+2}\left(b_{11}e_{m+1}{}^2+b_{11}{}^2e_{m+1}e_m{}^2+\hat{e}_{m+2}{}^2\right)+ \\ a_{m+3}\left(b_{11}e_{m+2}{}^2+b_{11}{}^2e_{m+2}e_{m+1}{}^2+b_{11}{}^3e_{m+2}e_{m+1}e_m{}^2+e_{m+3}^2\right)+\ldots \end{array}\right]^2$$

$$=3\sigma^4b_{11}{}^2\sum_{k=m+1}^{n}a_k^2+\sum_{k=m+1}^{n}a_k^2\sigma^2$$

Hence equation (7) can be simplified as

$$disp \; x_m = \sigma^2 + 3\sigma^4 b_{11}{}^2 \sum_{k=m+1}^{n} a_k^2 + \sum_{k=m+1}^{n} a_k^2 \sigma^2 \tag{8}$$

Now differentiating equation (8) with respect to $a_k$ and equating to zero, we obtain

$$\begin{aligned} \frac{d}{da_k} disp\, x_m &= \frac{d}{da_k}\left\{ \sigma^2 + 3\sigma^4 b_{11}{}^2 \sum_{k=m+1}^{n} a_k^2 + \sum_{k=m+1}^{n} a_k^2 \sigma^2 \right\} = 0 \\ \Rightarrow \qquad & \hat{a}_k = 0 \end{aligned}$$

Substituting the values of $a_k$ in equation (3), we obtain optimal estimator of the missing value as

$$\begin{aligned} x_m^* &= \hat{x}_m \\ &= E(x_t / x_{m-1}, x_{m-2}, .....x_1) \\ &= \hat{b}_{11} x_{m-1} \hat{e}_{m-1} \end{aligned}$$

This result shows that the missing value is a one step-ahead prediction based on the past observations collected before the missing value. This is similar to the findings of Nassiuma (1994) which found $\hat{a}_k$ =0 for missing values of ARMA stable processes in some cases.

### 3.3.2 Estimating missing values for pure bilinear time series model with normal innovations

The pure bilinear time series model BL (0, 0, p, p) with normal innovations is given by

$$x_t = \sum_{i=1}^{p} b_{ii} x_{t-i} e_{t-i} + e_t, \qquad where \quad e_t \sim N(0,1) \tag{9}$$

The missing value estimate for this model is based on the following theorem 3.2.

**Theorem 3.2**

The optimal linear estimate of the missing value for the pure bilinear time series model BL (0, 0, p, p) is given by

$$x_m^* = \sum_{i=1}^{p} b_{ii} x_{t-i} e_{t-i} + e_t \qquad (10)$$

**Proof**

The stationary pure model is obtained by recursive substitution of equation (9), leading to

$$x_t = \sum_{s=1}^{p}\sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(b_{ss}e_{t-sj}\right)\right)e_{t-is} + e_t + \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} e_{t-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{t-s_1} + e_{t-s_2}\right)\right) + o(h) \qquad (11)$$

Setting t at t+h in equation (11), the h-steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(b_{ss}e_{t+h-sj}\right)\right)e_{t+h-is} + e_t$$

$$+\sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} e_{t+h-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{t+h-s_1} + e_{t+h-s_2}\right)\right) + o(h) \qquad (12)$$

Thus h-steps ahead forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{s=1}^{p}\sum_{i=1}^{h-1}\left(\prod_{j=1}^{i}\left(b_{ss}e_{t+h-sj}\right)\right)e_{t+h-is} + e_{t+h} + \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} e_{t+h-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{t+h-s_1} + e_{t+h-s_2}\right)\right)$$

or

$$x_k - \hat{x}_k = \sum_{s=1}^{p}\sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(b_{ss}e_{k-sj}\right)\right)e_{k-is} + e_k + \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} e_{k-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{k-s_1} + e_{k-s_2}\right)\right) \qquad (13)$$

Now substituting equation (13) in equation (3) and setting h-1=k-m, we get

$$
\begin{aligned}
disp\ x_m &= E(x_m - \hat{x}_m)^2 - 2E(x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k \sum_{s=1}^{p}\sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(b_{ss}e_{k-sj}\right)\right)e_{k-is} + e_k \\
&+ \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1}b_{s_2s_2}e_{k-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{k-s_1} + e_{k-s_2}\right)\right) + \\
&+ E\left\{\sum_{k=m+1}^{n} a_k \sum_{s=1}^{p}\sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(b_{ss}e_{k-sj}\right)\right)e_{k-is} + e_k + \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1}b_{s_2s_2}e_{k-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{k-s_1} + e_{k-s_2}\right)\right))\right\}^2
\end{aligned}
\tag{14}
$$

Substituting equation (14) in equation (3) we have

$$
\begin{aligned}
disp\ x_m &= E\left(\hat{e}_m{}^2\right) \\
&- 2E\hat{e}_m \bullet \left\{\begin{array}{l}\left(\sum_{k=m+1}^{n} a_k \sum_{s=1}^{p}\sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(b_{ss}e_{k-sj}\right)\right)e_{k-is} + \sum_{k=m+1}^{n} a_k e_k\right) + \\ \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1}b_{s_2s_2}e_{k-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{k-s_1} + e_{k-s_2}\right)\right)\end{array}\right\} \\
&+ E\left[\begin{array}{l}\left(\sum_{k=m+1}^{n} a_k \sum_{s=1}^{p}\sum_{i=1}^{k-m1}\left(\prod_{j=1}^{i}\left(b_{ss}e_{k-sj}\right)\right)e_{k-is} + \sum_{k=m+1}^{n} a_k e_k\right) + \\ \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1}b_{s_2s_2}e_{k-s_1-s_2}{}^2\left(\sum_{s_1=1,s_2>s_1}^{p}\left(e_{t+h-s_1} + e_{k-s_2}\right)\right)\end{array}\right]^2
\end{aligned}
\tag{15}
$$

Simplifying each term in equation (15), we obtain

$$First\quad term: E\left(\hat{e}_m{}^2\right) = \sigma^2$$

$$
\begin{aligned}
Second\ \ term:\quad & -E2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{s=1}^{p}\sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(b_{ss}e_{k-sj}\right)\right)e_{k-is} \\
& = \text{-2E}\hat{\text{e}}_{\text{m}} \bullet \sum_{\text{s}=1}^{p}(b_{ss}^{k-m}e_{k-s}e_{k-2s}....e_m) = 0
\end{aligned}
$$

$$Third\ \ term: \ E\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k e_k = 0$$

$$Fourth\ \ term \quad 2E\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1} \sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} e_{k-s_1-s_2}{}^2 \left( \sum_{s_1=1, s_2>s_1}^{p} \left( e_{k-s_1} + e_{k-s_2} \right) \right)$$

$$= +2 \bullet \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1} \sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} E\hat{e}_m \left( \sum_{s_1=1, s_2>s_1}^{p} \left( e_{k-s_1} + e_{k-s_2} \right) \right)$$

$$= 2 \bullet \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1} \sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} \sigma^2 E\hat{e}_m \left( \sum_{s_1=1, s_2>s_1}^{p} \left( e_{k-s_1} + e_{k-s_2} \right) \right)$$

$$= \begin{cases} 2 \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1} \sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} \sigma^4 & , for \ \ m = k - s_1 \ or \ \ m = k - s_2 \\ 0 & , otherwise \end{cases}$$

$$Fifth\ term : E\left( \begin{array}{l} \left( \sum_{k=m+1}^{n} a_k \sum_{s=1}^{p} \sum_{i=1}^{k-m} \left( \prod_{j=1}^{i} \left( b_{ss} e_{k-sj} \right) \right) e_{k-is} + \sum_{k=m+1}^{n} a_k e_k \right) + \\ \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1} \sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} e_{k-s_1-s_2}{}^2 \left( \sum_{s_1=1, s_2>s_1}^{p} \left( e_{k-s_1} + e_{k-s_2} \right) \right) \end{array} \right)^2$$

$$= \sum_{k=m+1}^{n} a_k^2 (3\sigma^4 b_{ss}{}^2) + \sum_{k=m+1}^{n} a_k^2 \sigma^2$$

Therefore equation (15) becomes

$$Disp\ \ x_m = \begin{cases} \sigma^2 + \sum_{k=m+1}^{n} a_k^2 (3\sigma^4 b_{ss}{}^2) + \sum_{k=m+1}^{n} a_k^2 \sigma^2 & for \ \ s_1 \neq k - m, \ s_2 \neq k - m \\ \sigma^2 - 2 \sum_{k=m+1}^{n} a_k \sum_{s_1=1}^{p-1} \sum_{s_2>s_1}^{p} b_{s_1s_1} b_{s_2s_2} \sigma^4 + \sigma^2 \sum_{k=m+1}^{n} a_k^2 (3\sigma^2 b_{ss}{}^2 + 1) & \end{cases} \tag{16}$$

Differentiating equation (16) with respect to $a_k$ and setting the result to zero we get,

$$
dispx_m = \frac{d}{da_k}\begin{cases}(\sum_{k=m+1}^{n} 2a_k(3\sigma^2 b_{ss}^2) + 2\sum_{k=m+1}^{n} a_k\sigma^2 = 0 & for\ s_1 \neq k-m, s_2 \neq k-m \\ -2\sum_{k=m+1}^{n}\sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p}(b_{s_1s_1}b_{s_2s_2}\sigma^4 + 2\sigma^2\sum_{s=1}^{p} a_k(3\sigma^2 b_{ss}^2+1) = 0 & otherwise\end{cases} \qquad (17)
$$

Hence from equation (17), we have

$$
\hat{a}_k = \begin{cases}\hat{a}_k = 0 \quad for \quad for\ s_1 \neq k-m, s_2 \neq k-m \\ \dfrac{\sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p}(b_{s_1s_1}b_{s_2s_2}\sigma^2}{(3\sigma^2 b_{ss}^2+1)} \quad for\ s_1 \neq k-m, s_2 \neq k-m\end{cases}
$$

These results indicate that the observed values after the missing point may play a role in the estimation of missing value(s). In most cases there is a weight assigned to these values; values near the missing values are assigned higher weights. Therefore the optimal linear estimator of a pure bilinear time series model is

$$
x_m^* = \sum_{i=1}^{p}\hat{b}_{ii}x_{t-i}\hat{e}_{t-i} \qquad for \quad s_1 \neq k-m,\ s_2 \neq k-m
$$

or

$$
x_m^* = \sum_{i=1}^{p}\hat{b}_{ii}x_{t-i}\hat{e}_{t-i} + \sum_{k=m+1}^{n}\frac{\sum_{s_1=1}^{p-1}\sum_{s_2>s_1}^{p} b_{s_1s_1}b_{s_2s_2}\sigma^2}{\sum_{s=1}^{p}(3\sigma^2 b_{ss}{}^2+1)}(x_k - \hat{x}_k), \quad for \quad s_1, s_2 = k-m
$$

(18)

**Corollary**

i) For the case $b_{s_2s_2} = 0\ \ or\ b_{s_1s_1} = 0$ then $\hat{a}_k = 0$

ii) When k-m=0, it means the missing data point is the last data. For this case, $\hat{a}_k = 0$.

iii) When k-m=1, it means the missing data point is the second last data. For this case

Therefore k-m cannot be equal to $s_2s_2$. Therefore, $\hat{a}_k = 0$.

iv) For most cases, k-m cannot be equal to $s_1$ or $s_2$ and therefore for pure bilinear time series, we can generally say that $\hat{a}_k = 0$.This means that for all pure bilinear models, we have $\hat{a}_k = 0$.

### 3.3.3 Estimating missing values for BL (1, 0, 1, 1) with normal innovations

The stationary bilinear time series model of order BL (1, 0, 1, 1) is given by

$$x_t = \phi_1 x_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t, \quad e_t \sim N(0,1) \tag{19}$$

The optimal linear estimates for this model is obtained using theorem 3.3.

**Theorem 3.3**

The optimal linear estimate for the bilinear time series model with normal errors, BL (1, 0, 1, 1), is given by

$$x_m^* = \hat{\phi}_1 x_{m-1} + \hat{b}_{11} x_{m-1} \hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}^{(k-m)}}{(1+\hat{\phi}_1^{\,2} + 3\hat{b}_{11}^{\,2}\hat{\sigma}^2)} (x_k - \hat{x}_k)$$

**Proof**

Performing recursive substitution of equation (19), the stationary BL (1, 0, 1, 1) can be expressed as

$$x_t = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{t-j}\right)\right)e_{t-i} + e_t . \tag{20}$$

The h-steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h} \tag{21}$$

and the h-steps ahead forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h} \tag{22}$$

substituting equation (22) in equation (3) we have

$$\begin{aligned} disp\ x_m &= E\left(\hat{e}_m{}^2\right) - 2E\left\{\hat{e}_m \bullet \left(\sum_{k=m+1}^{n}a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n}a_k e_k\right)\right\} + \\ &+ E\left(\sum_{k=m+1}^{n}a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n}a_k e_k\right)^2 \end{aligned} \tag{23}$$

Simplifying each of the terms on the RHS of equation (18), we obtain

First term: $E\left(\hat{e}_m{}^2\right) = \sigma^2$

$$\text{Second term}: -\left\{2\hat{e}_m \bullet \left( \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{k-j}\right)\right) e_{k-i} + \sum_{k=m+1}^{n} a_k e_k \right)\right\}$$

$$= -E\hat{e}_m \bullet 2 \left[ \begin{array}{l} a_{m+1}(\phi_1 + b_{11}e_m)e_m + a_{m+1}e_{m+1} + a_{m+2}(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_{m+2})e_m + a_{m+2}e_{m+2} \\ + a_{m+3}(\phi_1 + b_{11}e_{m+2})(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_m)e_m + a_{m+3}e_{m+3} + \ldots) \end{array} \right]$$

$$= -2 \sum_{k=m+1}^{n} a_k \phi_1^{\,k-m}$$

$$\text{Third term}: \quad E\left( \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{k-j}\right)\right) e_{k-i} + \sum_{k=m+1}^{n} a_k e_k \right)^2$$

$$= E\left\{ \left( \sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right)^2 e_{k-j}^{\;\;2} + 2\left( \sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right) e_{k-j} \sum_{k=m+1}^{n} a_k e_k + \left( \sum_{k=m+1}^{n} a_k e_k \right)^2 \right\}$$

$$= \left\{ E\left( \sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right)^2 e_{k-j}^{\;\;2} + E2\left( \sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right) e_{k-j} \sum_{k=m+1}^{n} a_k e_k + E\left( \sum_{k=m+1}^{n} a_k e_k \right)^2 \right\} \quad (24)$$

Equation (24) is simplified as follows:

$$E2\left( \sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right) e_{k-j} \sum_{k=m+1}^{n} a_k e_k = 0$$

$$E\left( \sum_{k=m+1}^{n} a_k e_k \right)^2 = \sum_{k=m+1}^{n} a_k^{\;2} \sigma^2$$

$$E\left( \sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right)^2 e_{k-j}^{\;\;2} = \sum_{k=m+1}^{n} a_k^2 (\phi_1^{\;2} \sigma^2 + 3\sigma^4 b_{11}^{\;\;2})$$

Hence equation (24) becomes

$$dispx_m = \sigma^2 - 2\sum_{k=m+1}^{n} a_k \phi^{k-m} \sigma^{22} + \sum_{k=m+1}^{n} a_k^2 (\phi_1^{\;2} \sigma^2 + 3\sigma^4 b_{11}^{\;\;2}) + \sum_{k=m+1}^{n} a_k^2 \sigma^2 \qquad (25)$$

Differentiating equation (25) with respect to the coefficient $a_k$, we get

$$\frac{d}{da_k}\left[\sigma^2 - 2\sigma^2 \sum_{k=m+1}^{n} a_k b_{11}{}^{(k-m)} + \sum_{k=m+1}^{n} a_k{}^2(\phi_1{}^2 + b_{11}{}^2 3\sigma^2 + 1)\sigma^2\right] = 0$$

$$\Rightarrow \quad 0 - 2\phi^{(k-m)})\sigma^2 + 2a_k(\phi_1{}^2 + 3b_{11}{}^2\sigma^2 + 1)\sigma^2 = 0$$

Therefore we have

$$\hat{a}_k = \frac{\hat{\phi}_1{}^{(k-m)}}{(1 + \hat{\phi}_1{}^2 + \hat{b}_{11}{}^2 3\hat{\sigma}^2)}$$

The optimal linear estimate of $x_m$, *given by* $x_m^*$ that minimizes the error dispersion error is thus given as

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

$$= \hat{x}_m + \sum_{k=m+1}^{n} \frac{\phi^{(k-m)}}{(1 + \phi_1{}^2 + 3b_{11}{}^2\hat{\sigma}^2)}(x_{m+1} - \hat{x}_{m+1})$$

$$= \hat{\phi}_1 x_{m-1} + \hat{b}_{11} x_{m-1}\hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}^{(k-m)}}{(1 + \hat{\phi}_1{}^2 + 3\hat{b}_{11}{}^2\hat{\sigma}^2)}(x_k - \hat{x}_k)$$

### 3.3.4 Estimating missing values for BL (0, 1, 1, 1) with normal innovations

The bilinear time series model BL (0, 1, 1, 1) is given by

$$x_t = b_{11} x_{t-1} e_{t-1} + \theta e_{t-1} + e_t, \ \textit{where}\ e_t \sim N(0,1) \qquad (26)$$

Theorem 3.4 can is used to estimate the missing value for BL (0, 1, 1, 1) with normally distributed innovations.

**Theorem 3.4**

The optimal linear estimate of a missing value for BL (0, 1, 1,1) is given by

$$\hat{x}_m = \hat{b}_{11} x_{m-1} \hat{e}_{t-1} + \theta e_{m-1}$$

**Proof**

The stationary bilinear BL (0, 1, 1, 1) is expressed as

$$x_t = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} \theta e_{t-j}\right) e_{t-i-1} + \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} b_{11} e_{t-j}\right) e_{t-j} + e_t \qquad (27)$$

The h-steps ahead forecast based on equation (27) is given by

$$x_{t+h} = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} e_{t-j}\right) e_{t+h-i-1} + \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} b_{11} e_{t+h-j}\right) e_{t+h-j} + e_{t+h}$$

and the forecast error is expressed as

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1}\left(\theta \prod_{j=1}^{i} e_{t-j}\right) e_{t+h-i-1} + \sum_{i=1}^{h-1}\left(\prod_{j=1}^{i} b_{11} e_{t+h-j}\right) e_{t+h-j} + e_{t+h} \qquad (28)$$

Substituting equation (28) in equation (3), we have

$$disp\ \ x_m = E\left\{(\hat{e}_m{}^2) - \sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\theta\prod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]\right\}^2$$

$$= \left\{\begin{array}{l} E(\hat{e}_m)^2 - E\ \ 2(\hat{e}_m)\bullet \sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\theta\prod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right] + \\ E\left\{\sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\theta\prod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]\right\}^2 \end{array}\right\}$$

Simplifying each of the terms in equation (28), we obtain the following:

$$E(\hat{e}_m)^2 = \hat{\sigma}^2$$

$$E\ \ 2(\hat{e}_m)\bullet \sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]$$

$$= E\ \ 2(\hat{e}_m)\bullet\left(a_{m+1}\left\{\theta e_m e_{m-1} + b_{11}e_m{}^2 + e_{m+1}\right\}\right) + a_{m+2}\left\{\theta e_{m+1}e_m e_{m-1} + b_{11}{}^2 e_{m+1}e_m{}^2 + e_{m+2}\right\}$$

$$+ a_{m+3}\left\{\theta e_{m+2}e_{m+1}e_m + b_{11}{}^3 e_{m+2}e_{m+1}e_m{}^2 + e_{m+3}\right\} + a_{m+4}\left\{\theta e_{m+3}e_{m+2}e_{m+1}e_m e_{m-1} + b_{11}{}^3 e_{m+3}e_{m+2}e_{m+1}e_m{}^2 + e_{m+4}\right\}$$

$$= 0$$

*This implies that* $\hat{a}_k = 0$. *Therefore the best linear estimate is by*

$$\hat{x}_m = \hat{b}_{11}x_{m-1}\hat{e}_{t-1} + \hat{\theta}\hat{e}_{m-1}$$

### 3.3.5 Estimating missing values for BL (1, 1, 1, 1) with normal innovations

The bilinear time series model of BL (1, 1, 1, 1) is given by

$$x_t = \phi_1 x_{t-1} + b_{11}x_{t-1}e_{t-1} + \theta e_{t-1} + e_t,\ \ where\ e_t \sim N(0,1) \qquad (29)$$

The missing value estimate for the BL (1, 1, 1, 1) model is based on the following theorem 3.5.

**Theorem 3.5**

The optimal linear estimate for BL (1, 1, 1, 1) is given by

$$x_m^* = \hat{\phi}_1 x_{m-1} + \theta e_{t-1} + \hat{b}_{11} x_{m-1}\hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}^{(k-m)}}{(\hat{\theta}^2 + 1)(1+\hat{\phi}_1{}^2 + 3b_{11}{}^2\hat{\sigma}^2)}(x_k - \hat{x}_k)$$

**Proof**

Thus the stationary BL (1, 1, 1, 1) can be expressed as

$$x_t = \sum_{i=1}^{\infty}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{t-j}\right](\theta e_{t-i-1} + e_{t-i}) + (\theta e_{t-1} + e_t) \tag{30}$$

and h-steps ahead is forecast based on equation (30) is

$$x_{t+h} = \sum_{i=1}^{\infty}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{t+h-j}\right](\theta e_{t+h-i-1} + e_{t+h-i}) + (\theta e_{t+h-1} + e_{t+h})$$

The h-steps ahead forecast error

$$\begin{aligned} x_{t+h} - \hat{x}_{t+h} &= \sum_{i=1}^{h-1}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{t+h-j}\right](\theta e_{t+h-i-1} + e_{t+h-i}) + (\theta e_{t+h-1} + e_{t+h}) \\ &= \sum_{i=1}^{h-1}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{k-j}\right](\theta e_{k-i-1} + e_{k-i}) + (\theta e_{k-1} + e_k) \end{aligned}$$

The dispersion error is

$$E(x_m - x_m^*)^2 = E\left[(x_m - \hat{x}_m) - \sum_{k=m+1}^{n} a_k \sum_{i=1}^{h-1}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{k-j}\right](\theta e_{k-i-1} + e_{k-i}) + (\theta e_{k-1} + e_k)\right]^2$$

$$= E(\hat{e}_m{}^2) - E2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \left\{\sum_{i=1}^{h-1}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{k-j}\right](\theta e_{k-i-1} + e_{k-i}) + (\theta e_{k-1} + e_k)\right\}$$

$$+ E\left\{\sum_{k=m+1}^{n} a_k \sum_{i=1}^{h-1}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{k-j}\right](\theta e_{k-i-1} \quad + e_{t+h-i}) + (\theta e_{k-1} + e_k)\right\}^2 + E\sum_{k=m+1}^{n}(\theta e_{k-1} + e_k)^2 \qquad (31)$$

Simplifying the terms in equation (31) we obtain,

$$E(\hat{e}_m{}^2) = \sigma^2$$

$$E2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{k-j}\right](\theta e_{k-i-1} \quad + e_{k-i}) + \sum_{k=m+1}^{n} a_k(\theta e_{k-1} + e_k) +$$

$$= 2a_k \phi_1{}^{k-m}\sigma^2$$

$$E \bullet \left\{\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\prod_{j=1}^{i}\left[\phi_1 + b_{11}e_{k-j}\right](\theta e_{k-i-1} + e_{k-i}) + (\theta e_{k-1} + e_k)\right\}^2$$

$$= \sum_{k=m+1}^{n} a_k{}^2(\phi_1^2 + 3b_{11}{}^2\sigma^2)(\theta^2 + 1)\sigma^2 + \sum_{k=m+1}^{n} a_k{}^2(\theta^2 + 1)\sigma^2$$

$$= (\theta^2 + 1)\sigma^2 \sum_{k=m+1}^{n} a_k{}^2\left\{\phi_1^2 + 3b_{11}{}^2\sigma^2 + 1\right\}$$

$$Dispx_m = \sigma^2 - 2a_k\phi_1{}^{k-m}\sigma^2 + (\theta^2 + 1)\sigma^2 \sum_{k=m+1}^{n} a_k{}^2\left\{\phi_1^2 + 3b_{11}{}^2\sigma^2 + 1\right\}$$

$$\frac{dDisp}{da_k} = -2\phi_1{}^{k-m}\sigma^2 + 2a_k(\theta^2 + 1)\sigma^2 \sum_{k=m+1}^{n} a_k{}^2\left\{\phi_1^2 + 3b_{11}{}^2\sigma^2 + 1\right\} = 0$$

$$a_k = \frac{\hat{\phi}_1{}^{k-m}}{(\hat{\theta}^2 + 1)(\hat{\phi}_1^2 + 3\hat{b}_{11}\hat{\sigma}^2 + 1)\hat{\sigma}^2}$$

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

$$= \hat{x}_m + \sum_{k=m+1}^{n} \frac{\phi^{(k-m)}}{(\theta^2+1)(1+\phi_1{}^2+3b_{11}{}^2\hat{\sigma}^2)}(x_{m+1} - \hat{x}_{m+1})$$

$$= \hat{\phi}_1 x_{m-1} + \theta e_{t-1} + \hat{b}_{11} x_{m-1} \hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}^{(k-m)}}{(\hat{\theta}^2+1)(1+\hat{\phi}_1{}^2+3b_{11}{}^2\hat{\sigma}^2)}(x_k - \hat{x}_k)$$

### 3.3.6 Estimating missing values for BL (p, 0, p, p) with normal innovation

The bilinear time series model of BL (p, 0, p, p) is given by

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{i=1}^{p} \sum_{i=1}^{p} b_{i1} x_{t-i} e_{t-i} + e_t,\ e_t \sim N(0,1) \tag{32}$$

The missing value estimate is based on the following theorem 3.6.

**Theorem 3.6**

The optimal linear estimate for one missing value $x_m$ for the general bilinear time series model

BL (p, 0, p, p) is given by

$$x_m^* = \sum_{i=1}^{p} \hat{\phi}_i x_{t-i} + \sum_{i=1}^{p}\sum_{j=1}^{p} \hat{b}_{ij} x_{t-i} \hat{e}_{t-j} + \sum_{k=m+1}^{n} \frac{\sum_{K=m+1}^{n} \hat{\phi}_s{}^{k-m} + 2\sum_{s\neq r}^{p} \hat{\phi}_s \hat{\phi}_r}{\left\{\sum_{s=1}^{p} \hat{\phi}_s^2 + \sum_{j=1,s\neq j}^{p} \hat{b}_{sj}{}^2 + \sum_{j=1}^{p} 3\hat{b}_{jj}{}^2 \sigma^2\right\}}(x_k - \hat{x}_k)$$

**Proof**

Through recursive substitution of equation (24), the stationary bilinear time series model

BL (p, 0, p, p) is expressed as

$$x_t = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}e_{t-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{t-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}e_{t-s-r} + e_t \quad (32)$$

The h-steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t+h-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{t+h-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t+h-s-r} + e_t$$

or

$$x_k = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + e_t \quad (33)$$

and the  forecast is

$$x_k = \sum_{i=1}^{p}\sum_{i=1}^{k-1}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + e_t \quad (34)$$

The forecast error is

$$x_k - \hat{x}_k = \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r}$$

$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} + e_t$$

$$+ \sum_{i=0}^{\infty}\prod_{j=1}^{i}(\phi_2 + b_{21} e_{k-2j+1} + b_{22} e_{k-2j})\left(\theta_2 e_{k-2i-2} + \theta_1 e_{k-2i-1} + e_{k-2i}\right) + \left\{\theta_2 e_{k-2} + \theta_1 e_{k-1} + e_k\right\} \quad (35)$$

Substituting in equation (35) in equation (3), we obtain

$$disp\ x_m = E(x_m - \hat{x}_m)^2 - 2E(x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) + E\left\{\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)\right\}^2 \qquad (36)$$

The terms on the RHS side of equation (36) are simplified as follows

$$\text{First term} \qquad E(x_m - \hat{x}_m)^2 = \sigma^2$$

$$\text{Second term} \qquad -2E(x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) =$$

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \left\{\begin{aligned} &\sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\} e_{t-sj} \\ &+ \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\} e_{t+h-s-r} \\ &+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{t+h-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\} e_{t+h-s-r} + e_t \end{aligned}\right\}$$

$$= -2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-si} = -2\sum_{k=m+1}^{n} \phi^{(k-m)/s}\sigma^2$$

Now

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k$$

$$\sum_{s<r \ and \ s>}^{p} \sum_{r}^{p} \left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} = \begin{cases} \sum_{k=m+1}^{n} 4 a_k \phi_r \phi_s \sigma^2 & for \ s+r=k-m \\ 0 & otherwise \end{cases}$$

Third term

$$E\left\{\begin{array}{l} \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} + \\ \qquad \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} + e_t \end{array}\right\}^2$$

$$= \sum_{k=m+1}^{n} a_k^2 \left\{\sum_{s=1}^{p} \phi_s^2 \sigma^2 + \sum_{j=1, s\neq j}^{p} b_{sj}{}^2 \sigma^2 + e_{k-j} + \sum_{j=1}^{p} 3 b_{jj}{}^2 \sigma^4\right\}$$

which simplifies to

$$-2\sigma^2 \sum_{k=m+1}^{n} \phi_s^{(k-m)/s} + \sum_{k=m+1}^{n} 4 a_k \phi_r \phi_s \sigma^2 + \sum_{k=m+1}^{n} a_k^2 \left\{\sum_{s=1}^{p} \phi_s^2 \sigma^2 + \sum_{j=1, s\neq j}^{p} b_{sj}{}^2 \sigma^2 + e_{k-j} + \sum_{j=1}^{p} 3 b_{jj}{}^2 \sigma^4\right\} = 0 \qquad (37)$$

Differentiating equation (37) with respect to $a_k$ and setting to zero, we obtain

$$\frac{d}{da_k} disp = -2\sigma^2 \sum_{k=m+1}^{n} \phi_s^{(k-m)/s} + 4\phi_r \phi_s \sigma^2 +$$

$$2 a_k \left\{\sum_{s=1}^{p} \phi_s^2 \sigma^2 + \sum_{j=1, s\neq j}^{p} b_{sj}{}^2 \sigma^2 + e_{k-j} + \sum_{j=1}^{p} 3 b_{jj}{}^2 \sigma^4\right\} = 0 \quad (38)$$

Solving equation (38) for $a_k$, we get

$$\hat{a}_k = \frac{\sum_{K=m+1}^{n} \hat{\phi}_s^{\,k-m} + 2\sum_{s\neq r}^{p} \hat{\phi}_s \hat{\phi}_r}{\left\{ \sum_{s=1}^{p} \hat{\phi}_s^2 + \sum_{j=1, s\neq j}^{p} \hat{b}_{sj}^{\,2} + \sum_{j=1}^{p} 3\hat{b}_{jj}^{\,2} \sigma^2 \right\}}$$

The optimal linear estimate is therefore given by

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

$$= \sum_{i=1}^{p} \hat{\phi}_i x_{t-i} + \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{b}_{ij} x_{t-i} \hat{e}_{t-j} + \sum_{k=m+1}^{n} \frac{\sum_{K=m+1}^{n} \hat{\phi}_s^{\,(k-m)/s} + 2\sum_{s\neq r}^{p} \hat{\phi}_s \hat{\phi}_r}{\left\{ \sum_{s=1}^{p} \hat{\phi}_s^2 + \sum_{j=1, s\neq j}^{p} \hat{b}_{sj}^{\,2} + \sum_{j=1}^{p} 3\hat{b}_{jj}^{\,2} \sigma^2 \right\}} (x_k - \hat{x}_k)$$

**Corollary**

For p=1, we have the bilinear model BL (1, 0, 1, 1). The optimal linear estimate is given by

$$x_m^* = \hat{\phi}_1 x_{m-1} + \hat{b}_{11} x_{m-1} \hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}^{(k-m)}}{(1 + \hat{\phi}_1^{\,2} + 3\hat{b}_{11}^{\,2} \hat{\sigma}^2)} (x_k - \hat{x}_k)$$

### 3.3.7 Estimation of two missing values for bilinear time series with normal innovations

We consider the situation where two consecutive observations are missing in the data. The case of non-consecutive missing values can easily be obtained using the previous results and so it is not necessary to be discussed here. Suppose two consecutive values $\xi_k$ and $\xi_{k+1}$ are missing. Now we could use a vector form such that $\xi_m = (\xi_k, \xi_{k+1})^T$ while the matrix with the coefficient expressed as a matrix of the form

$$a_s = \begin{bmatrix} a_k & 0 \\ 0 & a_{k+1} \end{bmatrix}$$

We can rewrite equation (3) in vector form as

$$\xi_m^* = \hat{\hat{\xi}}_m + \sum_{s=k+1}^{n} a_s(\xi_s - \hat{\hat{\xi}}_s)$$

The innovation for the missing value would now be a column vector given by

$$\xi_m - \xi_m^* = (\xi_m - \hat{\hat{\xi}}_m) - \sum_{S=k+2}^{n} a_s(\xi_s - \hat{\hat{\xi}}_s),$$

*where* $\xi_m - \hat{\hat{\xi}}_m$ *is a column vector. The dispersion is given by*

$$= E(\hat{e}_m \hat{e}_m^{\,T}) - 2\hat{e}_m \bullet \sum_{S=k+2}^{n} \left\{(\xi_s - \hat{\hat{\xi}}_s)^T a_s\right\} + \sum_{S=k+2}^{n} a_s(\xi_s - \hat{\hat{\xi}}_s)\left\{\sum_{S=k+2}^{n} (\xi_s - \hat{\hat{\xi}}_s)^T \{a_s\}\right\}. \quad (39)$$

*and*

$$E(\xi_m - \xi_m^*)(\xi_m - \xi_m^*)^T = E\left\{(\xi_m - \hat{\hat{\xi}}_m) - \sum_{S=k+2}^{n} a_s a_s(\xi_s - \hat{\hat{\xi}}_s)\right\}\left\{(\xi_m - \hat{\hat{\xi}}_m) - \sum_{S=k+2}^{n} a_s(\xi_s - \hat{\hat{\xi}}_s)\right\}^T.$$

### 3.3.8 Estimating two missing values for BL (1, 0, 1, 1) with normal innovations

The stationary bilinear time series model of order BL (1, 0, 1, 1) is given by

$$x_t = \phi_1 x_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t, \quad e_t \sim N(0,1) \quad (40)$$

The estimators for two missing values for BL (1, 0, 1, 1) with normal errors is given in theorem 3.7.

**Theorem 3.7**

The estimator of two missing values for BL (1, 0, 1, 1) is given by

$$\xi_m^* = \hat{\xi}_m + \sum_{s=k+1}^{n} a_s(\xi_s - \hat{\xi}_s)$$

$$\xi_m^* = \hat{\xi}_m + \sum_{s=k+2}^{n} I \bullet \left[ \frac{\left\{\hat{\phi}^{s-m}\right\}}{\left\{\phi_1^2 * I + 3b_{11}^2 \Sigma + I\right\}} \right]$$

**Proof**

From equation (22), the h-steps ahead forecast error for BL (1, 0, 1, 1) is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{t+h-j}\right) \right) e_{t+h-i} + e_{t+h} \tag{41}$$

Substituting equation (41) in equation in (39) we obtain,

$$dispV = E\left(\hat{e}_m \bullet \hat{e}_m{}^T\right) - 2E\hat{e}_m \bullet \left\{ \left( \sum_{S=m+1}^{n} a_k \sum_{i=1}^{s-m} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{s-j}\right) \right) e_{s-i} + \sum_{s=k+2}^{n} a_s e_s \right) \right\}^T +$$

$$+ E \left( \sum_{s=k+2}^{n} a_s \sum_{s=1}^{s-k} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{s-j}\right) \right) e_{s-i} \sum_{s=k+2}^{n} a_s e_s \right) \left( \sum_{s=k+2}^{n} a_s \sum_{i=1}^{s-k} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{s-j}\right) \right) e_{s-i} + \sum_{s=k+2}^{n} a_s e_s \right)^T \tag{42}$$

Simplifying equation (42) we obtain:

$$E\left(\hat{e}_m \hat{e}_m{}^T\right) = \Sigma$$

$$-2E\left\{\hat{e}_m \bullet \left( \sum_{s=k+2}^{n} a_s \sum_{i=1}^{s-k} \left( \prod_{j=1}^{i} \left( \phi_1 + b_{11} e_{s-j} \right) \right) e_{s-i} + \sum_{s=k+2}^{n} a_s e_s \right)^T \right\}$$

$$= -2E\left\{\hat{e}_m \bullet \left( \sum_{s=k+2}^{n} a_s \sum_{s=k+2}^{s-k} \left( \prod_{j=1}^{i} \left( \phi_1 + b_{11} e_{s-j} \right) \right) e_{s-i} + \sum_{s=k+2}^{n} a_s e_s \right)^T \right\}$$

$$= -2\left\{ \Sigma \sum_{s=k+2}^{n} a_s \phi^{s-k} \right\}$$

$$E\left( \sum_{s=k+2}^{n} a_s \sum_{s=1}^{s-k} \left( \prod_{j=1}^{i} \left( \phi_1 + b_{11} e_{s-j} \right) \right) e_{s-i} \sum_{s=k+2}^{n} a_s e_s \right) \left( \sum_{s=k+2}^{n} a_s \sum_{i=1}^{s-k} \left( \prod_{j=1}^{i} \left( \phi_1 + b_{11} e_{s-j} \right) \right) e_{s-i} + + \sum_{s=k+2}^{n} a_s e_s \right)^T$$

$$= \sum_{s=k+2}^{n} a_s a_s^2 \left\{ \phi_1^2 \, \Sigma + 3 \bullet b_{11}^2 \, \Sigma \Sigma^T + \Sigma \right\} \tag{43}$$

Therefore equation (43) becomes

$$\text{dispV} = \Sigma - 2\left\{ \Sigma \sum_{s=k+2}^{n} a_s \phi_1^{s-k} \right\} +$$

$$I \bullet \sum_{s=k+2}^{n} a_s a_s^T \left\{ \phi_1^2 \, \Sigma + 3 \bullet b_{11}^2 \, \Sigma \Sigma^T + \Sigma \right\} \tag{44}$$

Differentiating equation (44) with respect to the coefficients $a_s$, we get

$$\frac{dispV}{da_k} = 0 - 2\Sigma \left\{ \phi_1^{s-k} \right\} + \sum_{s=k+2}^{n} a_s a_s^T \left\{ \phi_1^2 + 3 \bullet b_{11}^2 \, \Sigma^T + I \right\}$$

$$\hat{a}_s = \left[ \frac{\Sigma \phi_1^{s-k}}{\left\{ \phi_1^2 \, \Sigma + 3 b_{11}^2 \, \Sigma \Sigma^T + \Sigma \right\}} = \frac{\Sigma \phi_1^{s-k}}{\left\{ \phi_1^2 \, \Sigma + 3 b_{11}^2 \, \Sigma \Sigma^T + \Sigma \right\}} \right] \tag{45}$$

Therefore the estimate of the missing value is given by

$$\xi_m^* = \hat{\xi}_m + \sum_{s=k+2}^{n}\left[\frac{\left\{\hat{\phi}^{s-m}\right\}}{\left\{\phi_1^2 I + 3b_{11}^2 \Sigma^T + I\right\}}\right]$$

where I is an identity matrix.

**Corollary**

When we have only one missing value, $\hat{a}_s = a_k$, $\Sigma = \sigma^2$, $a_s = a_k$ .Therefore we have

$$\begin{bmatrix}\xi_k^* \\ \xi_{k+1}^*\end{bmatrix} = \begin{bmatrix}\hat{\xi}_k \\ \hat{\xi}_{k+1}\end{bmatrix} + \sum_{s=k+2}^{n}\left[\frac{\left\{\hat{\phi}^{s-m}\right\}}{\left\{\phi_1^2 + 3b_{11}^2\sigma^2 + 1\right\}}(\xi_s - \hat{\xi}_s)\right]$$

this can be expressed as

$$\begin{bmatrix}\xi_k^* \\ \xi_{k+1}^*\end{bmatrix} = \begin{bmatrix}\hat{\xi}_k \\ \hat{\xi}_{k+1}\end{bmatrix} + \begin{bmatrix}\frac{\phi^1}{\left\{\phi_1^2 + 3b_{11}^2\sigma^2 + 1\right\}}(\xi_{k+2} - \hat{\xi}_{k+2}) \\ \frac{\phi^2}{\left\{\phi_1^2 + 3b_{11}^2\sigma^2 + 1\right\}}(\xi_{k+2} - \hat{\xi}_{k+2})\end{bmatrix} + \begin{bmatrix}\frac{\phi^2}{\left\{\phi_1^2 + 3b_{11}^2\sigma^2 + 1\right\}}(\xi_{k+3} - \hat{\xi}_{k+3}) \\ \frac{\phi^3}{\left\{\phi_1^2 + 3b_{11}^2\sigma^2 + 1\right\}}(\xi_{k+3} - \hat{\xi}_{k+3})\end{bmatrix} + \ldots.$$

**3.4 Estimates of missing values for pure bilinear time series with student t-errors**

The missing values can be obtained using the following theorems.

**3.4.1 Estimates of missing values BL (0, 0, 1, 1) with student t errors**

The pure bilinear time series process of order one BL (0, 0, 1, 1) is of the form

$$x_t = b_{11}x_{t-1}e_{t-1} + e_t, \quad e_t \sim t(0,1) \tag{46}$$

The estimate of the missing value for this model is given by theorem 3.8.

**Theorem 3.8**

The optimal linear estimate for missing observation for BL (0, 0, 1, 1) with student errors is

**Proof**

The stationary BL (0, 0, 1, 1) is given by

$$x_t = \sum_{i=1}^{\infty} \left\{ \prod_{j=1}^{i} b_{11} e_{t-j} \right\} e_{t-i} + e_t \qquad (47)$$

and the h-steps ahead forecast is

$$x_{t+h} = \sum_{i=1}^{\infty} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{t+h-i} + e_{t+h}$$

Therefore the forecast error is

$$x_{t+h} - \hat{x}_{t+h} =$$

$$\sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{t+h-i} + e_{t+h} \qquad (48)$$

Substituting equation (32) in equation (3), the dispersion error is given by

$$E(x_m - x_m^*)^2 = E\left[(x_m - \hat{x}_m) - \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)\right]^2$$

$$= E(x_m - \hat{x}_m)^2 - 2E(x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) + E\left\{\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)\right\}^2 \qquad (49)$$

Simplifying each of the terms of equation (49), we obtain the following:

$$First\ \ Term: \quad E(x_m - \hat{x}_m)^2 = \tfrac{n}{n-2}$$

$$2nd\ Term: \quad E2((x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) = E\ 2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i} b_{11} e_{k-j}\right\} e_{k-i}$$

$$= 2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i} b_{11} e_{k-j}\right\} e_{k-i}$$

$$= 2E\left(\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i} b_{11} e_{k-j}\right\} e_{k-i}\right)$$

$$= 2Ee_m \bullet a_{k-m}(b_{11}{}^{k-m} e_{k-1}.e_{k-2} \ldots e_{k-m+1}$$

$$= 2.Ee_m^2 a_{k-m} b_{11}{}^{k-m} E\prod_{i=1}^{m-1} e_{k-i}$$

$$= 0$$

$$2E\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=0}^{k-m}\left(\prod_{j}^{i} b_{11} e_{k-j}\right) e_{k-i}$$

$$= 2E\hat{e}_m \bullet \begin{bmatrix} a_{m+1} b_{11} e_m{}^2 + a_{m+2}\left(b_{11} e_{m+1}{}^2 + b_{11}{}^2 e_{m+1} e_m{}^2\right) + \\ a_{m+3}\left(b_{11} e_{m+2}{}^2 + b_{11}{}^2 e_{m+2} e_{m+1}{}^2 + b_{11}{}^3 e_{m+2} e_{m+1} e_m{}^2\right) + \ldots \end{bmatrix} = 0$$

$$Third \quad term: E\left[\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i} b_{11} e_{k-j}\right\} e_{k-i} + e_k\right]^2$$

$$= E\begin{bmatrix} a_{m+1} b_{11} e_m{}^2 + a_{m+2}\left(b_{11} e_{m+1}{}^2 + b_{11}{}^2 e_{m+1} e_m{}^2\right) \\ + a_{m+3}\left(b_{11} e_{m+2}{}^2 + b_{11}{}^2 e_{m+2} e_{m+1}{}^2 + b_{11}{}^3 e_{m+2} e_{m+1} e_m{}^2\right) + \ldots \end{bmatrix}^2$$

$$= E\begin{bmatrix} a_{m+1}{}^2 b_{11}{}^2 e_m{}^4 + a_{m+2}{}^2\left(b_{11}{}^2 e_{m+1}{}^4 + b_{11}{}^4 e_{m+1}{}^2 e_m{}^4\right) \\ + a_{m+3}{}^2\left(b_{11}{}^2 e_{m+2}{}^4 + b_{11}{}^4 e_{m+2}{}^2 e_{m+1}{}^4 + b_{11}{}^4 e_{m+2}{}^2 e_{m+1}{}^2 e_m{}^4\right) + \ldots \end{bmatrix}$$

$$\approx a_{m+1}{}^2 b_{11}{}^2 3\upsilon_x(4) + a_{m+2}{}^2\left(3 b_{11}{}^2 \upsilon_x(4)\right) + a_{m+3}{}^2\left(b_{11}{}^2 \upsilon_x(4)\right) + \sum_{k=m+1}^{n} a_k^2 \frac{n}{n-2}$$

$$= \upsilon_x(4) b_{11}{}^2 \sum_{k=m+1}^{n} a_k^2 + \sum_{k=m+1}^{n} a_k^2 \frac{n}{n-2} \qquad (50)$$

Hence equation (50) can be simplified as

$$disp \quad x_m = \tfrac{n}{n-2} - 0 + 3\upsilon_x(4) b_{11}{}^2 \sum_{k=m+1}^{n} a_k^2 + \sum_{k=m+1}^{n} a_k^2 \tfrac{n}{n-2} \qquad (51)$$

Now differentiating equation (51) with respect to $a_k$ and equating to zero, we obtain

$$\frac{d}{da_k} disp\, x_m = \frac{d}{da_k}\left(\tfrac{n}{n-2} - 0 + \upsilon_x(4) b_{11}{}^2 \sum_{k=m+1}^{n} a_k^2\right) + \sum_{k=m+1}^{n} a_k^2 \tfrac{n}{n-2} = 0$$

$$\Rightarrow \qquad 0 - 0 + 2 a_k \upsilon_x(4) b_{11}{}^2 = 0$$

$$\Rightarrow \qquad \hat{a}_k = 0$$

Substituting the values of $a_k$ in equation (3), we obtain optimal estimator of the missing value as

$$x_m^* = \hat{x}_m$$
$$= E(x_t / x_{m-1}, x_{m-2}, \ldots.. x_1)$$
$$= \hat{b}_{11} x_{m-1} \hat{e}_{m-1}$$

This shows that the missing value is a one-step-ahead prediction based on the past observations collected before the missing value. This is in agreement with other studies that have estimated missing values using forecasting (Nassiuma, 1994).

### 3.4.2 Estimating missing values for BL (1, 0, 1, 1) with student-t errors

The bilinear model BL (1, 0, 1, 1) with t- errors is expressed as

$$x_t = \phi x_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t, \quad where \;\; e_t \sim t(0,1) \tag{52}$$

The missing value is obtained using theorem 3.9.

**Theorem 3.9**

The optimal linear estimate for missing value for BL (1, 0, 1, 1) with student errors is given by

$$x_m^* = \hat{\phi} x_{m-1} + \hat{b}_{11} x_{m-1} e_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{\,k-m}}{\left[\frac{n}{n-2}(\hat{\phi}_1^{\,2} + 1) + b_{11}^{\;2} v(4)\right]} (x_{m+1} - \hat{x}_{m+1})$$

**Proof**

The stationary BL (1 ,0 ,1 ,1) can be expressed as

$$x_t = \sum_{i=1}^{\infty} \left( \prod_{j=1}^{i} \left(\phi + b_{11} e_{t-j}\right) \right) e_{t-i} + e_t \tag{53}$$

The h-steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(\phi + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h}$$

and the h-steps ahead forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1}\left(\prod_{j=1}^{i}\left(\phi + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h} \qquad (54)$$

Substituting equation (54) in equation (3), we obtain

$$Disp\ x_m = E\left(\hat{e}_m{}^2\right) - 2E\left\{\hat{e}_m \bullet \left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)\right\} +$$

$$+\ E\left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)^2 \qquad (55)$$

Simplifying the terms RHS of equation (55), we obtain

$$E\left(\hat{e}_m{}^2\right) = \frac{n}{n-2}$$

$$E\left\{2\hat{e}_m \bullet \left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)\right\}$$

$$= E\left\{2\hat{e}_m \bullet \left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)\right\}$$

$$= E\hat{e}_m \bullet 2\left[\begin{array}{l} a_{m+1}(\phi_1 + b_{11}e_m)e_m + a_{m+1}e_{m+1} + a_{m+2}(\phi_1 + b_{11}e_{k-j})(\phi_1 + b_{11}e_{k-j})e_m \\ + a_{m+2}e_{m+2} + + a_{m+3}(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_{m+1)})(\phi_1 + b_{11}e_m)e_m + a_{m+3}e_{m+3} + \ldots) \end{array}\right]$$

$$= \sum_{k=m+1}^{n} a_k b^{k-m} \frac{n}{n-2}$$

$$E\left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)^2 =$$

$$= \left[\begin{array}{l} a_{m+1}(\phi_1 + b_{11}e_m)e_m + a_{m+1}e_{m+1} + a_{m+2}(\phi_1 + b_{11}e_{m+1)})(\phi_1 + b_{11}e_m)e_m + a_{m+2}e_{m+2} \\ + a_{m+3}(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_{m+1)})(\phi_1 + b_{11}e_m)e_m + a_{m+3}e_{m+3} + \ldots) \end{array}\right]^2$$

$$= \left[\begin{array}{l} a_{m+1}{}^2(\phi_1 + b_{11}e_m)^2 e_m{}^2 + a_{m+1}{}^2 e_{m+1}{}^2 + a_{m+2}{}^2(\phi_1 + ce_{m+1)})^2(\phi_1 + b_{11}e_m)^2 e_m + a_{m+2}{}^2 e_{m+2}{}^2 \\ + a_{m+3}{}^2(\phi_1 + b_{11}e_{m+1})^2(\phi_1 + b_{11}e_{m+1)})^2(\phi_1 + b_{11}e_m)^2 e_m{}^2 + a_{m+3}{}^2 e_{m+3}{}^2 + .. \\ + a_{m+1}\left[(\phi_1 + b_{11}e_m)e_m + e_{m+1})\right]\left[a_{m+2}(\phi_1 + b_{11}e_{m+1)})(\phi_1 + b_{11}e_m)e_m + a_{m+2}e_{m+2}\right] + \end{array}\right]$$

$$\approx \left[\sum_{k=m+1}^{n} a_k{}^2(\phi^2 \frac{n}{n-2} + b_{11}{}^2 v(4) + \frac{n}{n-2})\right]$$

where

$$v(4) = v(4) = E(e_t{}^4) = \frac{n^2\Gamma(\frac{5}{2})\Gamma\frac{(n-4)}{2}}{\Gamma(\frac{n}{2})\Gamma(\frac{1)}{2}}.$$

Hence equation (56) becomes

$$disp\, x_m = \frac{n}{n-2} - 2\sum_{k=m+1}^{n} a_k \phi^{(k-m)} \frac{n}{n-2} + \sum_{k=m+1}^{n} a_k{}^2 (\phi^2 \frac{n}{n-2} + b_{11}{}^2 v(4) + \frac{n}{n-2}). \quad (56)$$

Differentiating equation (56) with respect to the coefficients, we get

$$\frac{d}{da_k}\left[-2\frac{n}{n-2}\sum_{k=m+1}^{m+1} a_k b^{k-m} + 2\sum_{k=m+1}^{n} a_k (\phi^2 \frac{n}{n-2} + b_{11} v(4) + \frac{n}{n-2})\right] = 0$$

$$\Rightarrow \quad 0 - 2\frac{n}{n-2}\phi^{k-m} + 2a_k (\phi^2 \frac{n}{n-2} + b_{11} v(4) + \frac{n}{n-2}) = 0$$

$$\Rightarrow \quad a_k = \frac{\hat{\phi}_1^{\,k-m}}{\left[\frac{n}{n-2}(\hat{\phi}_1{}^2 + 1) + b_{11}{}^2 v(4)\right]}$$

The optimal linear estimate of $x_m$, *denoted by* $x_m^*$, that minimizes the error dispersion error of the estimate is thus given as

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

$$= \hat{x}_m + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{\,k-m}}{\left[\frac{n}{n-2}(\phi_1{}^2 + 1) + b_{11}^2 v(4)\right]}(x_{m+1} - \hat{x}_{m+1})$$

$$= \hat{\phi} x_{m-1} + \hat{b}_{11} x_{m-1} e_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{\,k-m}}{\left[\frac{n}{n-2}(\hat{\phi}_1{}^2 + 1) + c^2 v(4)\right]}(x_{m+1} - \hat{x}_{m+1})$$

### 3.4.3 Estimating missing values for BL (0, 1 ,1 ,1) with student-t errors

The BL (0, 1, 1, 1) model with student-t errors is given by

$$x_t = b_{11} x_{t-1} e_{t-1} + \theta e_{t-1} + e_t \qquad where \qquad e_t \sim t(0,1) \tag{57}$$

The missing values are obtained from the following theorem 3.10

**Theorem 3.10**

The optimal linear estimate for BL (0, 1, 1, 1) with student errors is given by

$$\hat{x}_m = \hat{b}_{11} x_{m-1} \hat{e}_{t-1} + \hat{\theta} \hat{e}_{m-1}$$

**Proof**

The stationary bilinear BL (0,1,1,1) is expressed as

$$x_t \quad = \sum_{i=1}^{\infty} \left( \coprod_{j=1}^{i} \theta e_{t-j} \right) e_{t-i-1} + \sum_{i=1}^{\infty} \left( \prod_{j=1}^{i} b_{11} e_{t-j} \right) e_{t-j} + e_t \tag{58}$$

The h-steps ahead forecast is given by

$$x_{t+h} \quad = \sum_{i=1}^{\infty} \left( \coprod_{j=1}^{i} e_{t-j} \right) e_{t+h-i-1} + \sum_{i=1}^{\infty} \left( \prod_{j=1}^{i} b_{11} e_{t+h-j} \right) e_{t+h-j} + e_{t+h}$$

and the forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} \quad = \sum_{i=1}^{h-1} \left( \theta \coprod_{j=1}^{i} e_{t-j} \right) e_{t+h-i-1} + \sum_{i=1}^{h-1} \left( \prod_{j=1}^{i} b_{11} e_{t+h-j} \right) e_{t+h-j} + e_{t+h} \tag{59}$$

Substituting equation (59) in equation (3), we get

$$E(x_m - x_m^*)^2 = E\left\{(x_m - \hat{x}_m) - \sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\coprod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{h-1}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]\right\}^2$$

$$disp\ \ x_m = E\left\{(\hat{e}_m) - \sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\theta\coprod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]\right\}^2$$

$$= \left\{\begin{array}{l} E(\hat{e}_m)^2 - E\ \ 2(\hat{e}_m) \bullet \sum_{k=m+1}^{n} a_k \left[\begin{array}{l}\sum_{i=1}^{k-m}\left(\theta\coprod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} \\ + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\end{array}\right] + \\ E\left\{\sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\theta\coprod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]\right\}^2 \end{array}\right\} \qquad (60)$$

Simplifying each of the terms of equation (60), we obtain

$$E(\hat{e}_m)^2 = \frac{n}{n-2}$$

$$E\ 2(\hat{e}_m) \bullet \sum_{k=m+1}^{n} a_k \left[\sum_{i=1}^{k-m}\left(\coprod_{j=1}^{i} e_{k-j}\right)e_{k-i-1} + \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i} b_{11}e_{k-j}\right)e_{k-i} + e_k\right]$$

$$\begin{aligned} &= E\ 2(\hat{e}_m) \bullet \left(a_{m+1}\left\{\theta e_m e_{m-1} + ce_m{}^2 + e_{m+1}\right\}\right) + E\ 2(\hat{e}_m) \bullet a_{m+2}\left\{\theta e_{m+1}e_m e_{m-1} + b_{11}{}^2 e_{m+1}e_m{}^2 + e_{m+2}\right\} \\ &\quad + E\ 2(\hat{e}_m) \bullet a_{m+3}\left\{\theta e_{m+2}e_{m+1}e_m + b_{11}{}^3 e_{m+2}e_{m+1}e_m{}^2 + e_{m+3}\right\} \\ &\quad + E\ 2(\hat{e}_m) \bullet a_{m+4}\left\{\theta e_{m+3}e_{m+2}e_{m+1}e_m e_{m-1} + b_{11}{}^3 e_{m+3}e_{m+2}e_{m+1}e_m{}^2 + e_{m+4}\right\} \\ &= 0 \end{aligned}$$

*This implies that* $\hat{a}_k = 0$. *Therefore the best linear estimate is by*

$$\hat{x}_m = b_{11}x_{m-1}e_{t-1} + \theta e_{m-1}$$

### 3.4.4 Estimating missing values for BL (1, 0, 1, 1) with student-t errors

The bilinear BL (1,0,1,1) model with student t errors is

$$x_t = \phi_1 x_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t, \ \ where \ \ e_t \sim t(0,1) \tag{61}$$

The missing values are estimated based on the following theorem 3.11.

**Theorem 3.11**

The optimal linear estimate for BL (1, 0, 1, 1) with student errors is given by

$$x_m^* = \hat{\phi} x_{m-1} + \hat{b}_{11} x_{m-1} \hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{\,k-m}}{\left[ \frac{n}{n-2} (\hat{\phi}_1^{\,2} + 1) + b_{11}^{\,2} v(4) \right]} (x_{m+1} - \hat{x}_{m+1})$$

Where

$$v(4) = E(e_t^{\,4}) = \frac{n^2 \Gamma(\frac{5}{2}) \Gamma \frac{(n-4)}{2}}{\Gamma(\frac{n}{2}) \Gamma(\frac{1}{2}}$$

**Proof**

Performing recursive substitution on equation (61), the stationary BL (1, 0, 1, 1) can be expresses as

$$x_t = \sum_{i=1}^{\infty} \left( \prod_{j=1}^{i} \left( \phi_1 + b_{11} e_{t-j} \right) \right) e_{t-i} + e_t . \tag{62}$$

The h-steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{\infty} \left( \prod_{j=1}^{i} \left( \phi_1 + b_{11} e_{t+h-j} \right) \right) e_{t+h-i} + e_{t+h}$$

and the h-steps ahead forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h} \tag{63}$$

Substituting equation (63) in equation (3), we have

$$disp\ x_m = E\left(\hat{e}_m^{\ 2}\right) - 2E\left\{\hat{e}_m \bullet \left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)\right\} +$$

$$+ E\left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)^2 \tag{64}$$

Simplifying equation (64), we obtain

First term: $E\left(\hat{e}_m^{\ 2}\right) = \sigma^2$

$$Second\ \ term: -\left\{2\hat{e}_m \bullet \left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)\right\}$$

$$= -E\hat{e}_m \bullet 2\left[\begin{array}{l} a_{m+1}(\phi_1 + b_{11}e_m)e_m + a_{m+1}e_{m+1} + a_{m+2}(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_{m+2})e_m + a_{m+2}e_{m+2} \\ + a_{m+3}(\phi_1 + b_{11}e_{m+2})(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_m)e_m + a_{m+3}e_{m+3} + ...) \end{array}\right]$$

$$= -2\sum_{k=m+1}^{n} a_k \phi_1^{\ k-m}$$

$$Third\ \ term: \ E\left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)^2$$

$$= E\left\{\left(\sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m}(\phi_1 + b_{11}e_{k-j}\right)^2 e_{k-j}^{\ 2} + 2\left(\sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m}(\phi_1 + b_{11}e_{k-j}\right)e_{k-j}\sum_{k=m+1}^{n} a_k e_k + \left(\sum_{k=m+1}^{n} a_k e_k\right)^2\right\}$$

$$\left\{E\left(\sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m}(\phi_1 + b_{11}e_{k-j}\right)^2 e_{k-j}^{\ 2} + E2\left(\sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m}(\phi_1 + b_{11}e_{k-j}\right)e_{k-j}\sum_{k=m+1}^{n} a_k e_k + E\left(\sum_{k=m+1}^{n} a_k e_k\right)^2\right\}$$

$$E2\left(\sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m}(\phi_1 + b_{11}e_{k-j}\right)e_{k-j}\sum_{k=m+1}^{n} a_k e_k = 0$$

$$E\left(\sum_{k=m+1}^{n} a_k e_k\right)^2 = \sum_{k=m+1}^{n} \frac{n}{n-2} a_k{}^2$$

$$E\left(\sum_{k=m+a}^{n} a_k \sum_{i=1}^{k-m}(\phi_1 + b_{11}e_{k-j}\right)^2 e_{k-j}{}^2 = \sum_{k=m+1}^{n} a_k^2(\phi_1{}^2\sigma^2 + 3\sigma^4 b_{11}{}^2)$$

Hence equation (65) becomes

$$dispx_m = \sigma^2 - 2\sum_{k=m+1}^{n} a_k \phi^{k-m}\sigma^2 + \sum_{k=m+1}^{n} a_k^2(\phi_1{}^2\sigma^2 + 3\sigma^4 b_{11}{}^2) + \sum_{k=m+1}^{n} a_k^2\sigma^2 \qquad (66)$$

Differentiating equation (66 with respect to the coefficients $a_k$ and setting it to zero, we get

$$\frac{d}{da_k}\left[\sigma^2 - 2\frac{n}{n-2}\sum_{k=m+1}^{n} a_k b_{11}{}^{(k-m)} + \sum_{k=m+1}^{n} a_k{}^2\left(\frac{n}{n-2}\phi_1{}^2 + v(4)b_{11}{}^2 + \frac{n}{n-2}\right)\right] = 0$$

$$\Rightarrow \quad 0 - 2\phi^{(k-m)})\frac{n}{n-2} + 2a_k\left(\frac{n}{n-2}\phi_1{}^2 + v(4)b_{11}{}^2 + \frac{n}{n-2}\right) = 0$$

where

$$v(4) = E(e_t{}^4) = \frac{n^2\Gamma(\frac{5}{2})\Gamma\frac{(n-4)}{2}}{\Gamma(\frac{n}{2})\Gamma(\frac{1)}{2}}$$

$$Therefore \qquad \hat{a}_k = \frac{\hat{\phi}_1{}^{(k-m)}}{\left\{\frac{n}{n-2}\hat{\phi}_1{}^2 + v(4)b_{11}{}^2 + \frac{n}{n-2}\right\}}$$

The optimal linear estimate of $x_m$, *given by* $x_m^*$ that minimizes the error dispersion error of the estimate is thus given as

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

$$= \hat{x}_m + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{(k-m)}}{\left\{ \frac{n}{n-2} \hat{\phi}_1^{\,2} + v(4) b_{11}^{\,2} + \frac{n}{n-2} \right\}} (x_{m+1} - \hat{x}_{m+1})$$

$$= \hat{\phi}_1 x_{m-1} + \hat{b}_{11} x_{m-1} \hat{e}_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{(k-m)}}{\left\{ \frac{n}{n-2} \hat{\phi}_1^{\,2} + v(4) b_{11}^{\,2} + \frac{n}{n-2} \right\}} (x_k - \hat{x}_k)$$

### 3.4.5 Estimating missing values BL (p, 0, p, p) with Student-t innovations

The pure bilinear time series model BL (p, 0, p, p) with student-t errors is

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{i=1}^{p} \sum_{i=1}^{p} b_{i1} x_{t-i} e_{t-i} + e_t \quad where \qquad e_t \sim t(0,1)$$

The missing values can be estimated using the following theorem 3.12.

**Theorem 3.12**

The optimal linear estimate for one missing value $x_m$ for the general bilinear time series model BL (p, 0, p, p) with student t-errors is given by

$$x_m^* = \sum_{i=1}^{p} \hat{\phi}_i x_{t-i} + \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{b}_{ij} x_{t-i} \hat{e}_{t-j} + \sum_{k=m+1}^{n} \frac{\frac{n}{n-2} \left\{ \hat{\phi}_s^{\,k-m} + 2 \sum_{s \neq r}^{p} \hat{\phi}_s \hat{\phi}_r \right\} \{x_k - x_k\}}{\left\{ \frac{n}{n-2} \left\{ \sum_{s=1}^{p} \hat{\phi}_s^2 + \sum_{j=1, s \neq j}^{p} \hat{b}_{sj}^{\,2} \right\} + \sum_{j=1}^{p} v(4) \hat{b}_{jj}^{\,2} \right\}}$$

Where v(4) is the fourth moment of the data given a

$$v(4) = E(e_t^{\ 4}) = \frac{n^2\Gamma(\frac{5}{2})\Gamma\frac{(n-4)}{2}}{\Gamma(\frac{n}{2})\Gamma(\frac{1)}{2}}$$

It can be estimated by v(4)=kurtosis*(variance)$^2$

**Proof**

The stationary bilinear time series model BL (p, 0, p, p) is of the form

$$x_t = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}e_{t-s-r} + \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{t-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t-j}\right\}e_{t-s-r} + e_t \quad (67)$$

The h steps ahead forecast for equation (67) is given by

$$x_{t+h} = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t+h-s-r} + \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{t+h-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t+h-s-r} + e_t$$

or

$$x_k = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + e_t \quad (68)$$

and the forecast is

$$x_k = \sum_{i=1}^{p}\sum_{i=1}^{k-1}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} +$$

$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} + e_t$$

There forecast error is

$$x_k - \hat{x}_k = \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r}$$

$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} + e_t$$

$$+\sum_{i=0}^{\infty}\prod_{j=1}^{i}(\phi_2 + b_{21} e_{k-2j+1} + b_{22} e_{k-2j})\left(\theta_2 e_{k-2i-2} + \theta_1 e_{k-2i-1} + e_{k-2i}\right) + \left\{\theta_2 e_{k-2} + \theta_1 e_{k-1} + e_k\right\} \quad (69)$$

Substituting equation (69) in equation (3) and simplifying, we obtain

$$\text{First term} \qquad E\left(\hat{e}_m{}^2\right) = \frac{n}{n-2}$$

$$\text{sec}\,ond\ \ term \qquad -2E(x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \bullet \left\{ \begin{array}{l} \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\} e_{t-sj} \\ + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\} e_{t+h-s-r} \\ + \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{t+h-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j}\right\} e_{t+h-s-r} + e_t \end{array} \right\} =$$

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-si} = -2\sum_{k=m+1}^{n}\frac{n}{n-2}\phi^{(k-m)/s}$$

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{s<r\ and\ s>}^{p} \sum_{r}^{p} \left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r}$$

$$= \begin{cases} \sum_{k=m+1}^{n} \frac{4n}{n-2} a_k \phi_r \phi_s & for\ s+r=k-m \\ 0 & otherwise \end{cases}$$

Third term

$$E\left\{\begin{array}{l} \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} \\ \qquad + \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj} e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j}\right\} e_{k-s-r} + e_t \end{array}\right\}^2$$

$$= \sum_{k=m+1}^{n} a_k^2 \left\{ \sum_{s=1}^{p} \frac{n}{n-2} \phi_s^2 + \sum_{j=1, s\neq j}^{p} b_{sj}{}^2 \frac{n}{n-2} + e_{k-j} + \sum_{j=1}^{p} v(4) b_{jj}{}^2 \right\}$$

Equation (69) simplifies to

$$\begin{aligned} dispx_m &= -2\frac{n}{n-2}\sum_{k=m+1}^{n}\phi_s^{(k-m)/s} + \sum_{k=m+1}^{n}\frac{n}{n-2}4a_k\phi_r\phi_s \\ &+ \sum_{k=m+1}^{n} a_k^2\left\{\sum_{s=1}^{p}\frac{n}{n-2}\phi_s^2 + \sum_{j=1,s\neq j}^{p}\frac{n}{n-2}b_{sj}{}^2 + e_{k-j} + \sum_{j=1}^{p} v(4) b_{jj}{}^2\right\} \end{aligned} \qquad (70)$$

where

$$v(4) = E(e_t{}^4) = \frac{n^2\Gamma(\frac{5}{2})\Gamma\frac{(n-4)}{2}}{\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})}$$

Differentiating equation (71) with respect to $a_k$, we have

$$\frac{d}{da_k} disp = -2\frac{n}{n-2}\sum_{k=m+1}^{n}\phi_s^{(k-m)/s} + \frac{n}{n-2}\phi_r\phi_s +$$

$$2a_k\left\{\sum_{s=1}^{p}\frac{n}{n-2}\phi_s^2 + \sum_{j=1,s\neq j}^{p}\frac{n}{n-2}b_{sj}^{\ 2} + e_{k-j} + \sum_{j=1}^{p}v(4)b_{jj}^{\ 2}\right\} = 0$$

Solving for $a_k$, we get

$$\hat{a}_k = \frac{\sum_{K=m+1}^{n}\frac{n}{n-2}\hat{\phi}_s^{\ k-m} + 2\sum_{s\neq r}^{p}\frac{n}{n-2}\hat{\phi}_s\hat{\phi}_r}{\left\{\frac{n}{n-2}\left\{\sum_{s=1}^{p}\hat{\phi}_s^2 + \sum_{j=1,s\neq j}^{p}\hat{b}_{sj}^{\ 2}\right\} + \sum_{j=1}^{p}v(4)\hat{b}_{jj}^{\ 2}\right\}}$$

**Corollary**

For p=1, we have the bilinear model BL (1, 0, 1, 1). The optimal linear estimate is given by

$$x_m^* = \hat{\phi}_1 x_{m-1} + \hat{b}_{11}x_{m-1}\hat{e}_{m-1} + \sum_{k=m+1}^{n}\frac{\hat{\phi}^{(k-m)}}{(1+\hat{\phi}_1^{\ 2} + v(4)\hat{b}_{11}^{\ 2})}(x_k - \hat{x}_k)$$

### 3.5 Estimating missing values for bilinear time series model with GARCH innovations

The bilinear time series $x_t$ of order BL (p, q, m, k) with GARCH innovations satisfies the difference equation

$$x_t = \sum_{i=1}^{p}\phi_i x_{t-i} + \sum_{j=1}\theta e_{t-j} + \sum\sum b_{ij}x_{t-i}e_{t-j} + e_t \quad . \qquad (71)$$

where $\theta, \phi$ and $b_{ij}$ are constants while $e_t$ is a purely random process and $\theta_o$ =1 and

$$e_t/\psi_{t-1} \sim N(0,h_t),\ e_t = \eta_t h_t^{\frac{1}{2}},\ \ h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i e_{t-i} + \sum_{i=1}^{p}\beta_i h_{t-i} \qquad (72)$$

with the inequality conditions $\alpha_0 > 0$, $\alpha_i \geq 0$ for i=1,…,q, $\beta_i \geq 0$, for i=1,…,p to ensure that the conditional variance is strictly positive. Missing values for BL (1, 0, 1, 1) can obtain from the following theorem 3.13.

### 3.5.1 Estimating missing values for BL (0, 0, 1, 1) time series model with GARCH innovations

The simplest pure bilinear time series model of order one, BL(0, 0, 1, 1) is of the form

$$x_t = b_{11} x_{t-1} e_{t-1} + e_t \tag{73}$$

with $e_t$ distributed as specified in equation (43).

**Theorem 3.13**

The optimal linear estimate for missing observation for BL (0, 0, 1, 1) with GARCH errors is given by

$$x_m^* = \hat{b}_{11} x_{m-1} \hat{e}_{m-1}$$

**Proof**

Through recursive substitution of equation (73), the stationary BL (0, 0, 1, 1) is obtained as

$$x_t = \sum_{i=1}^{\infty} \left\{ \prod_{j=1}^{i} b_{11} e_{t-j} \right\} e_{t-i} + e_t$$

By adding h to t in the equation, the h-steps ahead forecast is

$$x_{t+h} = \sum_{i=1}^{\infty} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{t+h-i} + e_{t+h}$$

Therefore the forecast error is

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{t+h-i} + e_{t+h} \qquad (74).$$

or it can also be represented as

$$x_k - \hat{x}_k = \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{k-i} + e_k \qquad (75).$$

Substituting equation (75) in equation (3), we have

$$\begin{aligned} disp\ x_m &= E(x_m - \hat{x}_m)^2 - 2E(x_m - \hat{x}_m) \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) + E\left\{ \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) \right\}^2 \\ &= E(\hat{e}_m{}^2) \quad - 2E(\hat{e}_m) \sum_{k=m+1}^{n} a_k \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{k-i} + e_k \\ &+ E\left\{ \sum_{k=m+1}^{n} a_k \sum_{k=m+1}^{n} a_k \sum_{i=1}^{h-1} \left\{ \prod_{j=1}^{i} b_{11} e_{t+h-i} \right\} e_{k-i} + e_k \right\}^2 \qquad (76) \end{aligned}$$

Simplifying each term of equation (76) separately, we have

$$First\ Term: \quad E(x_m - \hat{x}_m)^2 = h_m$$

$$2nd\ Term: \quad E2((x_m - \hat{x}_m) \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) = E\ 2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m} \left\{ \prod_{j=1}^{i} b_{11} e_{k-j} \right\} e_{k-i}$$

$$Third \quad term: E\left[\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left\{\prod_{j=1}^{i} b_{11} e_{k-j}\right\} e_{k-i} + e_k\right]^2$$

$$= E\left[a_{m+1} b_{11} e_m{}^2 + a_{m+2}\left(b_{11}{}^2 e_{m+1} e_m{}^2\right) + a_{m+3}\left(b_{11} e_{m+2}{}^2 + b_{11}{}^2 e_{m+2} e_{m+1}{}^2 + b_{11}{}^3 e_{m+2} e_{m+1} e_m{}^2\right) + \sum_{k=m+1}^{n} a_k e_k .\right]^2$$

$$= E\begin{bmatrix} a_{m+1}{}^2 b_{11}{}^2 e_m{}^4 + a_{m+2}{}^2\left(b_{11}{}^2 e_{m+1}{}^4 + b_{11}{}^4 e_{m+1}{}^2 e_m{}^4\right) \\ + a_{m+3}{}^2\left(b_{11}{}^2 e_{m+2}{}^4 + b_{11}{}^4 e_{m+2}{}^2 e_{m+1}{}^4 + b_{11}{}^4 e_{m+2}{}^2 e_{m+1}{}^2 e_m{}^4\right) \\ + (\sum_{k=m+1}^{n} a_k e_k)^2 \end{bmatrix}$$

$$\approx a_{m+1}{}^2 b_{11}{}^2 (3h_m{}^2) + a_{m+2}{}^2\left(b_{11}{}^2 3h_{m+1}{}^2\right) + a_{m+3}{}^2\left(b_{11}{}^2 3h_{m+2}{}^2)\right) + \sum_{k=m+1}^{n} a_k^2 h_k^2$$

$$= b_{11}{}^2 \sum_{k=m+1}^{n} 3a_k^2 h_k{}^2 + \sum_{k=m+1}^{n} a_k^2 h_k^2 \tag{77}$$

Hence equation (77) can be simplified as

$$disp \quad x_m = h_m + b_{11}{}^2 \sum_{k=m+1}^{n} 3a_k^2 h_k{}^2 + \sum_{k=m+1}^{n} a_k^2 h_k^2 \tag{78}$$

Now differentiating equation (78) with respect to $a_k$ and equating to zero, we obtain

$$\begin{aligned} \frac{d}{da_k} disp\, x_m &= \frac{d}{da_k}(h_m + b_{11}{}^2 \sum_{k=m+1}^{n} 3a_k^2 h_k{}^2 + \sum_{k=m+1}^{n} a_k^2 h_k^2) = 0 \\ \Rightarrow \qquad & 0 - 0 + 6a_k h_k b_{11}{}^2 + a_k h_k b_{11}{}^2 = 0 \\ \Rightarrow \qquad & \hat{a}_k = 0 \end{aligned}$$

Substituting the values of $a_k$ in equation (3), we obtain optimal estimator of the missing value as

$$
\begin{aligned}
x_m^* &= \hat{x}_m \\
&= E(x_t / x_{m-1}, x_{m-2}, .....x_1) \\
&= \hat{b}_{11} x_{m-1} \hat{e}_{m-1}
\end{aligned}
$$

This is the same result we obtained for pure bilinear time series model whose innovations are distributed. This shows that the missing value is a prediction based on the past observations collected before the missing value. This is in agreement with other studies that have estimated missing values using forecasting.

### 3.5.2 Estimating missing values for BL (1, 0, 1, 1) with GARCH innovations

The bilinear BL (1,0,1,1) model with GARCH innovations errors is

$$
x_t = \phi_1 x_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t \tag{79}
$$

Where $e_t$ is distributed as specified in equation (79). The estimate of the missing value is obtained using theorem 3.14.

**Theorem 3.14**

The optimal linear estimate for BL (1, 0, 1, 1) with GARCH errors is given by

$$
x_m^* = \hat{\phi}_1 x_{m-1} + b_{11} x_{m-1} e_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1^{\,k-m}}{(\hat{\phi}_{11}^{\;2} + 3b_{11}^{\;2} h_k + 1)} (x_k - \hat{x}_m)
$$

**Proof**

The stationary bilinear time series model with GARCH errors of order BL (1, 0, 1, 1) can be expressed as

$$x_t = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{t-j}\right)\right)e_{t-i} + e_t \qquad (90)$$

The h-steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}\left(\phi + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h} \qquad (91)$$

and the h-steps ahead forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{t+h-j}\right)\right)e_{t+h-i} + e_{t+h} \qquad (92)$$

Substituting equation (92) in equation (3), we have,

$$\begin{aligned} disp\ x_m &= E\left(\hat{e}_m{}^2\right) - 2E\left\{\hat{e}_m \bullet \left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)\right\} \\ &\quad + E\left(\sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m}\left(\prod_{j=1}^{i}\left(\phi_1 + b_{11}e_{k-j}\right)\right)e_{k-i} + \sum_{k=m+1}^{n} a_k e_k\right)^2 \end{aligned} \qquad (93)$$

Simplifying equation (93), we obtain

$$\text{Now} \quad E\left(\hat{e}_m{}^2\right) = h_m$$

$$E\left\{ 2\hat{e}_m \bullet \left( \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{k-j}\right) \right) e_{k-i} + \sum_{k=m+1}^{n} a_k e_k \right) \right\}$$

$$= E\left\{ 2\hat{e}_m \bullet \left( \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^{i} \left(\phi_1 + b_{11} e_{k-j}\right) \right) e_{k-i} + \sum_{k=m+1}^{n} a_k e_k \right) \right\} = 2 a_k \phi_1{}^{k-m} h_m$$

$$E\left( \sum_{k=m+1}^{n} a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^{i} \left(b + c e_{k-j}\right) \right) e_{k-i} + \sum_{k=m+1}^{n} a_k e_k \right)^2 =$$

$$= E\begin{bmatrix} a_{m+1}(\phi_1 + b_{11}e_m)e_m + a_{m+1}e_{m+1} + a_{m+2}(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_m)e_m + a_{m+2}e_{m+2} \\ + a_{m+3}(\phi_1 + b_{11}e_{m+2})(\phi_1 + b_{11}e_{m+1})(\phi_1 + b_{11}e_m)e_m + a_{m+3}e_{m+3} + \ldots) \end{bmatrix}^2$$

$$= E\begin{bmatrix} a_{m+1}{}^2(b + ce_m)^2 e_m{}^2 + a_{m+1}{}^2 e_{m+1}{}^2 + a_{m+2}{}^2 (b + ce_{m+1)})^2 (b + ce_m)^2 e_m + a_{m+2}{}^2 e_{m+2}{}^2 \\ + a_{m+3}{}^2 (b + ce_{m+1})^2 (b + ce_{m+1)})^2 (b + ce_m)^2 e_m{}^2 + a_{m+3}{}^2 e_{m+3}{}^2 + .. \\ + a_{m+1}\left[(b + ce_m)e_m + e_{m+1})\right]\left[a_{m+2}(b + ce_{m+1)})(b + ce_m)e_m + a_{m+2}e_{m+2}\right] + \end{bmatrix}$$

$$= \sum_{k=m+1}^{n} \left[ a_k{}^2 (\phi_{11}{}^2 h_k + 3 b_{11}{}^2 h_k{}^2 + h_k) \right]$$

Hence equation (88) becomes

$$disp\, x_m = h_m - 2 \sum_{k=m+1}^{n} a_k b^{k-m} h_k + \sum_{k=m+1}^{n} \left[ a_k{}^2 (\phi_{11}{}^2 h_k + 3 b_{11}{}^2 h_k{}^2 + h_k) \right] \qquad (94).$$

Differentiating equation (94) with respect to the coefficients, we get

$$\frac{d}{da_k}\left[h_m - 2\sum_{k=m+1}^{n} a_k b^{k-m} h_k + \sum_{k=m+1}^{n} a_k{}^2(\phi_1{}^2 h_k + b_{11}{}^2 3h_k^2 + h_k)\right] = 0$$

$$\Rightarrow \quad 0 - 2b^{k-m}h_k + 2a_k h_k(\phi_{11}{}^2 + 3b_{11}{}^2 h_k + 1) = 0$$

$$\Rightarrow \quad a_k = \frac{\hat{\phi}_1{}^{k-m}}{(\hat{\phi}_{11}{}^2 + 3b_{11}{}^2 h_k + 1)}$$

Therefore the estimate of the missing value for the BL (1, 0, 1, 1) is

$$\Rightarrow \quad x_m^* = \hat{\phi}_1 x_{m-1} + b_{11} x_{m-1} e_{m-1} + \sum_{k=m+1}^{n} \frac{\hat{\phi}_1{}^{k-m}}{(\hat{\phi}_{11}{}^2 + 3b_{11}{}^2 h_k + 1)}(x_k - \hat{x}_m)$$

### 3.5.3 Estimating missing values for BL (p, 0, p, p) with GARCH errors

The bilinear time series BL (p, 0, p, p) is given by

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{i=1}^{p}\sum_{i=1}^{p} b_{ii} x_{t-i} e_{t-i} + e_t \qquad (95)$$

where

$$e_t / \psi_{t-1} \sim N(0, h_t),\ e_t = \eta_t h_t^{\frac{1}{2}},\ h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i e_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

with the inequality conditions $\alpha_0 > 0$, $\alpha_i \geq 0$ for i=1,…,q, $\beta_i \geq 0$, for i=1,…,p to ensure that the conditional variance is strictly positive. The estimate of the missing value for BL (p, 0, p, p) with GARCH errors is given in theorem 3.15.

**Theorem 3.15**

The optimal linear estimate for one missing value $x_m$ for the general bilinear time series model BL (p, 0, p, p) is given by

$$x_m^* = \sum_{i=1}^{p} \hat{\phi}_i x_{t-i} + \sum_{i=1}^{p}\sum_{j=1}^{p} \hat{b}_{ij} x_{t-i}\hat{e}_{t-j} + \sum_{k=m+1}^{n} \frac{\sum_{K=m+1}^{n} \hat{\phi}_s^{\,k-m} + 2\sum_{s\neq r}^{p} \hat{\phi}_s \hat{\phi}_r}{\left\{ \sum_{s=1}^{p} \hat{\phi}_s^2 + \sum_{j=1, s\neq j}^{p} \hat{b}_{sj}^{\ 2} + \sum_{j=1}^{p} 3\hat{b}_{jj}^{\ 2} h_t \right\}} (x_k - \hat{x}_k)$$

**Proof**

The stationary bilinear time series BL (p, 0, p, p) can be expressed as

$$x_t = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i} \left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{t-j} \right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p} \left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{t-j} \right\}\left\{ \phi_r + \sum_{j=1}^{p} b_{sj} e_{t-j} \right\} e_{t-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p} \left\{ \phi_r + \sum_{j=1}^{p} b_{rj} e_{t-j} \right\}\left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{t-j} \right\} e_{t-s-r} + e_t$$

Therefore h steps ahead forecast is given by

$$x_{t+h} = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i} \left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j} \right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p} \left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j} \right\}\left\{ \phi_r + \sum_{j=1}^{p} b_{sj} e_{t+h-j} \right\} e_{t+h-s-r} +$$
$$\sum_{r>}^{p}\sum_{s}^{p} \left\{ \phi_r + \sum_{j=1}^{p} b_{rj} e_{t+h-j} \right\}\left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{t+h-j} \right\} e_{t+h-s-r} + e_t$$

or

$$x_k = \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i} \left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j} \right\} e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p} \left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j} \right\}\left\{ \phi_r + \sum_{j=1}^{p} b_{sj} e_{k-j} \right\} e_{k-s-r} +$$
$$\sum_{r>}^{p}\sum_{s}^{p} \left\{ \phi_r + \sum_{j=1}^{p} b_{rj} e_{k-j} \right\}\left\{ \phi_s + \sum_{j=1}^{p} b_{sj} e_{k-j} \right\} e_{k-s-r} + e_t$$

The forecast can be expressed as

$$x_k = \sum_{i=1}^{p}\sum_{i=1}^{k-1}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + e_t \qquad (96)$$

and the h-steps ahead forecast error is

$$x_k - \hat{x}_k = \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r}$$
$$+ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + e_t$$
$$\sum_{i=0}^{\infty}\prod_{j=1}^{i}(\phi_2 + b_{21}e_{k-2j+1} + b_{22}e_{k-2j})\left(\theta_2 e_{k-2i-2} + \theta_1 e_{k-2i-1} + e_{k-2i}\right) + \left\{\theta_2 e_{k-2} + \theta_1 e_{k-1} + e_k\right\}$$

(97)

Substituting equation (97) in equation (3) and simplifying, we get

First term $\quad E(x_m - \hat{x}_m)^2 = h_m$

Second term $\quad -2E(x_m - \hat{x}_m)\sum_{k=m+1}^{n} a_k(x_k - \hat{x}_k) =$

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \left\{\begin{array}{l} \sum_{i=1}^{p}\sum_{i=1}^{\infty}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t-sj} \\ + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t+h-s-r} + \\ \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{t+h-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{t+h-j}\right\}e_{t+h-s-r} + e_t \end{array}\right\}$$

$$= -2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-si} = -2\sum_{k=m+1}^{n}\phi^{(k-m)/s}h_k$$

$$-2\hat{e}_m \bullet \sum_{k=m+1}^{n} a_k \sum_{s<r \text{ and } s>}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r}$$

$$= \sum_{k=m+1}^{n} 4a_k\phi_r\phi_s h_k \qquad for\ s + r = k - m$$

$$= -2\hat{e}_m \bullet \sum_{k=m+1}^{p} e_k = 0$$

Third term

$$E\left\{\begin{array}{l}\sum_{i=1}^{p}\sum_{i=1}^{k-m}\prod_{j=1}^{i}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{t-sj} + \sum_{s<}^{p}\sum_{r}^{p}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}\left\{\phi_r + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + \\ \qquad \sum_{r>}^{p}\sum_{s}^{p}\left\{\phi_r + \sum_{j=1}^{p} b_{rj}e_{k-j}\right\}\left\{\phi_s + \sum_{j=1}^{p} b_{sj}e_{k-j}\right\}e_{k-s-r} + e_t\end{array}\right\}^2$$

$$= \sum_{k=m+1}^{n} a_k^2\left\{\sum_{s=1}^{p}\phi_s^2\sigma^2 + \sum_{j=1, s\neq j}^{p} b_{sj}{}^2\sigma^2 + e_{k-j} + \sum_{j=1}^{p} 3b_{jj}{}^2 h\sigma^4\right\}$$

which simplifies to

$$-2\sum_{k=m+1}^{n} h_k\phi_s^{(k-m)/s} + \sum_{k=m+1}^{n} 4a_k\phi_r\phi_s h_k + \sum_{k=m+1}^{n} a_k^2\left\{\sum_{s=1}^{p}\phi_s^2 h_k + \sum_{j=1, s\neq j}^{p} b_{sj}{}^2 h_k + e_{k-j} + \sum_{j=1}^{p} 3b_{jj}{}^2 h_k^2\right\} = 0 \qquad (98)$$

Differentiating equation (98) with respect to $a_k$ and setting the result to zero, we obtain

$$\frac{d}{da_k} disp = -2\sum_{k=m+1}^{n} h_k(\phi_s^{(k-m)/s} + \sum_{k=m+1}^{n} 2a_k\phi_r\phi_s) +$$

$$2a_k\left\{\sum_{s=1}^{p}\phi_s^2 h_k + \sum_{j=1,s\neq j}^{p} b_{sj}{}^2 h_k + \sum_{j=1}^{p} 3b_{jj}{}^2 h_k^2\right\} = 0 \qquad (99)$$

Solving for $a_k$, we get

$$\hat{a}_k = \frac{\sum_{k=m+1}^{n} h_k(\phi_s^{(k-m)/s} + \sum_{k=m+1}^{n} 2a_k\phi_r\phi_s)}{\left\{\sum_{s=1}^{p}\phi_s^2 h_k + \sum_{j=1,s\neq j}^{p} b_{sj}{}^2 h_k + e_{k-j} + \sum_{j=1}^{p} 3b_{jj}{}^2 h_k^2\right\}}$$

The missing value estimate is therefore given by

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k(x_k - \hat{x}_k)$$

$$= \sum_{i=1}^{p}\hat{\phi}_i x_{t-i} + \sum_{i=1}^{p}\sum_{j=1}^{p}\hat{b}_{ij}x_{t-i}\hat{e}_{t-j} + \sum_{k=m+1}^{n}\frac{\sum_{K=m+1}^{n} h_k(\hat{\phi}_s{}^{(k-m)/s} + 2\sum_{s\neq r}^{p}\hat{\phi}_s\hat{\phi}_r)}{\left\{\sum_{s=1}^{p} h_k\hat{\phi}_s^2 + \sum_{j=1,s\neq j}^{p}\hat{b}_{sj}{}^2 + \sum_{j=1}^{p}\hat{b}_{jj}{}^2 h_k\right\}}(x_k - \hat{x}_k)$$

**Corollary**

For p=1, we have the bilinear model BL (1, 0, 1, 1). The optimal linear estimate is given by

$$x_m^* = \hat{\phi}_1 x_{m-1} + \hat{b}_{11}x_{m-1}\hat{e}_{m-1} + \sum_{k=m+1}^{n}\frac{\hat{\phi}^{(k-m)}}{(1+\hat{\phi}_1{}^2 + 3\hat{b}_{11}{}^2 h_k)}(x_k - \hat{x}_k)$$

Finally, one observation that we can make about the derived optimal linear estimates is that for pure bilinear time series models, the estimate of missing values is the one-step-ahead forecast error. For the other models, the estimate is also a function of observations after the missing value point which are given weights depending on how close they are to the  to the missing value point.

# CHAPTER FOUR

# ESTIMATION OF MISSING VALUES (RESULTS)

## 4.1 Introduction

In this section, the results of the estimates obtained using the optimal linear estimation, artificial neural networks and exponential smoothing methods based on data generated from bilinear models with different innovations are given. The data with normally distributed innovation was simulated from the models: BL(0,0,1,1), BL (0,0,2,1), and BL(1,0,1,1). The data simulated for student t-distribution included BL(0,0,1,1), BL(0,0,2,1) and BL(1,0,1,2). For GARCH distribution the data was simulated from the models: BL(0,0,1,1), BL(1,0,1,1) and BL(0,0,2,1). The R software was used to generate the bilinear random variables. The first 1000 observations were discarded to reduce the influence of the initial data value used in the simulation. One hundred samples of size 500 were generated for each model and missing values created at positions 48, 293 and 496. The mean absolute deviation and mean squared errors were calculated for each model used. Simulation results are given in Tables 4.1-4.9 and Figures 2-10.

## 4.2 Time series plots of the bilinear models based on simulated data

The data generated were plotted in graphs as depicted in Figures 2-10. These graphs are characterized by sharp outbursts. This is clearly evident in the graph of BL (0, 0, 1, 1). Sharp outburst is one of the characteristics of bilinear time series. For bilinear time series with student-t distributions, the range of the values is also large.

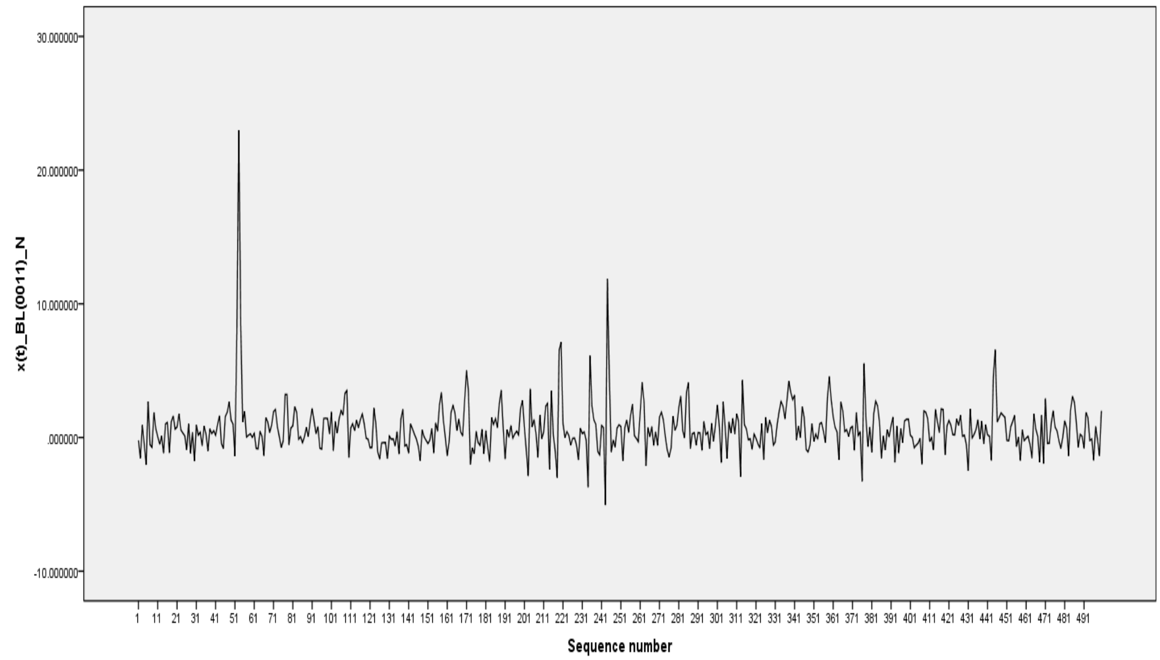



**Figure 2: BL (0, 0, 1, 1) with Normally Distributed Innovations**

Figure 2 displays a classic example of a bilinear time series data. It is evident that there are sharp out-bursts at position 50, 241 and 451. This is a pure bilinear series with the coefficient of the bilinear term, $b_{11} = 0.2$. It is also evident that the series is stationary since it has a constant trend.

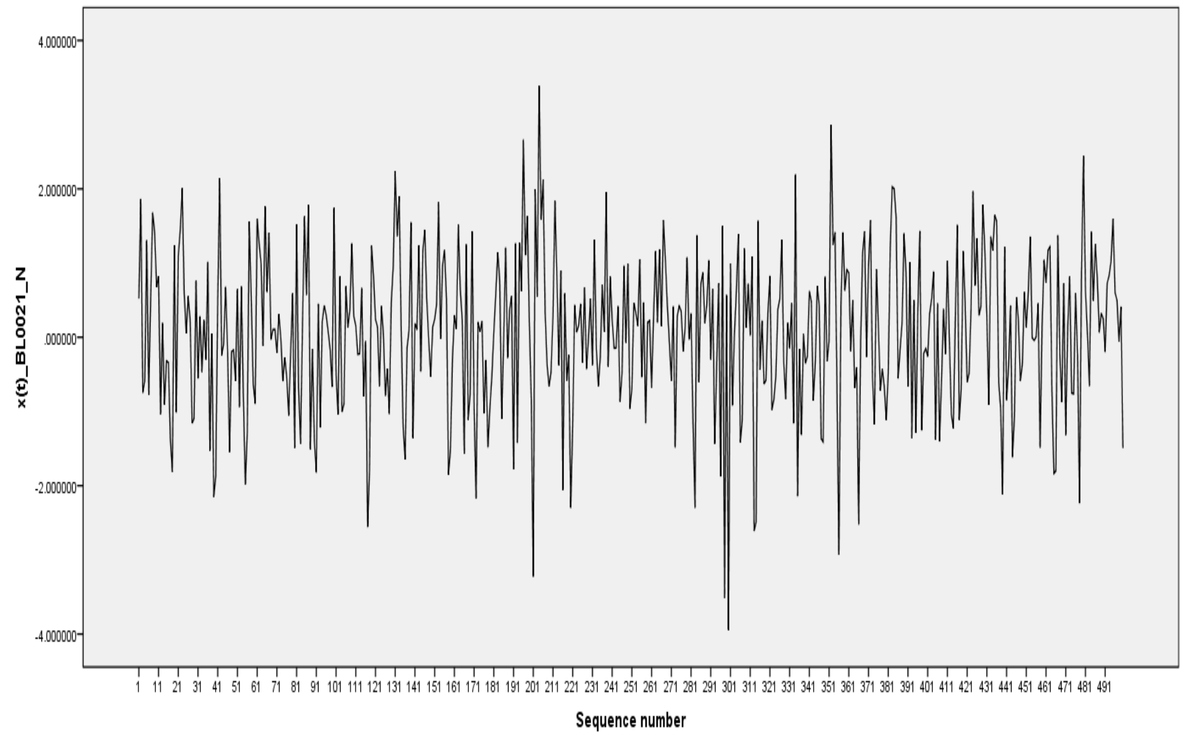



**Figure 3: BL (0, 0, 2, 1) with Normally Distributed Innovations**

Figure 3 displays a graph of a pure bilinear time series with more frequent outburst. The series has a constant trend hence it is stationary.

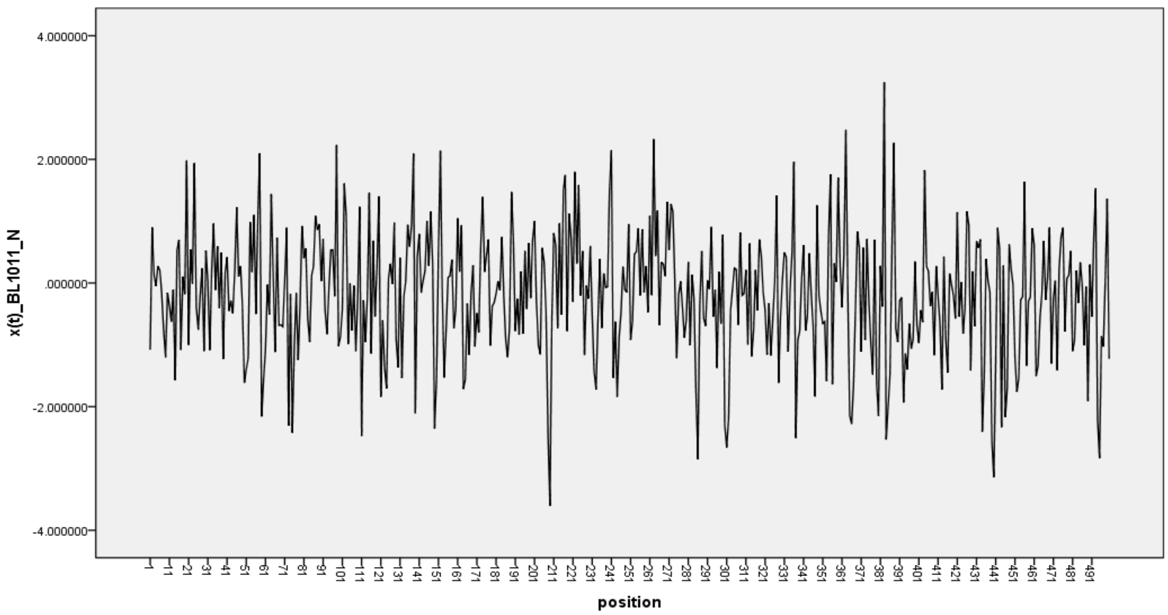



**Figure 4: BL (1, 0, 1, 1) with Normally Distributed Innovations**

The series is not only stationary but the outbursts are less conspicuous.

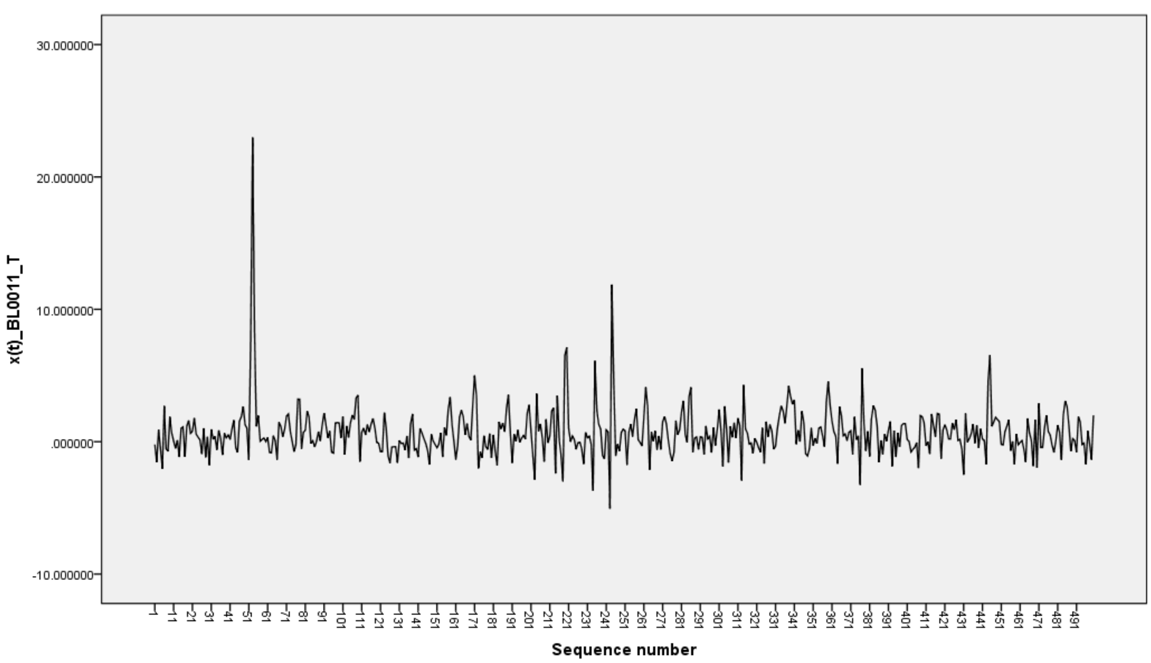



**Figure 5: BL (0, 0, 1, 1) with t-Distributed Innovations**

Figure 5 displays a bilinear time series with a few outbursts which are very conspicuous. This graph is similar to the one of the pure bilinear time series given in Figure 2 with normal innovations. The series has a constant trend hence it is stationary.

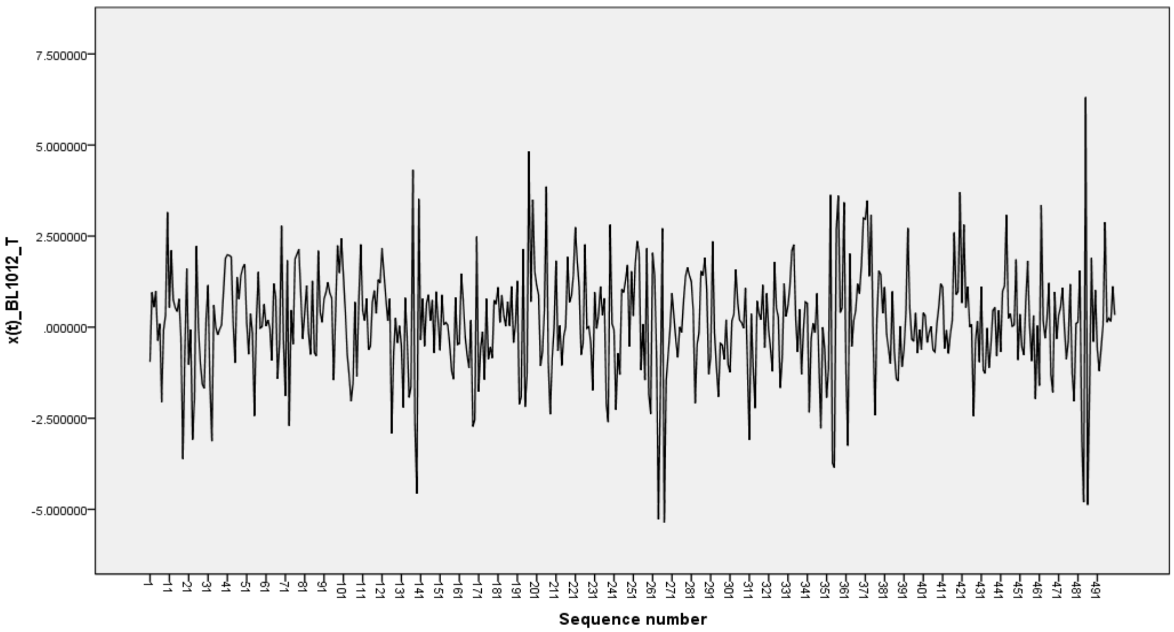



**Figure 6: BL (1, 0, 1, 2) with t-Distributed Innovations**

It is observable that the series is stationary and has sharp outburst of opposite signs. This is different from the other bilinear time series graphs discussed above.

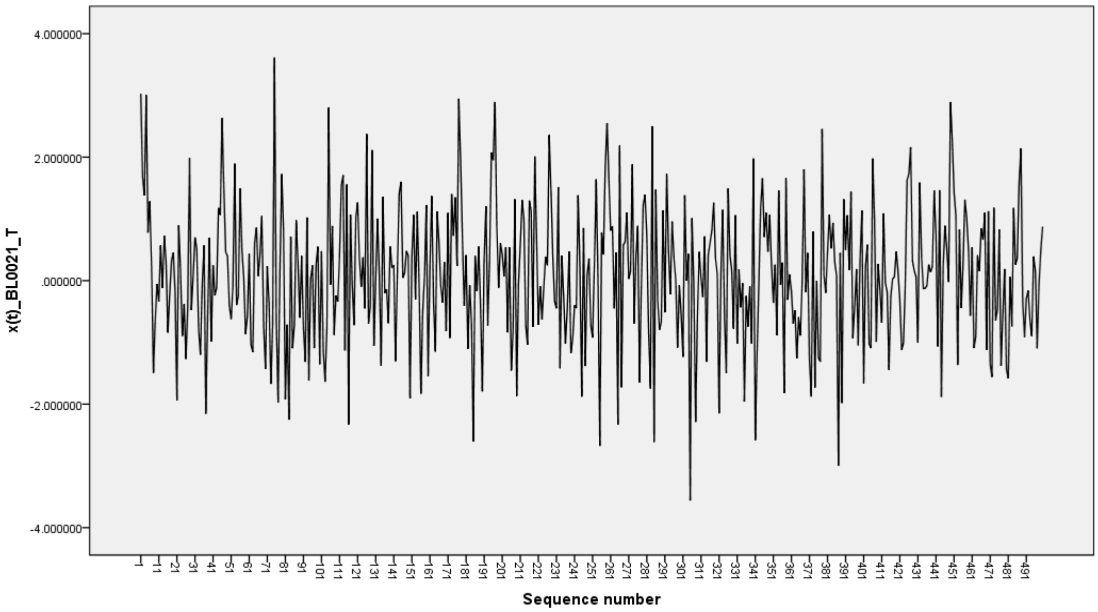



**Figure 7: BL (0,0,2,1) with t-Distributed Innovations**

Figure 7 displays the graph of a bilinear time series with more frequent outbursts of opposite signs. This figure is similar to the graph of BL (0, 0, 2, 1) model with normal distribution.

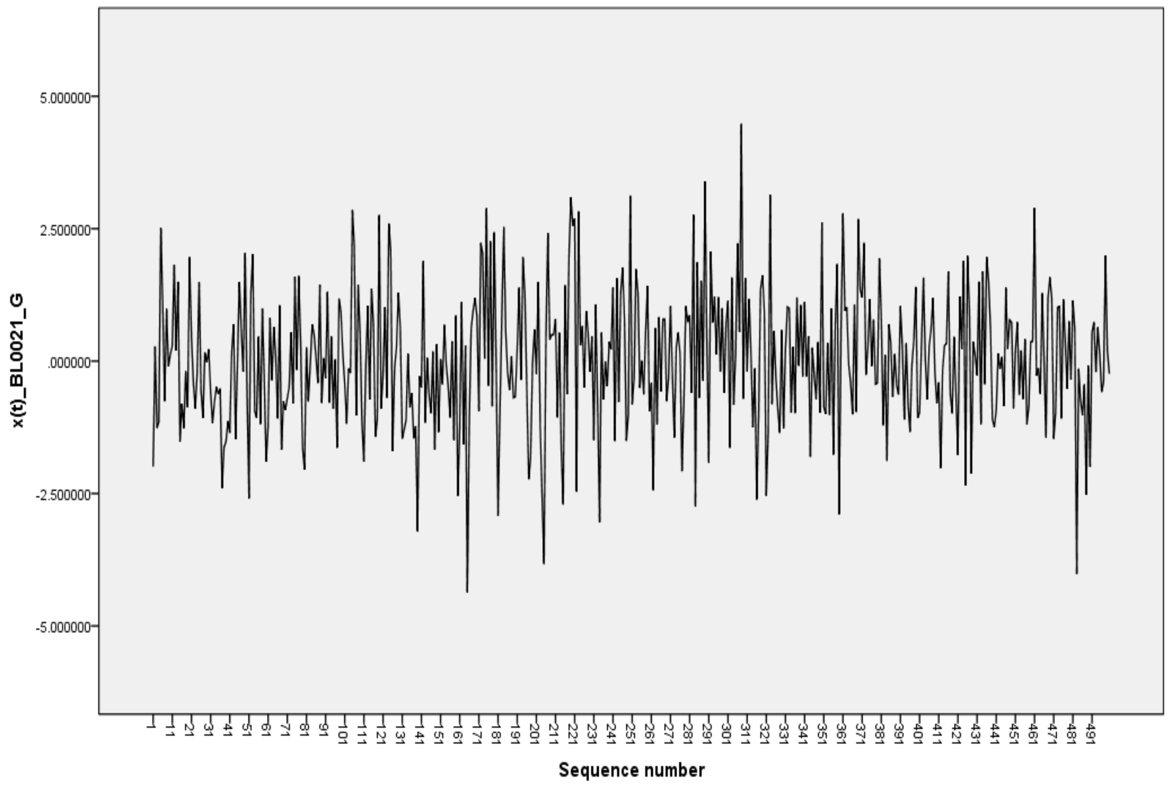



**Figure 8: BL(0, 0, 2, 1) with GARCH Distributed Innovations**

Figure 8 displays the graph of BL (0, 0, 2, 1). It has numerous sharp outburst of opposite signs. It is similar to the graph of BL (0, 0, 2, 1) with the student-t and normal innovations.

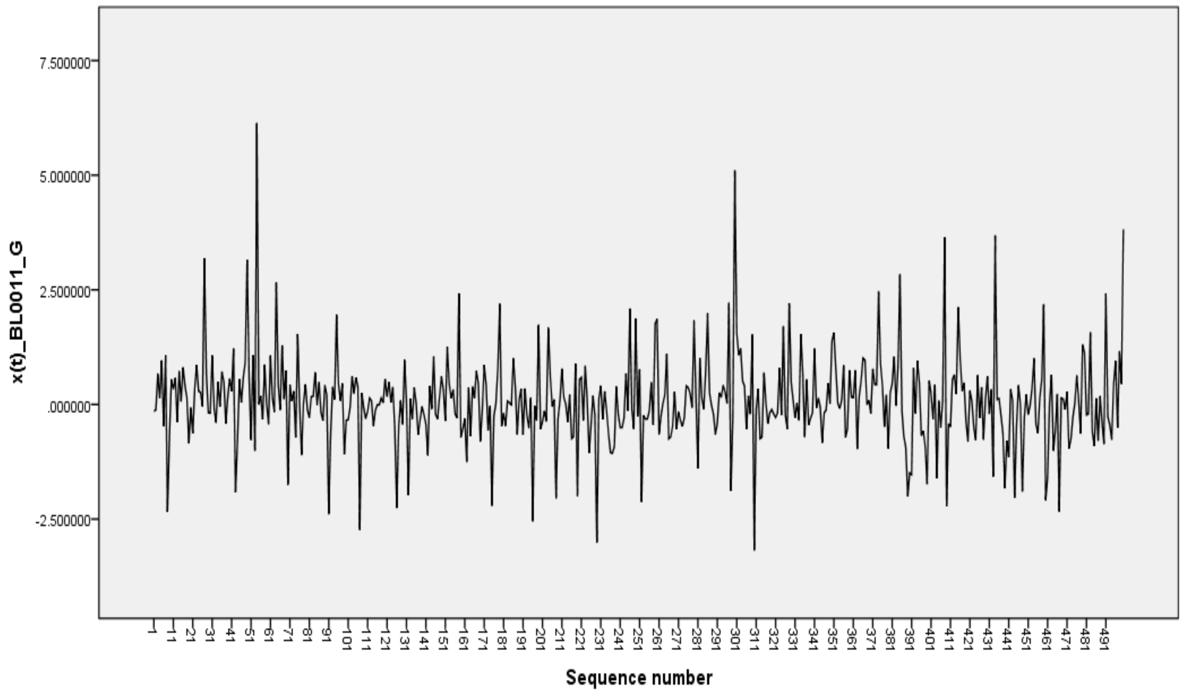



**Figure 9: BL (0, 0, 1, 1) with GARCH Innovations**

Figure 9 has numerous sharp outbursts of the opposite signs at various positions in the data. It is slightly different from BL (0, 0, 1, 1) with either the normal innovations or the student-t innovations since it has more sharp outbursts. These again are of opposite signs.

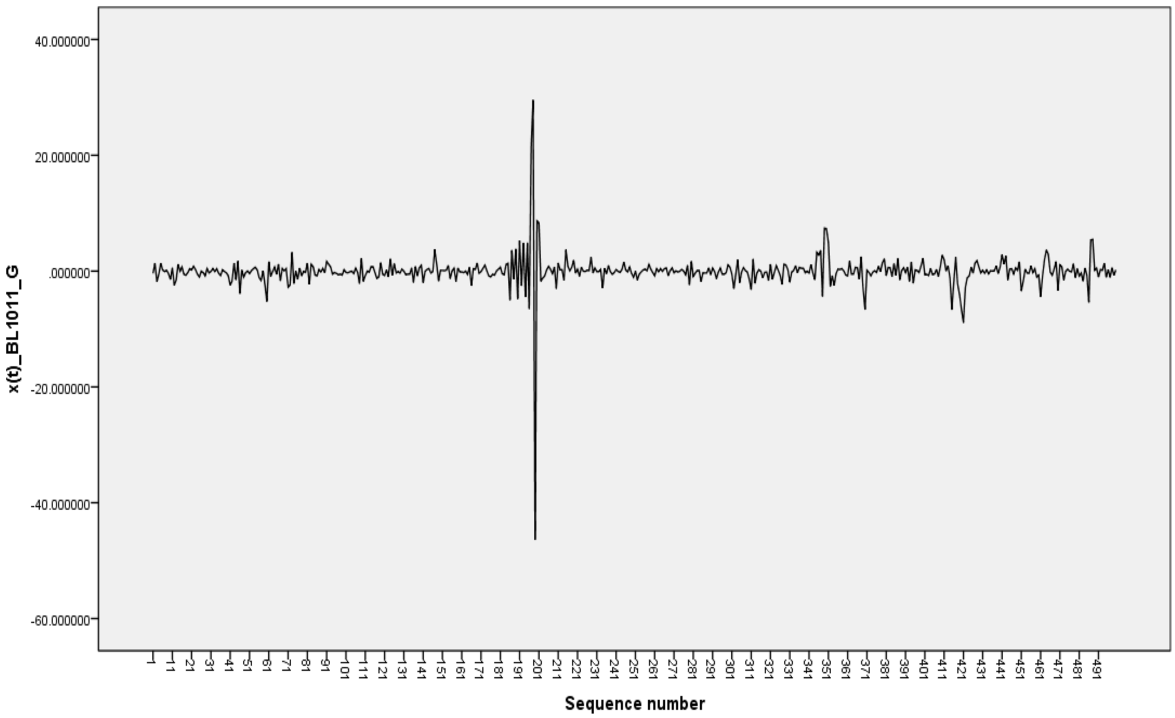



**Figure 10: BL (1, 0, 1, 1) with GARCH Innovations**

This graph is characterized by one conspicuous outburst at position 201. The data is stationary. The zero trends in the graph imply that the missing value estimate should be close to zero. This confirms one of the assumptions made in generating the data that the time series data are stationary. This implies that whereas the student-t and normal distributions are close in structure they are different from those of GARCH distributions. An interesting observation is that the time series models for normal and student distributions have similar structure which is quite different from that of the GARCH distributions. Hence estimates of missing values for normal and student-t distribution should have some common characteristics which are quite distinct from those of the GARCH distribution.

### 4.3 Efficiency measures for bilinear models with different distributions

Data was analyzed using several software and the results obtained are summarized in Tables (1-9). From the analysis, the study found that the method for imputing missing values was correlated with the probability distribution of the innovation sequence of the data. This is similar to the findings of Musial, et al. (2011) who used different nonparametric methods to estimate missing values and concluded that each method exhibited advantages and drawbacks, and that the choice of an approach largely depends on the properties of the underlying time series and the objective of the research. More detailed analysis for each table is given below.

**Table 1: Efficiency Measures for BL (0, 0, 1, 1) with normal innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 0.764572 | 0.84293 | 0.762066 | 1.033225 | 1.223745 | 1.054167 |
| 293 | 0.887287 | 0.900142 | 0.908151 | 1.166085 | 1.259637 | 1.215906 |
| 496 | 0.949875 | 0.93157 | 0.952156 | 1.497507 | 1.363035 | 1.474815 |
| Total | 2.601734 | 2.674642 | 2.622373 | 3.696817 | 3.846417 | 3.744888 |
| Mean | 0.867245 | 0.891547 | 0.874124 | 1.232272 | 1.282139 | 1.248296 |

From Table 1, it is evident that the OLE estimates had the lowest mean square error (MSE =1.232272) among all the estimates of the missing values for the different missing data points positions, followed by EXP smoothing estimates (MSE=1.248296). Estimates based on ANN had the highest mean square error (MSE=1.282139). This implies that OLE estimates were the most efficient estimators for the bilinear time series model BL (0, 0, 1, 1) with normal errors followed by the EXP estimates while ANN estimates were the least efficient. It is also evident that for all the estimators, the position of the missing value had a negative correlation with the efficiency of the estimates obtained.

The estimates for position 48 were based on data in the neighborhood of 48 for OLE and 48 for ANN. This also applied to the data point 293 and 496 where data points in the neighborhood of 293 and 496 were used respectively. It can be noted that the efficiency of the estimates generally did not improve with the sample size about the point of the missing value.

**Table 2: Efficiency Measures for BL (1, 0, 2, 1) with normal innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 0.792612 | 1.135251 | 0.98159 | 1.042752 | 2.620982 | 1.542025 |
| 293 | 0.759908 | 0.870468 | 0.811869 | 0.906356 | 1.602965 | 1.078504 |
| 496 | 0.803375 | 0.862982 | 0.932996 | 0.976211 | 1.214815 | 1.369218 |
| Total | 2.355895 | 2.868701 | 2.726455 | 2.925319 | 5.438762 | 3.989747 |
| Mean | 0.785298 | 0.956234 | 0.908818 | 0.975106 | 1.812921 | 1.329916 |

From Table 2, it is clear that the OLE estimates of missing values were the most efficient (MSE=0.975106) for the different missing data point positions. This was followed by EXP smoothing estimates (MSE=1.329916). It is also evident that the size of the set of values used to estimate the missing values had an effect on the efficiency of the estimates obtained. All estimates for data points 48 were less efficient than estimates obtained at data points 293. However, as the data point was increased to 496, the efficiency generally did not improve.

**Table 3: Efficiency Measures for BL (1, 0, 1, 1) with normal innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 0.842466 | 0.946487 | 0.872985 | 1.122368 | 1.57565 | 1.296111 |
| 293 | 0.893107 | 0.871257 | 0.907073 | 1.239053 | 1.246586 | 1.290295 |
| 496 | 0.951015 | 0.948569 | 0.914346 | 1.442964 | 1.46972 | 1.369107 |
| Total | 2.686588 | 2.766313 | 2.694404 | 3.804385 | 4.291956 | 3.955513 |
| Mean | 0.895529 | 0.922104 | 0.898135 | 1.268128 | 1.430652 | 1.318504 |

From Table 3, it can be concluded that based on MSE, the OLE estimates of missing values were the most efficient (MSE=1.268128) for the different missing data point positions followed by EXP smoothing estimates (MSE=1.318504). It is also evident that the size of set of values used to estimate the missing values had a negative correlation with the efficiency of the estimates obtained. All estimates for data points 48 were more efficient than estimates obtained at data points 496.

**Table 4: Efficiency Measures for BL (0, 0, 1, 1) with student-t innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 0.927299 | 1.252003 | 1.190896 | 1.718008 | 2.706119 | 2.631451 |
| 293 | 0.978332 | 0.961104 | 1.107915 | 2.178498 | 1.802398 | 2.35478 |
| 496 | 0.897052 | 1.188884 | 1.14698 | 1.347372 | 2.922418 | 2.977324 |
| Total | 2.802683 | 3.401991 | 3.445791 | 5.243878 | 7.430935 | 7.963555 |
| Mean | 0.934228 | 1.133997 | 1.148597 | 1.747959 | 2.476978 | 2.654518 |

From Table 4, it can be concluded that based on MSE, the OLE estimates of missing values were the most efficient (MSE=1.747959) for the different missing data point positions followed by ANN estimates (MSE=2.476978). It is also evident that the size of the set of values used to estimate the missing values had a positive correlation with the efficiency of the estimates obtained. All estimates for data points 48 were less efficient than estimates obtained at data points 496.

**Table 5: Efficiency Measures for BL (0, 0, 2, 1) with student-t innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 0.804529 | 0.992834 | 0.82809 | 1.211077 | 1.741109 | 1.293227 |
| 293 | 0.671182 | 0.800902 | 0.720927 | 0.914355 | 1.297314 | 1.088236 |
| 496 | 0.654064 | 0.64545 | 0.680498 | 0.836314 | 0.8832 | 0.863463 |
| Total | 2.129775 | 2.439186 | 2.229515 | 2.961746 | 3.921623 | 3.244926 |
| Mean | 0.709925 | 0.813062 | 0.743172 | 0.987249 | 1.307208 | 1.081642 |

From Table 5, it can be concluded that based on MSE, the OLE estimates of missing values were the most efficient (MSE=0.987249) for the different missing data point positions followed by EXP smoothing estimates (MSE=1.081642). It is also evident that the size of the set of values used to estimate the missing values had a positive correlation with the efficiency of the estimates obtained for all the estimators. All estimates for data points 48 were less efficient than estimates obtained at data points 293 and 496. The estimates became more efficient as the position of missing data increased.

**Table 6: Efficiency Measures for BL (1, 0, 1, 2) with Student-t innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 1.05252 | 1.05009 | 1.18255 | 2.49592 | 2.08018 | 2.97117 |
| 293 | 1.25008 | 1.06888 | 1.40457 | 2.63802 | 2.35919 | 3.44825 |
| 496 | 1.41045 | 0.94823 | 1.45937 | 4.1928 | 1.94371 | 4.3707 |
| Total | 3.71305 | 3.0672 | 4.04649 | 9.32674 | 6.38308 | 10.79012 |
| Mean | 1.237683 | 1.0224 | 1.34883 | 3.108913 | 2.127693 | 3.596707 |

From Table 6, it can be concluded that based on MSE, the ANN estimates of missing values were the most efficient (MSE=2.127693) for the different missing data point positions followed by OLE estimates (MSE=3.108913). It is also evident that the size of the set of values used to estimate the missing values had mixed results on the efficiency of the estimates obtained. Generally the estimates showed a negative correlation with the size of the position of the missing value.

We can conclude that OLE estimates generally give the optimal estimates of missing values and that the size of the position of the missing observation had a positive correlation with the efficiency of the estimates obtained.

**Table 7: Efficiency Measures for BL (0, 0, 1, 1) with GARCH innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 0.804529 | 0.992834 | 0.82809 | 1.211077 | 1.741109 | 1.293227 |
| 293 | 0.661487 | 0.788114 | 0.719507 | 0.851441 | 1.20582 | 1.023331 |
| 496 | 0.654064 | 0.64545 | 0.680498 | 0.836314 | 0.8832 | 0.863463 |
| Total | 2.12008 | 2.426398 | 2.228095 | 2.898831 | 3.83013 | 3.180021 |
| Mean | 0.706693 | 0.808799 | 0.742698 | 0.966277 | 1.27671 | 1.060007 |

From Table 7, it can be observed that based on MSE, OLE estimates of missing values were the most efficient (MSE=0.966277) for the different missing data point positions followed by EXP smoothing estimates (MSE=1.06007). It is also evident that the size of the set of values used to estimate the missing values had a positive correlation with the efficiency of the estimates obtained. The higher the position of missing data, the more efficient the estimates obtained.

**Table 8: Efficiency Measures for BL (0, 0, 2, 1) with GARCH innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 1.203124 | 1.146296 | 1.260539 | 2.3505 | 2.264935 | 2.609242 |
| 293 | 1.136119 | 0.978307 | 1.110822 | 3.409059 | 1.783063 | 2.251697 |
| 496 | 0.964016 | 0.887704 | 1.045276 | 2.108792 | 1.778419 | 2.052729 |
| Total | 3.303259 | 3.012307 | 3.416637 | 7.868351 | 5.826417 | 6.913668 |
| Mean | 1.101086 | 1.004102 | 1.138879 | 2.622784 | 1.942139 | 2.304556 |

From table 8, it can be concluded that based on MSE, the ANN estimates of missing values were the most efficient (MSE=1.942139) for the different missing data point positions followed by EXP estimates (MSE=2.304556). It is also evident that efficiency of the estimates improved as the position of the missing value increased.

**Table 9: Efficiency Measures for BL (1, 0, 1, 1) with GARCH innovations**

| MISSING | MAD | | | MSE | | |
|---|---|---|---|---|---|---|
| POSITION | OLE | ANN | EXP | OLE | ANN | EXP |
| 48 | 1.136119 | 0.978307 | 1.110822 | 3.409059 | 1.783063 | 2.251697 |
| 293 | 0.964016 | 0.887704 | 1.045276 | 2.108792 | 1.778419 | 2.052729 |
| 496 | 1.276248 | 1.003744 | 1.300622 | 3.591108 | 2.326084 | 3.63698 |
| Total | 3.376383 | 2.869755 | 3.45672 | 9.108959 | 5.887566 | 7.941406 |
| Mean | 1.125461 | 0.956585 | 1.15224 | 3.03632 | 1.962522 | 2.647135 |

From Table 9, it can be concluded that based on MSE, the ANN estimates of missing values were the most efficient (MSE=1.962522) for the different missing data point positions followed by EXP estimates (MSE=2.647135). It is also evident that efficiency of the estimates had a mixed correlation as the position of the missing data increased.

We can conclude that for bilinear time series with GARCH innovations, the ANN estimates were generally more efficient than the estimates obtained from the other estimators. It is evident that the efficiency of the estimator used was correlated with the distribution of the innovations of the bilinear time series. Further, the efficiency of the OLE estimators for normally distribution had a positive negative correlation with the position of the missing data.

For student distribution, the OLE estimators were the most efficient but not as remarkable as in the normal distribution. For the bilinear time series with GARCH innovations, the ANN estimators were generally more efficient than the other estimators used. The estimators had a positive negative correlation with the position of the missing data a positive negative correlation with the position of the missing data.

The figures below (Figure 11-19) show the deviations between the actual data and the estimated values of BL (1, 0, 1,1) for normally distributed innovations.

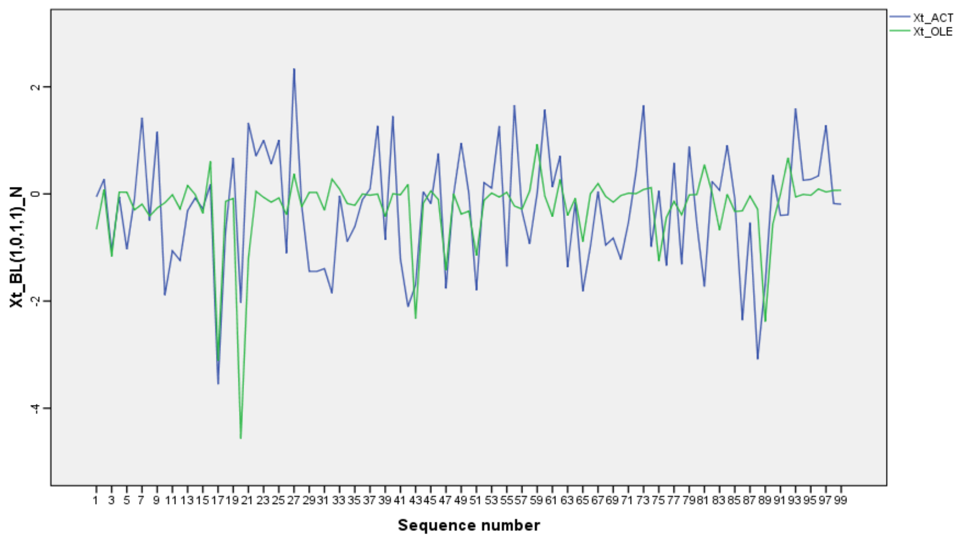



**Figure 11: Actual Values vs OLE Estimates for with Missing Value at 48 for BL (1, 0, 1, 1).**

The pattern of the actual values and the estimated values are similar indicating the efficiency of the results.

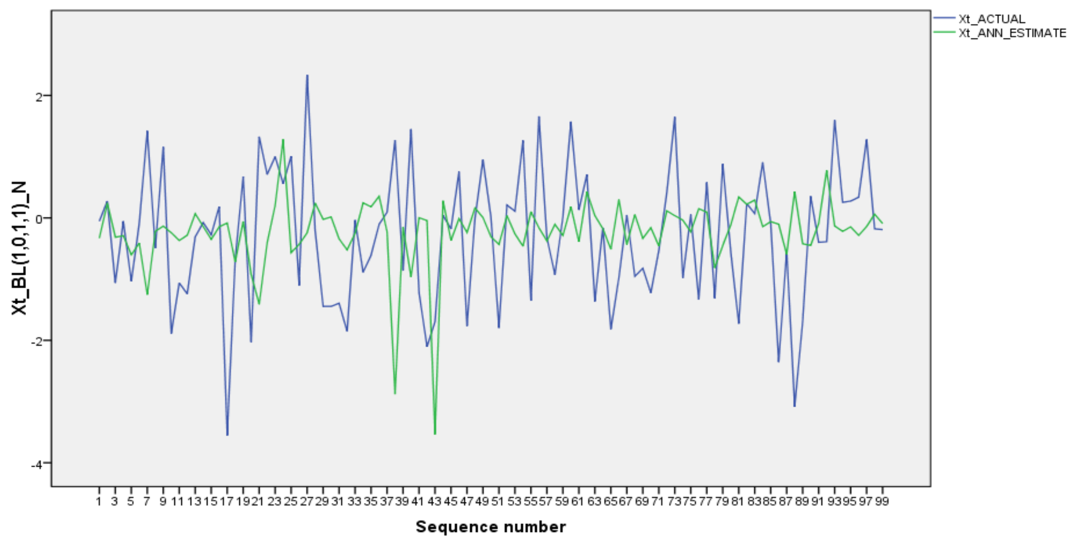



**Figure 12: Actual Values vs ANN Estimates for Missing Value at 48 BL (1, 0, 1, 1)**

The patterns of the two graphs are similar. There is a higher deviation between actual values and the actual values.

For

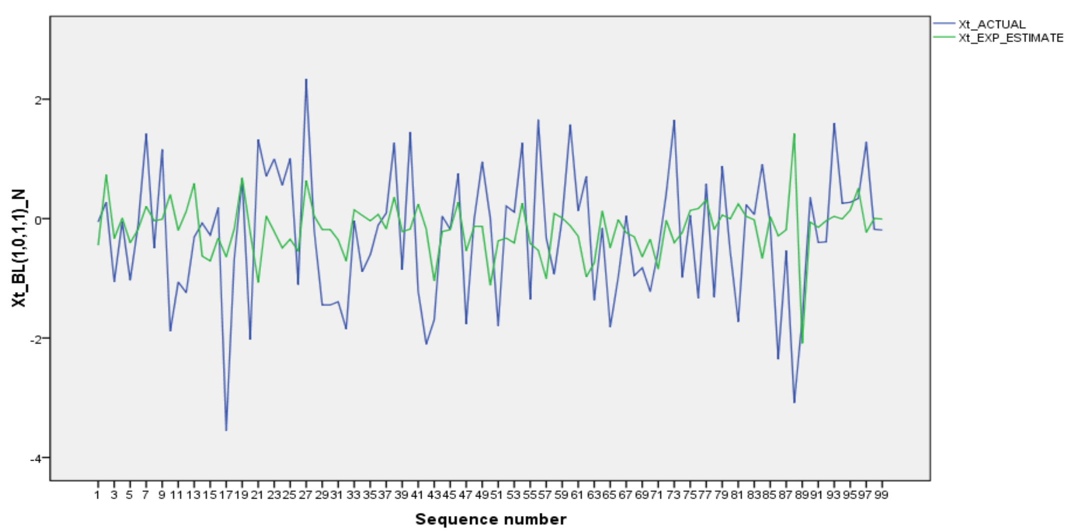



**Figure 13: Actual Values vs EXP Estimates for with Missing Value at 48 for BL (1, 0, 1, 1).**

There is higher deviation between the actual values and the estimated values as evidenced by higher disparity between the two graphs..

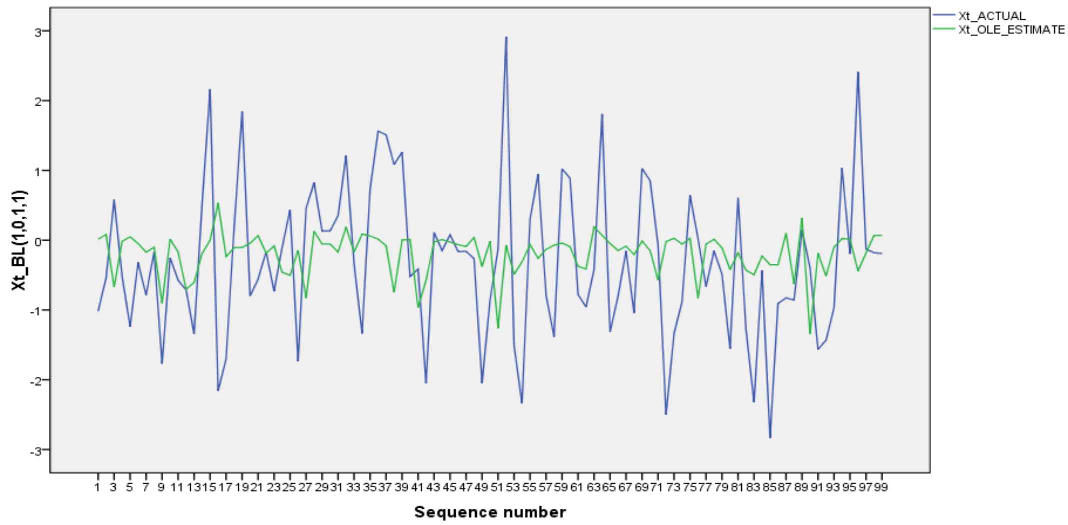



**Figure 14: Actual Values vs OLE Estimates for Missing Value at 293 for BL (1, 0, 1, 1)**

There is higher deviation between the actual values and the estimate.

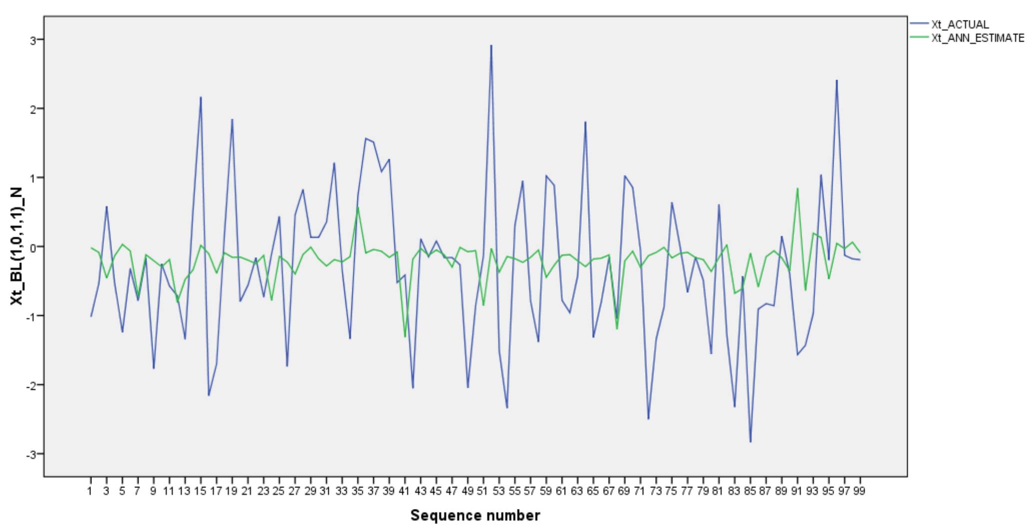



**Figure 15: Actual Values vs ANN Estimates for Missing Value at 293 for BL (1, 0, 1, 1)**

There is a significant disparity between the actual values and the estimates obtained.

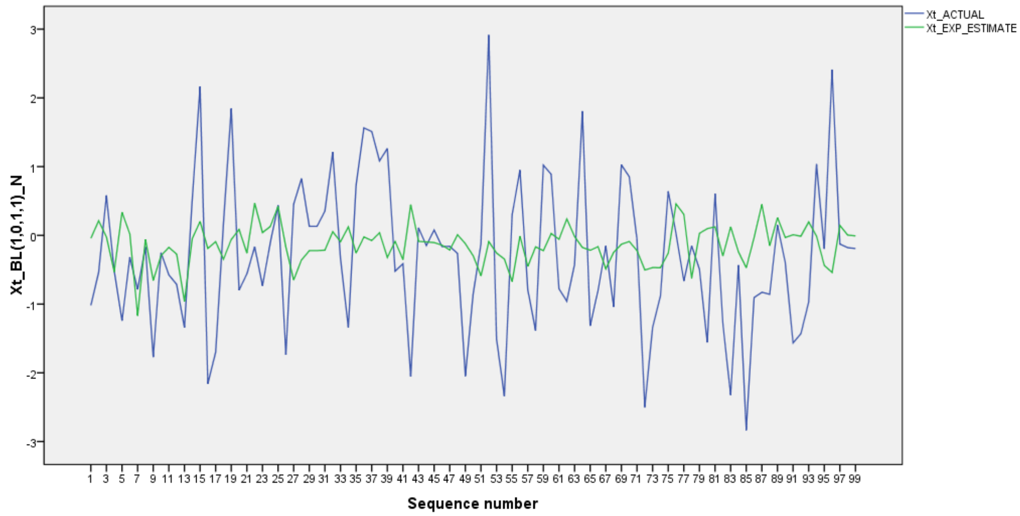



**Figure 16: Actual Values vs EXP Estimates for Missing Value at 293  for BL (1, 0, 1, 1)**

There is high disparity between the actual values and the estimated values.

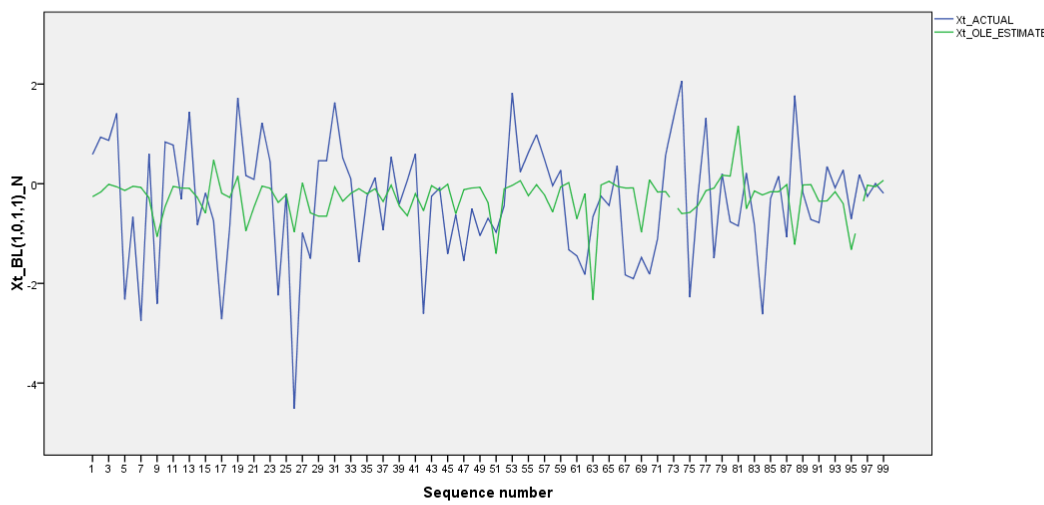



**Figure 17: Actual Values vs OLE Estimates for Missing Value at 496 BL (1, 0, 1, 1)**

There appears to be high disparity between the observed values and the estimated values.

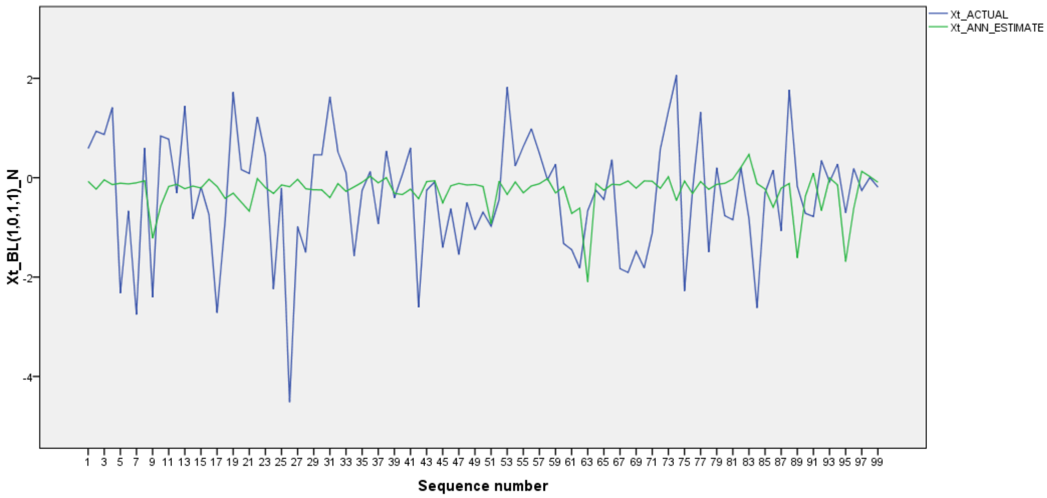



**Figure 18: Actual Values Vs ANN Estimates for Missing Value at 496 for BL (1, 0, 1, 1)**

Higher deviation between the estimated values and the actual values is an indicator that the estimates are less efficient.

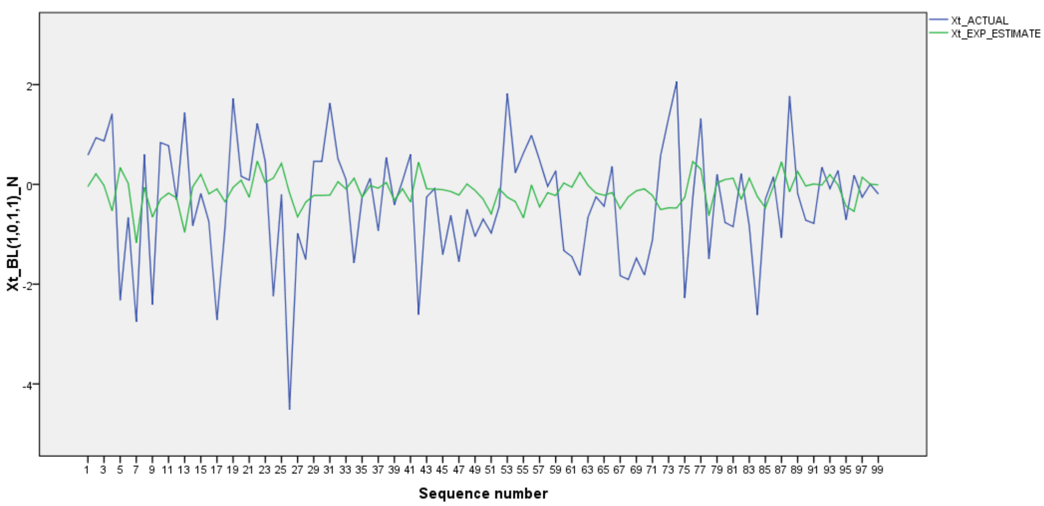



**Figure 19: Actual Values vs EXP Estimates for Missing Value at 496 for BL (1, 0, 1, 1)**

The estimates have the shape of the moving average values. We can conclude that for bilinear time series data with normal errors, the OLE estimates gave the most efficient estimates of the missing values. It is also evident that the efficiency of the estimates had a negative correlation worsens with the position of the missing data. For the EXP and ANN the results were mixed. In all the above cases ANN estimates had the least efficient estimates obtained.

# CHAPTER FIVE

# CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Conclusions

This study had five objectives; the first three was concerned with the derivation of the estimators; the fourth one dealt with nonparametric estimators while the last one involved the comparison of the efficiency of the estimates obtained. To measure the efficiency and accuracy of the estimates, simulation studies were conducted. One hundred samples of size 500 each was generated using the R software and missing values were created at random at data positions 48, 293 and 496. The mean square error was used to determine the efficiency of the estimates obtained using three techniques: optimal estimates, artificial neural networks and exponential smoothing.

For pure bilinear time series model BL (0, 0, p, p), the missing value was found to be equivalent to the one–step-ahead forecast based on the lagged observations before the point of the missing value. All the observations beyond the point of missing values played no role in estimating the missing value. For the general bilinear time series models BL (p, 0, p, p), the estimate not only consisted of the forecasted value based on the previous observations but in a few the observations after the missing value contributed to the estimate. Weights were attached to these observations with data closest to the point after the missing having higher more weights attached.

Since different distributions and missing positions were used, it was imperative to determine how these factors affected the efficiency of the estimates of missing values. The study found that artificial neural networks gave more efficient estimates for GARCH distribution compared to the optimal linear estimates and exponential smoothing. In fact, optimal linear estimates were the least efficient. For normally distributed data, optimal linear estimates were most efficient compared to both the artificial neural network estimates and exponential smoothing estimates.

As far as the relationship between the position of missing value and efficiency of the estimator is concerned, the study had mixed findings. The estimates based on the ANN generally improved when the position of the missing value was large. That is, estimates at position 496 were efficient than estimates at position 48 or 293. For GARCH distribution, cases of non-convergence of the estimates when the position of the missing value was low (48) were frequent, in fact in some cases it was 45%. This meant that only data of size 55% was used to compute the performance measures. This also occurred for missing value points at 293 and 496 where we had several failed convergences. When the position of the missing value was closer to 500, the efficiency of the estimates improved. For normally distributed data, position of the missing value gave mixed results on the efficiency of the estimates.

## 5.2. Recommendations

The study recommends that for bilinear time series data with normal and student t-innovations, OLE estimates be used in estimating missing values. For bilinear time series data with GARCH distribution, ANN estimates may be used. The study found that OLE estimates do not improve with the position of the missing data. That is the further the missing data point is from the first data collected, the less efficient the estimate becomes.

## 5.3 Recommendation for further research

- More research needs to be done on whether the accuracy of an imputation method depends on the distribution of the data.
- A more elaborate research should be done to compare the efficiency of several imputation methods such as K-NN, Kalman filter and estimating functions, genetic algorithms, besides the three used in this study.
- Derivation of estimates of missing value for ARMA time series models with different distributions such student-t, normal and GARCH distributions should be undertaken.
- Derivation of estimates of missing values for bilinear time series with infinite variance should be undertaken.
- Application of the derived estimates to real data

## 6.0 REFERENCES


Abdalla, M., and Marwalla, T. (2005). The use of Genetic Algorithms and neural networks to approximate missing data. *Computing and Informatics vol. 24, 5571-589.*

Abowd, J., Stinson, M., & Benedetto, G. (2006). Final Report to the Social Security Administration on the SIPP/SSA/IRS Public Use File Project. Tech. rept. U.S. Census Bureau Longitudinal Employer-Household Dynamics Program. Retrieved at http://www.bls.census.gov/sipp/synth.

Abraham, B. (1981). Missing observations in time series. *Comm. Statist. A-Theory Methods.*

Abraham, B.; Thavaneswaeran, A. (1991). A Nonlinear Time Series and Estimation of missing observations. *Ann. Inst.Statist.Math. Vol. 43, pp 493-504.*

Abrahantes, J. C., Sotto, C., Molenberghs, G., Vromman, G., and Bierinckx, B. (2011). A comparison of various software tools for dealing with missing data via imputation .*J. Stat. Comput. Simulation* 81, No. 11, 1653-1675.Am Stat. 2007 Feb; 61(1): 79–90.

Anders,U; Korn, O and Schmitt, C. (1998). “Improving the pricing of options: A neural network approach,” *Journal of Forecasting,vo*l. 17, no. 5-6, pp. 369-388.

Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P. (2003). "Modeling and Forecasting Realized Volatility,"Econometrica, 71, 529-626.

Anderson B., and Moore, J. (1979). Optimal Filtering. Prentice Hall, Englewood Cliffs, N.J. Appl. Prob., 14, 4ll-415.*Applied Probability*. Chapman and Hall series.

Background Facts on Economic.Statistics. (2003). A note on improving imputations using time series forecasts.13.Statistics Sweden, Publication Services, SE 701 89 ÖREBRO, Sweden. Doi http://www.scb.se/statistik/OV/AA9999/2003A01/AA9999_2003A01_BR_X100ST0313.pdf.

Beveridge, S. (1992). Least Squares Estimation of Missing Values in Time Series. *Biometrika*, *72, pp. 39-44*.

Bishop, C. M. (1995). Neural Networks for pattern recognition. *Oxford: Oxford University Press*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics.* Vol.3, issue 3..Elsevior.

Box, G. P., Jenkins,G. M., and Reinsel, G. (2008). Time Series Analysis: *Forecasting and Control, fourth* ed, Wiley, New Jersey.

Brockwell, P. J., and Davis, R. A. (1991). Time Series: Theory and Methods.*New York*, USA : Springer-Verlag.

Burden, F.R., Brereton, R. G ., and Walsh, P.T.(1997). Cross-validatory selection of test and validation sets in multivariate calibration and neural networks as applied to spectroscopy. *Analysts*, 122: 1015-1022.

Campozan, L.,Sanchez, E. Aviles, 3, Samaniego, E. (2014). Evaluation of infilling methods for time series of daily precipitation and temperature: The case of the Ecuadorian Andes. *Maskana*, Vol. 5. Revista semestral de la DIUC 99.

Cao, Y., Poh K, L., and Cui, W, J. (2008). A non-parametric regression approach for missing value imputation microarray. *Intelligent Information Systems*. pages 25–34.

Cheng and Titterington, D. M. (1994): Neural networks: review from a statistical perspective. *Statist. Science*. Vol 9, NO 1, 2-30.

Cheng, P. (1994). Non-parametric estimation of mean of functionals with data missing at random. *Journal of the American statistical association*, 89, 81-87.

Chng, E.S., Chen, S., Mulgrew, B. (1996). Gradient radial basis function networks for nonlinear and nonstationary time series prediction.*IEEE Transactions on Neural Networks* 7 (1), 191–194.

Cortiñas J,A., Sotto, C; Molenberghs, G; Vromman, G. (2011). A comparison of various software tools for dealing with missing data via imputation. *Bart Bierinckx* pages 1653-1675.

Guegan, D., and Pham, D. T. (1989). A Note on the Estimation of the Parameters of the Diagonal Bilinear Model by the Method of Least Squares. *Scandinavian Journal of Statistics* Vol. 16, No. 2 (1989), pp. 129-136 Published by Wiley.

Damsleth, E. (1979). Interpolating Missing Values in a Time Series. *Scand J Statist.*, 7, data via the EM algorithm," *J. Royal Statist. Soc.*, ser. B, vol.

De Gooijer, J.C. (1989) Testing Nonlinearities in World Stock Market Prices, *Economics Letters* v31, 31-35.

Degiannakis, S. and Xekalaki, E. (2004). Autoregressive Conditional Heteroscedasticity Models: A Review. *Quality Technology and Quantitative Management*1 (2),271-324.

Dempster, A., Laird, and Rubin. D. (1977). “Maximum likelihood from incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society.* Series B (Methodological).

Demuth, H.and Beale, M. (1997). Neural Network TOOLBOX for Use with Matlab.User’s Guide. http://www.mathworks.com Edition. The Math Works Inc. Engineering System Design, 5:3, 149-160.

Denk, M and Weber, M. (2011). Avoid Filling Swiss Cheese with Whipped Cream: Imputation Techniques and Evaluation Procedures for Cross-Country Time Series. IMF working paper.WP/11/I51.

Ferreiro, O. (1987). Methodologies for the estimation of missing observations in time series. *Statist. Probab. Lett.*, 5, 65-69.

Fiedler, R. and Schodl, P. (2008). Data Imputation and Estimation for the Austrian Register-Filter. *Cambridge,* UK : Cambridge University Press.

Flood, I and Kartam, N. (1994). Neural Networks in Civil Engineering. II: Systems and Application. *Journal of computing in civil engineering.* 8:2(149).

Fung, D.S.C. (2006). Methods for the Estimation of Missing Values in Time Series http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1063&context=theses retrieved on 11/2/2006.

Gaver, D.G and Lewis , P.A. (1980). First order autoregressive Gamma sequences and point processes. *Adv. Applic, Prob*. 12,727-745.

Golyandina, N., and Osipov, E. (2007). The Catterpillar"-SSA method for analysis of time series with missing values, *J. Stat. Plan.Inf*. 137 2642-2653.

Golyandina, N.; Nekrutkin, V.;Zhigljavsky, A. (2001). Analysis of Time Series Structure: SSA and Related Techniques, Chapman& Hall/CRC, London.

Gomez, I. A.; Burton, D. M; and Love, H. A. (1995). Imputing Missing Natural Resource Inventory Data and the Bootstrap. *Natural Resource Modelling,* 9(4), 299-328.

Granger, C. W., and Anderson, A.P. (1978a). An Introduction to Bilinear Time Series model. Vandenhoeck and Ruprecht: Guttingen.

Gupta, A., and Lam, M. S. (1995).Estimating Missing Values Using Neural Networks. *Journal of the Operational Research Society* (1996) **47,** 229–238. doi:10.1057/jors.1996.21

Haggan, V., and Ozaki, T. (1981). Modelling non-linear random vibrations in using anamplitude - dependent autoregressive time series model. Bíometríka, 68, 189 - L96.

Hannan, E J. (1982). "A Note on Bilinear Time Series Models", *Stochastic Processes and their Applications*, vol. 12, p. 221-24.

Harvey, A. C. (2001). Forecasting, Structural Time Series Models and the Kalman, Filter, ‘Company Press, Cambridge.

Harvey, A. C. ; Pierse, R. G. (1984). Estimating missing observations in Economic time series. *Journal of the American Association, 79 (385), 125-131*. Midwest Political Science Association.

Hassani, H. (2007).Singular Spectrum Analysis: Methodology and comparison. *Journal of Data Science*, 5(21), pp. 239-257.

Haykin, S. (1999) Neural networks: A comprehensive foundation. *American Journal of Political Science* Vol. 54, No. 2, April 2010, Pp. 561–581

Hellem, B. T. (2004). Lsimpute accurate estimation of missing values in micro Array data with least squares method. *Nucleic Acids,* 32,e34.

Hirano, K ; Imbens, G, W.; Ridder (2002). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71,1161 -1189.

Honaker, J. ; King, G.(2010). What to Do about Missing Values in Time-Series? *American Journal of Political Science* Vol. 54, No. 2, April 2010, Pp. 561–581

Hornick, K. Stinchcombe, M.; White, M. (1989). Multi Layer feedforward networks are universal approximations. *Neural networks, 2, 359-366.*

Horton N J and Kleiman, K P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistical Association* Vol. 61, No. 1 79. DOI: 10.1198/000313007X172556.

Howitt, P. (1988). "Business Cycles with Costly Search and Recruiting", *Quarterly Journal of Economics*, vol.103(1), p. 147-65.

Irvine, K.N., and Eberhardt, A.J., (1992). Multiplicative, seasonal ARIMA models for Lake Erie and Lake Ontario water levels.*J.;Tíme Series Anal*.,10, 33-39.

Jacobs, P.A. and Lewis, P. A. V. (1977). A mixed autoregressive moving average exponential sequence and point processes (EARMA 7,1).*adv. Appl. Prob*.,9,87.

Jones, R H. (1985). Time series analysis with unequally spaced data, *Handbook of Statistics*, vol. 5 (eds, E.J. Hannan, P.R Krishnaiah and M.M. Rao), 157-177, North Holland, Amsterdam.

Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics,* 22, 389 -395.

Junninen, H : Niska H, Tuppuraine K,Ruuskanen, J. : Kolchmainen, M.(2004). Methods for imputation of missing values in quality data sets. *Atmost Environ* 38(18):2895-2907.

Kaastra, I. and Boyd, M. (1995). Designing a neural network for forecasting financial and economic time series. *Neurocomputing, 10, 215-236.*

Kaiser, J. (I2012). Algorithm for Missing Values Imputation in Categorical Data with Use of Association Rules. *ACEEE*.

Karunanithi, N., Grenney, W.J., Whitley, D., Bovee, K., 1994. Neural networks for river flow prediction. *Journal of Computing in Civil Engineeirng* 8 (2), 201–220.

Keenan, D M. (1985). A Tukey Non-additivity Type Test for Time Series Nonlinearities." Biometrika, 72 39-44.

Khashei, M. and, Mehdi, B. (2010). An artificial neural network (p,d,q) model for time series forecasting.*Expert Systems with Applications* 37 (2010) 479–489.

Kim, J K and Fuller. W. (2004). Fractional hot deck imputations. *Biometrika* 91, 559-578.

Kohn, R ; Ansley, C.F. (1986). Estimation, Prediction, and Interpolation for ARIMA models with missing data. *Journal of the American Statistical Association, 81(395) 751-761.Lecture*

Kondrasho,D. and Ghil, M.(2006). Spatio-temporal lling of missing points in geophysical data sets, *Nonlinear Proc. Geoph*. 13(2006) 151-159.

Kuligowski, R. J., and Barros. P. A.(1998a). Using Artificial Neural Networks To Estimate Missing Rainfall Data. *Journal of The American Water Resources Association* Vol. 34, No.6 American water Resources Association.

Kumar, K. (1986). On The Identification o Some Bilinear Time Series Model .*Journal of Time Series Analysis* Volume 7, Issue 2, Pages 117–122.

Kyung, W., Kim, L., Billard and Basawa, V. (1990). Estimation for The First-Order Diagonal Bilinear Time Series Model. *Journal of Time Series Analysis*. Volume 11, Issue 3, pages 215–229DOI: 10.1111/j.1467-9892.1990.tb00053.

Lawrence, R. (1997). Using Neural Networks to forecast stock market prices. R. Lawrence, "Using
University of Manitoba. https://people.ok.ubc.ca/rlawrenc/research/Papers/nn.pdf.

Lawrence, A.J. and P.A.W. Lewis (1980). The exponential autoregressive moving average EARMA(p,q) process. *./RS^S, Ser. B*, 42, I50 - 161.

Ledolter, J. (2008). Time Series Analysis Smoothing Time Series with Local Polynomial Regression on Time series. *Communications in Statistics—Theory and Methods*, 37: 959–971.

Liu, J., and Brockwell P. J. 1988. "On the general bilinear time series model." *Journal of Applied probability*, 25, 553–564.

Ljung, G. M. (1989). A note on the Estimation of missing Values in Time Series. *Communications in statistics simulation 18(2), 459-465.*

Luca, D. G. (2006). Forecasting Volatility using High-Frequency Data. *Statistica Applicata* Vol. 18, n. 2. 407-422.

Luceno, A. (1997). Estimation of Missing Values in Possibly Partially Nonstationary Vector Time Series. *Biometrika* Vol. 84, No. 2 (Jun., 1997), pp. 495-499. Oxford University Press.

Maier, H. R., and Dandy, G. C. (1996b). The use of artificial neural networks for the prediction of water quality parameters. *Water Resource. Research* 32(4), 1013-1022.

Maravall, A. (1983). "An application of nonlinear time series forecasting",*Journal of Businesa 6 Econamic Statistics*, 1, 66-74.

Masters, T. (1993). Practical neural network recipes in C++London: Academic Press.

Mcknight, E, P ; McKnight, M, K; Sidani, S.; Figueredo, A.(2007). Missing data. Guiford New York.

Miller, R.B. and Ferreiro,O. (1984). A strategy to complete a time series with missing, Lecture Notes in Statistics, 25, 251-275, Springer, New York.

Minns, A.W., and Hall, M. J. (1996). Artificial neural networks as rainfall–runoff models. *Hydrological Sciences Journal* 41 (3), 399–417.Models with irregular data. Commun. Statísts.Theory,2l .

Moffat, A.; Papale, D.; Reichstein, M.; Hollinger, D.Y; Richardson A,D.;. Barr A,G; B Eckstein ,C; Braswell B, H; Churkina, G; desay A, R;Falgae E.; Gove , JH.Heimann M; Hui D Javis D. (2014). Forecasting Home sales in the four census regions and the Aggregate US economy using singular spectrum Analysis.

Islam, M. A. (2013). Estimating Volatility of Stock Index Returns by Using Symmetric Garch models .*Middle-East Journal of Scientific Research* 18 (7): 991-999.DOI: 10.5829/idosi.mejsr.2013.18.7.11806.

Moreno, J. J. M., Pol, A. P., and Pilar M., and Gacia, M. P. (2011). Artificial neural networks applied to forecasting time series. *Psicothema* Vol. 23, nº 2, pp. 322-329.

Musial J. P., Verstraete1 M. M., and Gobron N. (2011). Technical Note: Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series.

Nassiuma, D.K and Thavaneswaran, A. (1992). Smoothed estimates for nonlinear time series models with irregular data. *Communications in Statistics-Theory and Methods 21*(8), 2247–2259.

Nassiuma, D. K. (1994). A Note on Interpolation of Stable Processes. *Handbook of statistics,Vol. 5 Journal of Agriculture*, *Science and Technology* Vol.3(1) 2001: 81-8.

Nassiuma, D. K. (1994). Symmetric stable sequence with missing observations. *J.T.S.A. volume 15, page 317*.

Nichols, D.F., and Quinn, D. G. (1982). Random coefficient Autoregressive models: An introduction Operating records," *Int. J. Control*, vol. 40, pp. 121–128, 1984.

Nieto, F. H., & Martinez, J. (1996). A Recursive Approach for Estimating Missing Observations in

Nieto, F. H., & Martfncz, J. (1996). A Recursive Approach for Estimating Missing Observations in An Univariate Time Series. *Communications in statistics Theory* A 25(9), 2101-2116.

Norazian, M. N., Shukri, Y. A., Azam, R. N., & Al Bakri, A. M. M. (2008). Estimation of missing values in air observations. *Lecture notes in Statistics* Vol. 25. Sprínger verlag. New York.

Oba, S., Sato, M.A, Takemasa I, et al. (2003). A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics* 2003; 19(16):2088-2096.

Önder, E., Bayır, F., & Hepsen, A. (2013). Forecasting Macroeconomic Variables using Artificial Neural Network and Traditional Smoothing Techniques. *Journal of Applied Finance and Banking*, *3*(4), 73-104.

Owili, P A., Nassiuma, D., and Orawo, L. (2015a). Efficiency of Imputation Techniques for Missing Values of Pure Bilinear Models with GARCH Innovations, *American Journal of Mathematics and Statistics*, Vol. 5 No. 5. pp. 316-324. doi:10.5923/j.ajms.20150505.13.

Owili, P A., Nassiuma, D., and Orawo, L. (2015b).Imputation of Missing Values for Pure Bilinear Time Series Models with Normally Distributed Innovations." *American Journal of Applied Mathematics and Statistics* 3.5 : 199-205.

Owili, P A., Nassiuma, D., and Orawo, L. (2015c). Estimation of Missing Values for Pure Bilinear Time Series Models with Student-t Innovations, *International Journal of Statistics and Applications*, Vol. 5 No. 6, 2015, pp. 293-301. doi: 10.5923/j.statistics.20150506.05.

Pascal, B. (2005) :Influence of Missing Values on the Prediction of a Stationary Time Series. *Journal of Time Series Analysis.*Volume 26, Issue 4, pages 519–525.

Pashova, L. & Popova, S. (2011). Daily sea level forecast at tide gauge Burgas, Bulgaria using artificial neural networks. *Journal of Sea Research* 66: 154–161.ew York.

Pena, D., & Tiao, G. C. (1991). A Note on Likelihood Estimation of Missing Values in perspective. *Multivariate Behavioral Research*, 33, 545−571.

Pigot, D. T. (2001). A Review of Methods for Missing Data. *Educational Research and Evaluation* Vol. 7, No. 4, pp. 353±383.

Popova, S., Iliev, S., and Trifonov, M. (2014).Neural Network Prediction of The Electricity consumption of Trolleybus and Tram Transport in Sofia City *Latest Trends in Energy, Environment and development.* http://www.wseas.us/elibrary/conferences/2014/Salerno/ENED/ENED-14.pdf

Pourahmadi, M. (1989). Estimation and interpolation of missing values of a stationary time series. *Journal of Time Series Analysis* 10(2), 149–69.

Gupta, P., and Srinivasan, R. (2011*).* Missing Data Prediction and Forecasting for Water Quantity Data. *International Conference on Modeling, Simulation and Control IPCSIT vol.10 IACSIT Press, Singapore.*

Kourentzes, N., and Crone, S. F. (2008). Forecasting high-frequency Time Series with Neural Networks – an analysis of modeling challenges from increasing data frequency.

http://www.DMIN08.20v3.pdf.

Pigott, D. T(2001). A Review of Methods for Missing Data. *Educational Research and Evaluation* 1380-3611/01/0704-353 2001, Vol. 7, No. 4, pp. 353±383

Priestley, M.B. (1980). State dependent models: A general approach to time series analysis. profile data. *Bioinformatics* 2003; 19(16):2088-2096.

Qi, M.and Zhang, G. (2001). "An investigation of model selection criteria for neural network time series forecasting," *European Journal of Operational Research*, vol. 132, no. 3, 2001, pp. 666-6807.0 Resources Bulletin 28 (2), 385–396.

Ripley, B. (1996). Pattern recognition and neural networks. *Cambridge:* Cambridge University Press.

Rubin, D. B. (1987). Multiple imputation for non-response in survey.New York. John Wiley and sons.

Rumelhart, D.E., Hinton, G. E and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323: 533-536.499-508.

Sande, I.G. (1982). Inputation in Surveys: Coping with Reality. The American Statistician.

Sarle, W.S. (1994). Neural networks and statistical models. In: Proceedings of the Nineteenth *Annual SAS Users Group International Conference*,pp. 1538–1550. SAS Institute.

Särndal, C.E. and Lundström, S. (2005). Estimation in Surveys with Nonresponse, Wiley, Chichester, England.

Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. *Ann. Vol 72 of monographs on satatistics.*

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, 33(4), 545-571.

Schoellhamer (2001). Singular spectrum analysis for time series with missing data. AGU publishers. *Statist. Sci. 1, 2-54.Statistical Association* 81, 131–41. series. J.Tíme Series Anal., 10, 149-169).

Sesay, S.A and Subba Rao, T. (1988): Yule Walker type difference equations for higher order moments and cumulants for bilinear time series models. *J. Time Ser. Anal*.9, 385-401.

Shiqing L, Liang, P, And Fukang Zhu. (2015). Inference For A Special Bilinear Time-Series Model. *Journal of Time Series Analysis J. Time. Ser. Anal*.36:61–66.

Subba Rao, T. (1981). On the Theory of Bilinear Time Series Models. *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 43, No. 2. (1981), pp.244-255.

Subba Rao, T., and Gabr, M. M. (1980). A test for non-linearity of stationary time series. Time Series Analysis, 1,145-158.

Subba, Rao, T. ; Gabr, M.M.(1984). An Introduction to Bispectral Analysis and Bilinear Time Series Models. *Lecture notes in statistics, 24.New York. Springer.*

Subba, Rao, T., and da Silva, A. E (1992). Identification Bilinear Time Series Models a BL(p,0,p,1). *Statistica Sinica* 2, 465-478.

Tamura, S., and Tateishi, M.(1997). Capabilities of a four –layed feedback neural networks, Four layers versus three *IEEE. Transactions and neural networks,* 8(2), 251-255.

Teegavarapu R. S. (2012). Spatial interpolation using nonlinear mathematical programming models for estimation of missing precipitation records, *Hydrological Sciences. Journal*, 57:3, 383-406.

Thavaneswaran, A., and Abraham (1987). Recursive estimation of Nonlinear Time series models. *Institute of statistical Mimeo series No 1835*.Time Series. The American statistician, 45(3), 212-213.

Tong, H. (1983). Threshold Models in Non-Linear Time Series analysis. Springer Verlag, Berlin.

Toth, E., Montanari, A., and Brath, A.( 2000.): Comparison of short-term rainfall prediction model for real-time flood forecasting, *J. Hydrol.*,239, 132–147.

Tsay, R. S. (1986). Time series model specification in the presence of outliers. Journal of the American USA : Springer-Verlag Vector Time Series. *Biometrika*, 84(2), 495-499.

Vautard, R and Ghil, M. (1989) Singular spectrum analysis in nonlinear dynamics, with applications to pale climatic time series, with missing data. AMS Subject Classification: 37M10; 62H25 with Missing Observations. *Technometrics,* 22(3), 389-395. N

Vijayakumar, N., and Plale, B. (2007, September). Tracking stream provenance in Complex Event Processing systems for workflow-driven computing. *In EDA-PS Workshop*. Vol. 39, No. 1,(1977), pp. 1-38. Royal Statistical Society.

Wagala, A., Nassiuma, D. K., Islam. A., and Mwangi.,J W. (2012). Volatility Modeling of the Nairobi Securities Exchange Weekly Returns Using the Arch-Type Models . *International Journal of Applied Science and Technology* Vol. 2 No.3;

Wallgren, A. and Wallgren, B. (2007). Register-based Statistics, Wiley, Chichester.*Journal of the Operational Research Society* (1996) **47,** 229–238. doi:10.1057/jors.1996.21.

Welch,G.,and Bishop, G.,(2011). An Introduc-tion to the Kalman Filter."Internet:http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf.

Weiss. A, A. (1984). Arma Models With Arch Errors. *Journal of Time Series Analysis* Volume 5, Issue 2, pages 129–143.

Wenzel, M. & Schröter, J. (2010). Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. *Journal of Geophysical Research* - Oceans 115, C08013, DOI: 10.1029/2009JC005630.

Wold, H.(1954). *A Study in the Analysis of Stationary Time Series*, Second revised edition, with an Appendix on "Recent Developments in Time Series Analysis" by Peter Whittle. *Almqvist and Wiksel*l Book Co., Uppsala.

Yozgatligil. C; Aslan, S; Iyigun, C.; Batmaz, I. (2012). Comparison of missing value imputation methods in time series: the case of Turkish meteorological data . *Theor Appl Climatol* DOI 10.1007/s00704-012-0723-x Springer-Verlag.

Zhang G. P. (2003). Time series forecasting using a Hybrid ARIMA and neural network model. *Neurocomputing* 50 (2003) pages 159–175.

Zhang, G., Patuwo, B. E., and Hu, M.Y. (1998), Forecasting with artificial neural networks: the state of the art, I*nt.J. Forecasting* 14 35–62- L04., 36:3 145-152.

# APPENDIX

## Appendix 1: Program codes used in simulation

```
# The R program BL(0,0,1,1)_Normal
  b1<-0.2
  e<-c()
  e[1]<- rnorm(1)
  x <-c()
  x[1]<-c(e[1])
 set.seed(0006412)
   for (i  in 2:1500) {
            e[i] = rnorm(1)              # generate noise value
            x[i] = b2*x[i-1]*e[i-1]+ e[i]    #  calculate x using the model
 }
t<- x[-1:-1000]
 y<-round(t,7)
z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
 n<-round(z,7)
summary(y)

  y

  n
```

```
 # The R program (BL0011_GARCH(1,1)
 b1 <-  0.9; b2<-0.64; s<-0.03; b3<-0.135;b4<- -0.17
h<-c()
x <-c()
 e<-c()
z<-c()
 e[1]<- rnorm(1)
z[1]<-rnorm(1)
h[1]<-c(sqrt(s))
x[1]<-c(z[1]*h[1])
set.seed(90033134)
 for (i  in 2:1500) {
            e[i] = rnorm(1)                # generate noise value
z[i]=rnorm(1)#generates noise value
h[i]= abs(sqrt(s+b2*e[i-1]^2+b3*h[i-1]))# conditional variance is strictly positive
            x[i] =b4* x[i-1]+b1*x[i-1]*e[i-1]+ z[i]*h[i]    #  calculate x using the model
}
    t<- x[-1:-1000]
  y<-round(t,7)


Y
```

```
# The R program (BL1012_studentErrors)
  b12<-0.4; b1=0.2
  sigma<-1
 h<-c()
  e<-c()
 z<-c()
  e[1:2]<- rt(2,7)
  x <-c()
  x[1:2]<-c(0,   e[1])
 set.seed(02848151)
   for (i  in 3:1500) {
            e[i] = rt(1,7)               # generate noise value
           x[i] = b1*x[i-1]+b12*x[i-1]*e[i-2]+ e[i]    #  calculate x using the model
 }
 t<- x[-1:-1000]
 y<-round(t,7)
z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
 n<-round(z,7)
 Y
```

```
# The R program (BL1011_Normal)
   b1<- 0.1; b2<- -0.2
  e<-c()
  e[1]<- rnorm(1)
  x <-c()
  x[1]<-c(0)
 set.seed(01009821)
   for (i  in 2:1500) {
            e[i] = rnorm(1)               # generate noise value
            x[i] = b1*x[i-1]+b2*x[i-1]*e[i-1]+ e[i]    #  calculate x using the model
 }
 t<- x[-1:-1000]
 y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
 n<-round(z,7)
y
n
```

```
# The R program BL(1,0,1,1)_GARCH(1,1)
 b1 <-  0.9; b2<-0.64; s<-0.03; b3<-0.135;b4<- -0.17
h<-c()
x <-c()
 e<-c()
z<-c()
 e[1]<- rnorm(1)
z[1]<-rnorm(1)
h[1]<-c(sqrt(s))
x[1]<-
c(z[1]*h[1])
set.seed(580173)
 for    (i         in
2:1500) {
          e[i] = rnorm(1)               # generate noise value
z[i]=rnorm(1)#generates noise value
h[i]= abs(sqrt(s+b2*e[i-1]^2+b3*h[i-1]))# conditional variance is strictly positive
          x[i] =b4* x[i-1]+b1*x[i-1]*e[i-1]+ z[i]*h[i]    #  calculate x using the model
}
   t<-        x[-1:-
1000]
  y<-round(t,7)
   summary(y)
```

```
# The R program (BL(0,0,2,1)_Normal_Errors)
    b21<-0.4; b1=0.2
  h<-c()
   e<-c()
  z<-c()
   e[1:2]<- rnorm(2)
  z[1:2]<-rnorm(2)
   x <-c()
   x[1:2]<-c(e[1],e[1])
  set.seed(239)
    for (i  in 3:1500) {
            e[i] = rnorm(1)              # generate noise value
            x[i] =b21*x[i-2]*e[i-1]+ e[i]    #  calculate x using the model
  }
  t<- x[-1:-1000]
  y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
  n<-round(z,7)
Y
```

```
# The R program (BL1011_Normal)
 b1<- 0.1; b2<- -0.2
 e<-c()
 e[1]<- rnorm(1)
 x <-c()
 x[1]<-c(e[1])
set.seed(01009821)
 for (i  in 2:1500) {
        e[i] = rnorm(1)            # generate noise value
        x[i] = b1*x[i-1]+b2*x[i-1]*e[i-1]+ e[i]    #  calculate x using the model
}
t<- x[-1:-1000]
y<-round(t,7)
z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
n<-round(z,7)
summary(y)
```

# The R program (BL0011_student)

```
 b11<-0.4
 e<-c()
z<-c()
 e[1:2]<- rt(2,7)
 x <-c()
 x[1:2]<-c( e[1], x[1]*e[1]+e[2])
set.seed(01018848)
 for (i  in 3:1500) {
          e[i] = rt(1,7)                # generate noise
value
          x[i]   =b11*x[i-1]*e[i-1]+   e[i]          #
calculate x using the odel
 }
t<- x[-1:-1000]
y<-round(t,7)
z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
n<-round(z,7)
y
```

```
# The R program General_BL(1,0,1,2)_student)
  b12<-0.3; b1=0.2; b11<-0.1
  e<-c()
 z<-c()
  e[1:2]<- rt(2,27)
  x <-c()
  x[1:2]<-c(0,   e[1])
 set.seed(007188)
   for (i  in 3:1500) {
            e[i] = rt(1,27)                # generate noise value
            x[i]  =b11*  x[i-1]*e[i-1]+b12*x[i-1]*e[i-2]+  e[i]       #   calculate  x  using  the
model
 }
 t<- x[-1:-1000]
 y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
 n<-round(z,7)
 y

 n
```

```
# The R program (BL0021_GARCH_Errors)
 b2<-0.6; b1=0.7 ; b3=0.2;c4<-0.5;c5<-0.6;c3<-0.3;s<-0.9
h<-c()
 e<-c()
 z<-c()
 e[1:2]<- rnorm(2)
z[1:2]<-rnorm(2)
h[1]<-c(sqrt(s))
h[2]<-c(sqrt(1+b3*e[1]^2+c3*h[1]))
 x <-c()
 x[1:2]<-c(0,   e[1])
 set.seed(6032609)
  for (i  in 3:1500) {
          e[i] = rnorm(1)              # generate noise value
 z[i]=rnorm(1)#generates noise value
 h[i]= (sqrt(s+b3*e[i-1]^2+c3*h[i-1]))# conditional variance is strictly positive
          x[i] = b2*x[i-2]*e[i-1]+ z[i]*h[i]    #  calculate x using the model
 }
 t<- x[-1:-1000]
 y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
 v<-round(z,7)
Y
```

```
# The R program (BL(0,0,1,1)_Normal)
   b1<-0.2
   e<-c()
   e[1]<- rnorm(1)
   x <-c()
   x[1]<-c(e[1])
  set.seed(0006412)
    for (i  in 2:1500) {
             e[i] = rnorm(1)                 # generate noise value
             x[i]  =  b2*x[i-1]*e[i-1]+  e[i]       #   calculate  x  using  the
model
  }
t<- x[-1:-1000]
  y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-
min(y))
  n<-round(z,7)
 summary(y)

y

n
```

```
# The R program (BL0011_GARCH(1,1)
b1 <-  0.9; b2<-0.64; s<-0.03; b3<-0.135;b4<- -0.17
h<-c()
x <-c()
e<-c()
z<-c()
e[1]<- rnorm(1)
z[1]<-rnorm(1)
h[1]<-c(sqrt(s))
x[1]<-c(z[1]*h[1])
set.seed(90033134)
for (i  in 2:1500) {
        e[i] = rnorm(1)             # generate noise value
z[i]=rnorm(1)#generates noise value
h[i]= abs(sqrt(s+b2*e[i-1]^2+b3*h[i-1]))# conditional variance is strictly positive
        x[i] =b4* x[i-1]+b1*x[i-1]*e[i-1]+ z[i]*h[i]    #  calculate x using the model
}
   t<- x[-1:-1000]
  y<-round(t,7)


Y
```

```
# The R program (BL1012_studentErrors)

b12<-0.4; b1=0.2

sigma<-1

h<-c()

e<-c()

z<-c()

e[1:2]<- rt(2,7)

x <-c()

x[1:2]<-c(0,   e[1])
```

```
set.seed(02848151)

  for (i  in 3:1500) {

        e[i] = rt(1,7)              # generate noise value

        x[i] = b1*x[i-1]+b12*x[i-1]*e[i-2]+ e[i]     #   calculate x using the
model

 }

 t<- x[-1:-1000]

 y<-round(t,7)

z<-0.1+0.8*(y-min(y))/(max(y)-min(y))

 n<-round(z,7)

 Y
```

```
# The R program (BL1011_Normal)
   b1<- 0.1; b2<- -0.2
  e<-c()
  e[1]<- rnorm(1)
  x <-c()
  x[1]<-c(0)
 set.seed(01009821)
   for (i  in 2:1500) {
          e[i] = rnorm(1)             # generate noise value
          x[i] = b1*x[i-1]+b2*x[i-1]*e[i-1]+ e[i]    #  calculate x using the model
 }
 t<- x[-1:-1000]
  y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
  n<-round(z,7)
y
n
```

```
# The R program (BL(1,0,1,1)_GARCH(1,1)
 b1 <-  0.9; b2<-0.64; s<-0.03; b3<-0.135;b4<- -0.17
h<-c()
x <-c()
 e<-c()
z<-c()
 e[1]<- rnorm(1)
z[1]<-rnorm(1)
h[1]<-c(sqrt(s))
x[1]<-
c(z[1]*h[1])
set.seed(580173)
 for     (i          in
2:1500) {
              e[i] = rnorm(1)                 # generate noise value
z[i]=rnorm(1)#generates noise value
h[i]= abs(sqrt(s+b2*e[i-1]^2+b3*h[i-1]))# conditional variance is strictly positive
              x[i] =b4* x[i-1]+b1*x[i-1]*e[i-1]+ z[i]*h[i]    #  calculate x using the model
}
   t<-         x[-1:-
1000]
  y<-round(t,7)
   summary(y)
```

# The R program (BL0021_Normal_Errors)

```
b21<-0.4; b1=0.2
h<-c()
e<-c()
z<-c()
e[1:2]<- rnorm(2)
z[1:2]<-rnorm(2)
x <-c()
x[1:2]<-c(e[1],e[1])
set.seed(239)
for (i  in 3:1500) {
        e[i] = rnorm(1)              # generate noise value
        x[i]  =b21*x[i-2]*e[i-1]+  e[i]       #   calculate  x  using  the
model
}
t<- x[-1:-1000]
y<-round(t,7)
z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
n<-round(z,7)
Y
```

# The R program (BL(1,0,1,1)_Normal)

```
b1<- 0.1; b2<- -0.2
e<-c()
```

```
  e[1]<- rnorm(1)
  x <-c()
  x[1]<-c(e[1])
 set.seed(01009821)
   for (i  in 2:1500) {
            e[i] = rnorm(1)                 # generate noise value
            x[i] = b1*x[i-1]+b2*x[i-1]*e[i-1]+ e[i]    #  calculate x using the
model
  }
 t<- x[-1:-1000]
 y<-round(t,7)
 z<-0.1+0.8*(y-min(y))/(max(y)-
min(y))
 n<-round(z,7)
 summary(y)
# The R program (BL0011_student)
    b11<-0.4
    e<-c()
   z<-c()
    e[1:2]<- rt(2,7)
    x <-c()
    x[1:2]<-c( e[1], x[1]*e[1]+e[2])
```

```
  set.seed(01018848)
      for (i  in 3:1500) {
               e[i]  =  rt(1,7)                       #  generate
noise value
               x[i]   =b11*x[i-1]*e[i-1]+   e[i]           #
calculate x using the odel
    }
    t<- x[-1:-1000]
   y<-round(t,7)
   z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
    n<-round(z,7)

y
# The R program General_BL(1,01,2)_student)
    b12<-0.3; b1=0.2; b11<-0.1
    e<-c()
   z<-c()
    e[1:2]<- rt(2,27)
    x <-c()
    x[1:2]<-c(0,   e[1])
  set.seed(007188)
     for (i  in 3:1500) {
              e[i] = rt(1,27)                # generate noise value
```

```
        x[i] =b11* x[i-1]*e[i-1]+b12*x[i-1]*e[i-2]+ e[i]     #   calculate x using the
model
 }
 t<- x[-1:-1000]
 y<-round(t,7)
z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
 n<-round(z,7)
 Y

 N

 # The R program (BL(0,0,2,1)_GARCH_Errors)
  b2<-0.6; b1=0.7 ; b3=0.2;c4<-0.5;c5<-0.6;c3<-0.3;s<-0.9
 h<-c()
   e<-c()
  z<-c()
   e[1:2]<- rnorm(2)
 z[1:2]<-rnorm(2)
 h[1]<-c(sqrt(s))
 h[2]<-c(sqrt(1+b3*e[1]^2+c3*h[1]))
   x <-c()
   x[1:2]<-c(0,   e[1])
  set.seed(6032609)
    for (i  in 3:1500) {
```

```
        e[i] = rnorm(1)           # generate noise value
  z[i]=rnorm(1)#generates noise value
  h[i]= (sqrt(s+b3*e[i-1]^2+c3*h[i-1]))# conditional variance is strictly positive
        x[i] = b2*x[i-2]*e[i-1]+ z[i]*h[i]    #  calculate x using the model
  }
  t<- x[-1:-1000]
  y<-round(t,7)
  z<-0.1+0.8*(y-min(y))/(max(y)-min(y))
  v<-round(z,7)
Y
```

## Appendix II: Moments of distributions used

### Moments of the standard normal distribution

$$E(e_{t)}) = 0$$
$$E(e_t^2) = \sigma^2$$
$$E(e_t^3) = 0$$
$$E(e_t^4) = 3\sigma^4$$

### Moments of the GARCHDistribution

*Let* $\psi$ *be the* $\sigma -$ *field generated by* $e_{t-1}, e_{t-2}, .....e_1$. *The conditional* exp*ectations or moments higher powers of* $e_t$, *are given by*

$$E(e_t / \psi_{t-1}) = E(\eta_t \sqrt{h_t} = \sqrt{h_t} E(\eta_t / \psi_{t-1}) = 0 \,,$$
$$E(e_t^{\,2} / \psi_{t-1}) = h_t E(\eta_t^{\,2} / \psi_{t-1}) = h_t ;$$
$$E(e_t^{\,3} / \psi_{t-1}) = h_t^{\frac{3}{2}} E(\eta_t^{\,3} / \psi_{t-1}) = 0;$$
$$E(e_t^{\,4} / \psi_{t-1}) = h_t^{\,2} E(\eta_t^{\,4} / \psi_{t-1}) = 3h_t^{\,2}$$

### Moments of the student distribution.

*The* exp*ectations or moments higher powers of* $e_t$, *are given by*

$$E(e_t) = 0$$
$$E(e_t^{\,2}) = \frac{n}{n-2};$$
$$E(e_t^{\,3}) = 0;$$
$$E(e_t^{\,4}) = \frac{n^2 \Gamma(\frac{5}{2}) \Gamma \frac{(n-4)}{2}}{\Gamma(\frac{n}{2}) \Gamma(\frac{1)}{2}}$$

**Appendix III: Optimal Linear Estimation Method**

According to Nassiuma (1994), suppose we have one value $x_m$ missing out of a set of an arbitrarily large number of n possible observations generated from a time series process $\{x_t\}$. Let the subspace $S_m^*$ be the allowable space of estimators of $x_m$ based on the observed values $\{x_{t,} x_{t-1}, x_{t-2},..., x_1\}$ i.e., $S_m^*$ =sp $\{x_t : 1 \le n, t \ne m\}$ where n, the sample size, is assumed large. The projection of $x_m$ onto $S_m^*$ (denoted $P_{S_m^*}^{x_m}$) such that the dispersion error of the estimate (written disp $(x_m - P_{S_m^*}^{x_m})$ is a minimum would simply be a minimum dispersion linear interpolator. Direct computation of the projection $x_m$ onto $S_m^*$ is complicated since the subspaces $S_1$ =sp $\{x_{m-1}, x_{m-2},...\}$and $S_m^*$ are not independent of each other. We thus consider evaluating the projection onto two disjoint subspaces of $S_m^*$. To achieve this, we express $S_m^*$ as a direct sum of the subspaces $S_1$ and another subspace, say $x_*$, such that $S_m^* = S_1 \oplus S_*$. A possible subspace is $S_* = sp\{x_i - \hat{x}_i : i \ge m+1\}$, where $\hat{x}_i$ is based on the values $\{x_{m-1}, x_{m-2},...\}$. The existence of the subspaces $S_1$ and $S_*$ is shown in the following lemma (Nassiuma, 1994)

Lemma

Suppose $\{x_t\}$ is a nondeterministic stationary process defined on the probability space $(\Omega, B, P)$ . Then the subspaces $S_1$ and $S_*$ defined in the norm of the $L^2$ are such that $S_m^* = S_1 \oplus S_*$.

**Proof:**

Suppose $x_* \in S_m^*$, then $x_*$ can be represented as

$$x_* = Z + \sum_{i=m+1}^{n} a_i x_i = (Z + \sum_{i=m+1}^{n} a_i \hat{x}_i) + \sum_{i=m+1}^{n} a_i (x_i - \hat{x}_i)$$

where $Z \in S_1$. Clearly the two components on the right hand side of the equality are disjoint and independent and hence the result. The optimal linear estimator of $x_m$ can be evaluated as the projection onto the subspaces $S_1$ and $S_*$ such that disp $(x_m - P_{S_m^*}^{x_m})$ is minimized. i.e.,

$$x_m^* = P_{S_m^*}^{x_m} = P_{S_1}^{x_m} + P_{S_*}^{x_m} = \hat{x}_m + P_{S_*}^{x_m}.$$

But $P_{S_*}^{x_m} = \left\{ \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k : \quad disp(x_m - P_{S_*}^{x_m} \right\}$ where the coefficients' are estimated such that the dispersion error is minimized. The resulting error of the estimate is evaluated as

$$x_m - x_m^* = (x_m - \hat{x}_m) - \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

Now squaring both sides and taking expectations, we obtain the dispersion error as

$$dispx_m = E(x_m - x_m^*)^2 = E\left\{ (x_m - \hat{x}_m) - \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k) \right\}^2 \qquad (1)$$

By minimizing the dispersion with respect to the coefficients (differentiating with respect to $a_k$ and solving for $a_k$) we should obtain the coefficients $a_k$, $for \;\; k \geq m+1$, which are used in estimating the missing value (Nassiuma,1994). The missing value $x_m$ is estimated as

$$x_m^* = \hat{x}_m + \sum_{k=m+1}^{n} a_k (x_k - \hat{x}_k)$$

## Appendix IV: Research Methodology

### 3.1 Methodology

The methodology used in this study s described below. It includes the derivation approaches, data generation method, choice of performance criteria.

### 3.2 Optimal linear interpolation method

In this study, the estimators of the missing values for bilinear time series models were derived using optimal linear interpolation method by minimizing the dispersion error. The estimates were derived for pure bilinear time series and general bilinear time series having different probability distributions.

#### 3.2.1 Data generalization

Data was obtained through simulation using computer codes written in R software. These codes are presented in the appendix.

#### 3.2.2 Missing data positions and softwares

Three data points 48, 293 and 496 selected at random and data at these positions removed to create a 'missing value(s)' at these points to be estimated. Data analysis was done using statistical and computer software which included Microsoft Excel, Time Series Modeling TSM and R and Matlab . R was used to generate the data, matlab was used in determining estimates based on artificial neural networks while Microsoft Excel was used in analysis of data to calculate the MAD and MSE as well as in obtaining estimates based on exponential smoothing.

### 3.2.3 Choosing a quality fit criterion

An important methodological issue that required careful attention is the selection of a measure of "goodness of fit" between the models and the data (time series), and of a criteria to judge when this measure is "good enough" for the stated purpose. The mean square error (MSE) and mean absolute deviation (MAD) were used as performance measures.

## 3.3 Methodology for ANN

Recent studies indicate that consideration of statistical principles in the ANN model building process may improve model performance ( Cheng and Titterington, 1994; Ripley, 1994; Sarle, 1994). Consequently, a systematic approach in the development of ANN models was adopted. The steps that were followed included: data pre-processing, the determination of adequate model inputs and suitable network architecture, parameter estimation (optimization) and model validation (Maier and Dandy, 1999b). In addition, careful selection of a number of internal model parameters required was undertaken.

### 3.3.1 Structure of the artificial neural network

For this analysis a basic multilayer perceptron (MLP) with a single hidden layer will be used, which is the most commonly employed form of ANN (see Zhang, 1998).To date, there is no simple clear-cut method for determination of input parameters and the procedure adopted was to test numerous networks with varying numbers of input and hidden units p and q, respectively.

### 3.3.2 Data pre-processing

The available data was divided into their respective subsets (e.g. training, testing and validation) before any data pre-processing is carried out (Burden et al., 1997).

This was in the ratio of 70% training, 15% validation and 15% testing. In order to ensure that all variables receive equal attention during the training process, data was standardized. The standardization was in the range 0.1–0.9. This achieved using the formula

$$z_t = 0.1 + 0.8\frac{(x_t - x_{\min})}{(x_{\max} - x_{\min})}$$

where $z_t$ - standardized form, $x_t$ -original data, $x_{\max}$ - maximum value in the sample, $x_{\min}$ - minimum value in the sample. For computation of performance measures, the estimates obtained from the artificial neural work were converted back to the original form usingthe formula

$$x_t = 1.25(z_t - 0.1)(x_{\max} - x_{\min}) - x_{\min}$$

### 3.3.3 Training

The data presented to the neural networks, $z_t$ , was scaled between [0.1, 0.9]. The numbers of hidden units were re-specified for every time series. Gradient descent back-propagation was used for the training. The learning rate was set to 0.5 with a cooling factor per epoch of 0.01. Momentum was set to 0.4 and the networks are trained for 1000 epochs or until an early stopping criterion was satisfied. For the early stopping criterion the mean squared error was evaluated in every epoch. Once a network structure (p,q) was specified, the network was ready for training, a process of parameter estimation such that mean squared error of the test data is minimized.

The momentum term may also be helpful to prevent the learning process from being trapped into poor local minima, and is usually chosen in the interval [0, 1].

Finally, the estimated model was evaluated using a separate hold-out sample that is not exposed to the training process.

In order to obtain the optimum network architecture, based on the concepts of artificial neural networks design and using pruning algorithms in MATLAB 7 package software, different network architectures were evaluated and used to compare the ANNs performance. The best-fitted network was selected, and used to estimate the missing values.

The test and train procedure involves training the network on most of the input data (around 70%) and testing on the remaining data. The network performance on the test set is a good indicator of its ability to generalize and handle data it has not been trained on. If the performance on the test was poor, the network configuration or learning parameters was changed. The network was then retrained until its performance was satisfactory.

### 3.3.4 Data and performance measures

The data series were simulated from different elementary and simple bilinear models which have normal, student and GARCH distributions using R-statistical software. A program codes in R were developed to assist in the simulation.

The seed in the R program code was changed to obtain a new sample. For each program code, a set of 100 samples will be generated and analyzed. Each sample was of size 500 and missing artificial points were created at data point 48, 293 and 496 (these points were selected at random).The models selected included

| **Normal** | **Student** | **GARCH** |
|---|---|---|
| BL_(0,0,1,1) | BL_(0,0,1,1) | BL_(0,0,1,1) |
| BL_(1,0,1,1) | Bl _ (1,0,1,2) | BL_(1,0,1,1) |
| BL_(1,0,2,1) | BL_(0,0,2,1) | BL_(0,0,2,1) |

### 3.4 Performance measures

The MAD (Mean Absolute Deviation) and MSE (Mean Squared Error) were used. These were obtained from equation (3) and equation (4) respectively.

### 3.5 Methodology for exponential smoothing

A simple exponential smoothing was used to estimate the missing values. For each sample data, the constant alpha was selected from a range of values between 0.1 and 0.9 in steps of 0.1. Based on the data values before the missing point, recursive estimates were obtained. The alpha that gave the least MAD was selected and used in forecasting the missing value. This was done with the aid of excel software.

### 3.5 Estimation of missing values using optimal estimation functions.

Let $x_1, x_2, ..., x_n$ be an observed time series with $x_m$ (1<m<n) missing. Then when considering $x_m$ as a parameter we can obtain its optimal estimate as in (Thavaneswaran and

Abraham, 1988). It can be shown that the optimal estimate of $x_m$ is obtained by solving for $x_m$ in the equation $\sum_{t=1}^{n} a_{t-1}^{0} h_t = 0$, where $h_t$ is a sequence of innovations of the form, such that,

$$\mathrm{E}\ (h_t / F_{t-1}^{\xi}) = 0$$

and

$$a_{t-1}^{0} = \frac{E\left[\frac{\partial}{\partial x_m} h_t / F_{t-1}^{x}\right]}{E(h_t^{2} / F_{t-1}^{x})}.$$

**3.61 Exponential smoothing**

There are two types of exponential smoothing that can be used namely: Simple Exponential Smoothing (Exponentially weighted moving average) and Brown's Linear (i.e., double) Exponential Smoothing

**3.6.11 Simple exponential smoothing (exponentially weighted moving average)**

Let α denote a "smoothing constant" (a number between 0 and 1). One way to write the model is to define a series L that represents the current level (i.e., local mean value) of the series as estimated from data up to the present. The value of L at time t is computed recursively from its own previous value like this:

$$L_t = \alpha Y_t + (1-\alpha) L_{t-1}$$

Thus, the current smoothed value is an interpolation between the previous smoothed value and the current observation, where α controls the closeness of the interpolated value to the most recent observation. The forecast for the next period is simply the current smoothed value:

$$\hat{Y}_{t+1} = L_t$$

Equivalently, we can express the next forecast directly in terms of previous forecasts and previous observations, in any of the following equivalent versions. In the first version, the forecast is an interpolation between previous forecast and previous observation.

$$\hat{Y}_{t+1} = \alpha\hat{Y}_t + (1-\alpha)\hat{Y}_t$$

In the second version, the next forecast is obtained by adjusting the previous forecast in the direction of the previous error by a fractional amount α:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha e_t$$

Where,

$$e_t = Y_t - \hat{Y}_t$$

is the error made at time t. In the third version, the forecast is an exponentially weighted (i.e. discounted) moving average with discount factor 1-α:

$$\hat{Y}_{t+1} = \alpha[Y_t + (1-\alpha)Y_{t-1} + (1-\alpha)^2 Y_{t-2} + (1-\alpha)^3 Y_{t-3} + ....]$$

The interpolation version of the forecasting formula is the simplest to use if you are implementing the model on a spreadsheet: it fits in a single cell and contains cell references pointing to the previous forecast, the previous observation, and the cell where the value of α is stored.

Another important advantage of the SES model over the SMA model is that the SES model uses a smoothing parameter which is continuously variable, so it can easily optimized by using a "solver" algorithm to minimize the mean squared error.

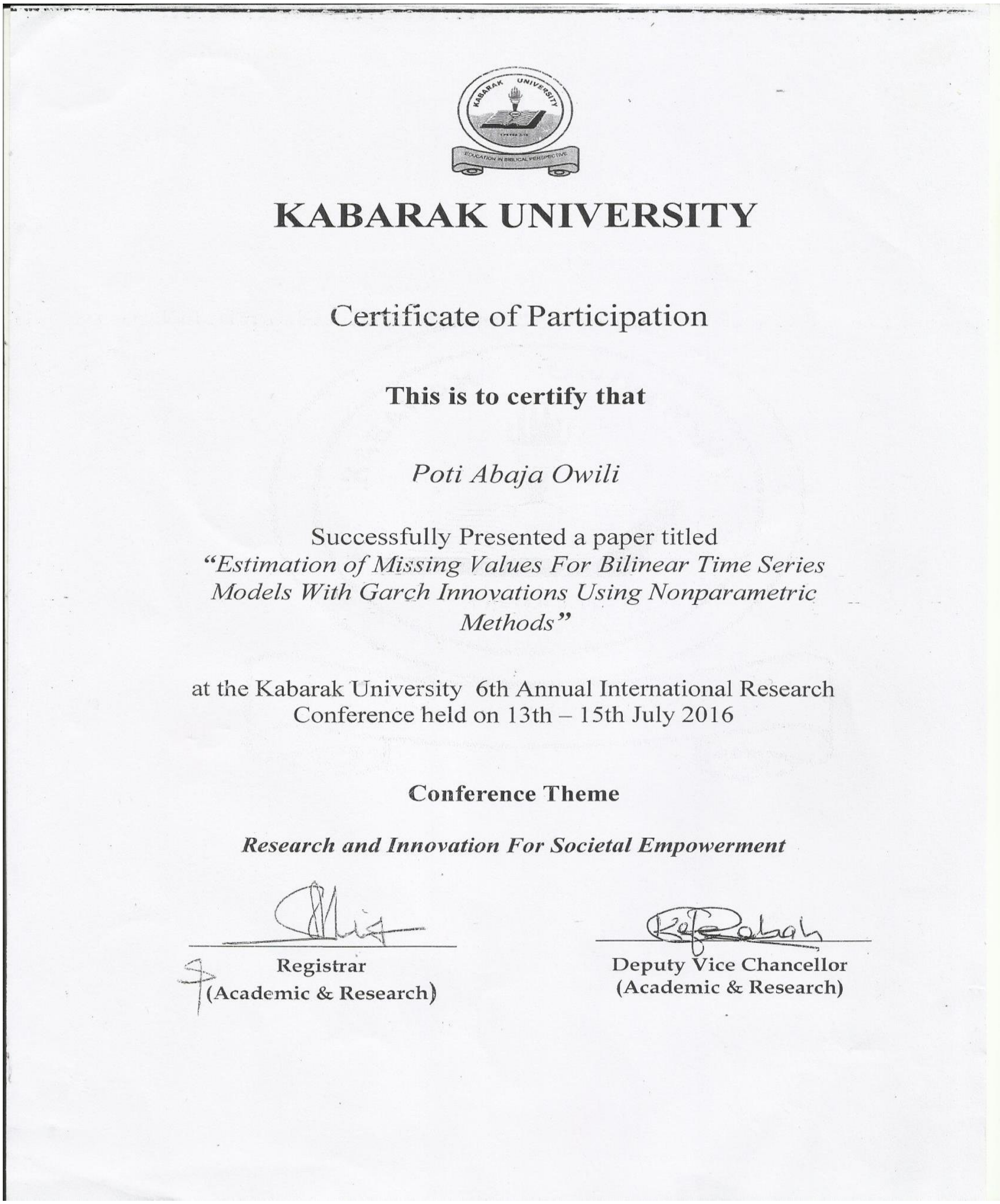
### 3.7 Brown's Linear exponential smoothing

The SMA models and SES models assume that there is no trend of any kind in the data (which is usually good or at least not-too-bad for 1-step-ahead forecasts when the data is relatively noisy), and they can be modified to incorporate a constant linear trend as shown above. What about short-term trends? If a series displays a varying rate of growth or a cyclical pattern that stands out clearly against the noise and if there is a need to forecast more than 1 period ahead, then estimation of a local trend might also be an issue. The simple exponential smoothing model can be generalized to obtain a linear exponential smoothing (LES) model that computes local estimates of both level and trend. The simplest time-varying trend model is Brown's linear exponential smoothing model, which uses two different smoothed series that are centered at different points in time. The forecasting formula is based on an extrapolation of a line through the two centers.

### 3.8 Exponential smoothing

Exponential Smoothing methods are the most common methods of forecasting. Their popularity can be attributed to several practical considerations. First, they are very simple in concept and easy to understand. Second, they require little computational effort and small data storage space. Third, they can achieve flexible adaptivity by varying smoothing parameters to account for changes in the behaviors of the time series being forecasted.

**Appendix V: Paper Presentation in International Conference**



**KABARAK UNIVERSITY**

Certificate of Participation

**This is to certify that**

*Poti Abaja Owili*

Successfully Presented a paper titled
*"Estimation of Missing Values For Bilinear Time Series Models With Garch Innovations Using Nonparametric Methods"*

at the Kabarak University 6th Annual International Research Conference held on 13th – 15th July 2016

**Conference Theme**

***Research and Innovation For Societal Empowerment***

Registrar
(Academic & Research)

Deputy Vice Chancellor
(Academic & Research)

**Appendix VI: Paper Presentation in International Conference**

EDUCATION IN BIBLICAL PERSPECTIVE

# KABARAK UNIVERSITY

## Certificate of Participation

**This is to certify that**

*Poti Abaja Owili*

Successfully Presented a paper titled
*"Efficiency of Nonparametric Estimators For Missing Observations of Bilinear Time Series With Gaussian Innovations"*

at the Kabarak University 6th Annual International Research Conference held on 13th – 15th July 2016

**Conference Theme**

***Research and Innovation For Societal Empowerment***

Registrar
(Academic & Research)

Deputy Vice Chancellor
(Academic & Research)

## Appendix VII : Publications

*American Journal of Applied Mathematics and Statistics, 2015, Vol. 3, No. 5, 199-205*
Available online at http://pubs.sciepub.com/ajams/3/5/4

DOI:10.12691/ajams-3-5-4

Science & Education Publishing

# Imputation of Missing Values for Pure Bilinear Time Series Models with Normally Distributed Innovations

**Poti Owili Abaja[1,*], Dankit Nassiuma[2], Luke Orawo[3]**

[1]Mathematics and Computer Science Department, Laikipia University, Nyahururu, Kenya
[2]Mathematics Department, Africa International University, Nairobi
[3]Mathematics Department, Egerton University, Private Bag, Egerton-Njoro, Nakuru, Kenya
*Corresponding author: abajapoti @gmail.com

**Abstract** In this study, estimates of missing values for bilinear time series models with normally distributed innovations were derived by minimizing the h-steps-ahead dispersion error. For comparison purposes, missing value estimates based on artificial neural network (ANN) and exponential smoothing (EXP) techniques were also obtained. Simulated data was used in the study. 100 samples of size 500 each were generated for different pure bilinear time series models using the R-statistical software. In each sample, artificial missing observations were created at data positions 48, 293 and 496 and estimated using these methods. The performance criteria used to ascertain the efficiency of these estimates were the mean absolute deviation (MAD) and mean squared error (MSE). The study found that optimal linear estimates were the most efficient estimates for estimating missing values. Further, the optimal linear estimates were equivalent to one step-ahead forecast of the missing value. The study recommends OLE estimates for estimating missing values for pure bilinear time series data with normally distributed innovations.

***Keywords:*** *optimal linear interpolation, simulation, MAD, innovations, ANN, exponential smoothing*

**Cite This Article:** Poti Owili Abaja, Dankit Nassiuma, and Luke Orawo, "Imputation of Missing Values for Pure Bilinear Time Series Models with Normally Distributed Innovations." *American Journal of Applied Mathematics and Statistics*, vol. 3, no. 5 (2015): 199-205. doi: 10.12691/ajams-3-5-4.

## 1. Introduction

A time series is data recorded sequentially over a specified time period. There are cases where some observations that were supposed to be collected are not obtained and this result in missing values. Being unable to account for missing observation may result in a severe mis-representation of the phenomenon under study. Further, it can cause havoc in the estimation and forecasting of linear and nonlinear time series as in [3]. This problem can be solved through missing value imputation.

Imputation of missing values has been done for several linear time series models. For non-linear time series models, imputation has been done for ARMA models with stable errors as in [24]. For other nonlinear models, such as bilinear time series models, there is no evidence to show that imputation of missing values has been explicitly done. Therefore this study derived estimates of missing values for the bilinear time series models with normally distributed innovations. The missing values were derived using optimal linear interpolation techniques based on minimizing the h-steps-ahead dispersion error. Other techniques for estimating missing values that were used included the non-parametric methods of artificial neural network as in [4] and [31] as well as exponential smoothing.

Interest in this study was also on the quality of the imputed values at the level of the individual, an issue that has received relatively little attention as in [5]. The basic idea of an imputation approach, in general, is to substitute a plausible value for a missing observation and to carry out the desired analysis on the completed data as in [22]. Here, imputation can be considered to be an estimation or interpolation technique.

The imputation of the missing value technique developed may be adopted by data analysts to improve on time series modeling.

## 2. Literature Review

Most of the real-life time series encountered in practice are neither Gaussian nor linear in nature and are adequately described by nonlinear models. One of the most important nonlinear models used in practice is the bilinear time series models. The nonlinearity of bilinear models can be approached in two ways. The first approach is to create a model that consist of a blend of non-Gaussian and nonlinearity which has been widely discussed as in [31] where he considers the existence of bilinear models with infinite variance innovations. The other approach is to introduce nonlinearity in the model but assume that the distribution of the innovation sequence is Gaussian as in [36]. Properties of these models have been extensively studied in the literature.

### 2.1. Bilinear Models

A discrete time series process $X_t$ is said to be a bilinear time series model of order BL (p,q , P, Q) if it satisfies the difference equation

## Appendix VIII: Pubblictions

International Journal of Statistics and Applications 2015, 5(6): 293-301
DOI: 10.5923/j.statistics.20150506.05

# Estimation of Missing Values for Pure Bilinear Time Series Models with Student-t Innovations

**Poti Abaja Owili[1,*], Luke Orawo[2], Dankit Nassiuma[3]**

[1]Mathematics and Computer Science Department, Laikipia University, Nyahururu, Kenya
[2]Mathematics Department, Egerton University, Nakuru, Private Bag, Egerton-Njoro, Kenya
[3]Mathematics Department, Africa International University, Nairobi, Kenya

**Abstract** In this study optimal linear estimates of missing values for pure bilinear time series models whose innovations have a student-t distribution are derived by minimizing the h-steps-ahead dispersion error. Data used in the study was simulated using the R-software where 100 samples of size 500 were generated for simple bilinear models. In each sample, three data positions 48, 293 and 496 were selected at random and artificial missing values created at these points. For comparison purposes, artificial neural network (ANN) and exponential smoothing (EXP) estimates were also computed. The performance criteria used to ascertain the efficiency of these estimates were the mean absolute deviation (MAD) and mean squared error (MSE). The study found that optimal linear estimates were the most efficient for estimating missing values of the pure bilinear time series followed by exponential smoothing estimates. Further, these estimates were equivalent to one-step-ahead forecast. The study recommends the use of optimal linear estimate for estimating missing values in pure bilinear time series data whose innovations have student-t distribution.

**Keywords** ANN, Exponential smoothing, MAD, Performance criterion, Simulation

## 1. Introduction

Data analysts are frequently faced with situations where one or several time series observations are missing at certain points within the data set collected for modeling. This creates missing values at such points. Being unable to account for missing values may result in a severe misrepresentation of the phenomenon under study or can cause havoc in the estimation and forecasting of linear and nonlinear time series as in [1]. The missing values can be reconstructed using imputation techniques. The basic idea of an imputation approach, in general, is to substitute a plausible value for a missing observation and to carry out the desired analysis on the completed data as in [2]. There are several methods that can be used for imputing missing values in numerical data. These methods include mean substitution, linear regression, neural networks and nearest neighbor approach and optimal linear interpolation.

In this study we were interested in estimating missing values for a class of bilinear nonlinear time series models called the pure bilinear time series models whose innovations have a student-t distribution using linear interpolation approach. The student-t distribution can be used to model the innovation of financial data which is known to be highly skewed. There is no evidence to show that optimal linear interpolation approach has been used to estimate missing values for bilinear time series and specifically, pure bilinear time series.

### 1.1. Bilinear Models

A discrete time series process $x_t$ is said to be a bilinear time series model of order BL (p, q, P, Q) if it satisfies the difference equation

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + \sum_{i=1}^{P}\sum_{j=1}^{Q} b_{ij} x_{t-i} e_{t-j} + e_t \quad (1)$$

where $\varphi_i$ (i=1,2,..p), $\theta_j$ (j=1,2,..q), $b_{ij}$ (i=1,2,..,P ; j=1,2,..,Q) are constants; the innovation sequence $\{e_t\}$ are i.i.d random process which has a student-t distribution and $\theta_o$ =1. For pure bilinear time series, only the bilinear coefficient is not equal to zero. Thus the pure bilinear time series model, BL(0,0,P,Q) is given by

$$x_t = \sum_{i=1}^{P}\sum_{j=1}^{Q} b_{ij} x_{t-i} e_{t-j} + e_t \quad (2)$$

where $b_{ij}$, $1 \le i \le P$ *and* $1 \le j \le Q$, are the coefficients of the model, and $\{e_t\}$ are i.i.d student-t process with zero mean and common finite variance. Some important

* Corresponding author:
abajapoti@gmail.com (Poti Abaja Owili)
Published online at http://journal.sapub.org/statistics

## Appendix: IX Publications

American Journal of Mathematics and Statistics 2015, 5(5): 316-324
DOI: 10.5923/j.ajms.20150505.13

# Efficiency of Imputation Techniques for Missing Values of Pure Bilinear Models with GARCH Innovations

**Poti Abaja Owili[1,*], Dankit Nassiuma[2], Luke Orawo[3]**

[1]Mathematics and Computer Science Department, Laikipia University, Nyahururu, Kenya
[2]Mathematics Department, Africa International University, Nairobi, Kenya
[3]Mathematics Department, Egerton University, Private bag, Egerton-Njoro, Kenya

**Abstract** A major problem facing data analyst is the development and determination of the most efficient imputation technique for imputing missing observations in data used for modeling. Several imputation techniques exist. However, most of them do not take into consideration the probability distribution of the innovation sequence of the time series model. Therefore this study derived optimal linear estimates of missing values for bilinear time series models with GARCH innovations based on minimizing the h-steps-ahead dispersion error. The efficiency of the estimates derived was compared with those obtained using nonparametric techniques of artificial neural network (ANN) and exponential smoothing (EXP) using simulated data. A hundred different samples of size 500 each were generated for two different pure bilinear models with GARCH innovations namely: BL(0,0,1,1) and BL(0,0,2,1). In each sample, artificial missing observations were created at data positions 48, 293 and 496 and estimated using these methods. The performance criteria used to measure the efficiency of these estimates were the mean absolute deviation (MAD) and mean squared error (MSE). The study found mixed results; for the BL(0,0,1,1) model, the optimal linear estimates were the most efficient while for the BL(0,0,2,1), ANN were the most efficient. The study recommends that both ANN and optimal linear estimates be used in estimating missing values for bilinear time series data with GARCH innovations.

**Keywords** Artificial Neural Networks, Exponential Smoothing, Optimal Linear Interpolation

## 1. Introduction

A time series is data recorded sequentially over a specified time period. Data analysts are frequently faced with cases of missing observation at certain points within the data set collected for modeling. Missing values may occur for various reasons which may include poor record keeping, lost records etc. In addition some data, suspected to be outliers, may be deleted because they were erroneously collected. A major problem with missing data is that it can cause havoc in the estimation and forecasting of linear and nonlinear time series as in [2]. Therefore, if data has missing values, it is necessary that the missing value be imputed first before the data is analyzed. Missing value imputation techniques have been developed for several linear and nonlinear time series models. Most of these methods do not consider the distribution of the innovation sequence of the time series data, especially the non-Gaussian distribution. Further, most of these methods only deal with linear ARMA models. In this study, we are interested in a class of nonlinear bilinear time series models whose innovations have GARCH errors. For bilinear time series models, estimation of missing values is still at its infancy stage. This is especially so for bilinear time series models whose innovations are non-Gaussian.

A discrete time series process $X_t$ is said to be a bilinear time series model, denoted byBL (p, q, P, Q), if it satisfies the difference equation

$$X_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + \sum_{i=1}^{P} \sum_{j=1}^{Q} b_{ij} x_{t-i} e_{t-j} + e_t$$

and $\theta, \phi$ and $b_{ij}$ are constant $\theta_o = 1$ and $e_t$ is the innovation sequence which is normally distributed.

For pure bilinear time series model, we have

$$x_t = \sum_{i=1}^{p} \sum_{j=1}^{q} b_{ij} x_{t-i} + e_t.$$

It was proposed by Granger and Andersen (1978a) and has been widely applied in many areas such as control theory, economics and finance. For bilinear time series model with GARCH (p, q) innovations, the innovation $e_t$ is expressed as

$$e_t = \eta_t \sqrt{h_t},$$

* Corresponding author:
abajapoti@gmail.com (Poti Abaja Owili)
Published online at http://journal.sapub.org/ajms

## Appendix X: Publication

**Science Journal of Applied Mathematics and Statistics**
2015; 3(6): 234-242
Published online October 29, 2015 (http://www.sciencepublishinggroup.com/j/sjams)
doi: 10.11648/j.sjams.20150306.12
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

SciencePG
Science Publishing Group

# Imputation of Missing Values for BL (P,0,P,P) Models with Normally Distributed Innovations

**Poti Abaja Owili**

Mathematics and Computer Science Department, Laikipia University, Nyahururu, Kenya

**Email address:**
abajapoti@gmail.com

**To cite this article:**
Poti Abaja Owili. Imputation of Missing Values for BL (P,0,P,P) Models with Normally Distributed Innovations. *Science Journal of Applied Mathematics and Statistics.* Vol. 3, No. 6, 2015, pp. 234-242. doi: 10.11648/j.sjams.20150306.12

**Abstract:** This study derived estimates of missing values for bilinear time series models BL (p, 0, p, p) with normally distributed innovations by minimizing the h-steps-ahead dispersion error. For comparison purposes, missing value estimates based on artificial neural network (ANN) and exponential smoothing (EXP) techniques were also obtained. Simulated data was used in the study. 100 samples of size 500 each were generated for bilinear time series models BL (1, 0, 1, 1) using the R-statistical software. In each sample, artificial missing observations were created at data positions 48, 293 and 496 and estimated using these methods. The performance criteria used to ascertain the efficiency of these estimates were the mean absolute deviation (MAD) and mean squared error (MSE). The study found that optimal linear estimates were the most efficient estimates for estimating missing values for BL (p, 0, p, p). The study recommends OLE estimates for estimating missing values for bilinear time series data with normally distributed innovations.

**Keywords:** Optimal Linear Interpolation, Simulation, MSE, Innovations, ANN, Exponential Smoothing

## 1. Introduction

A time series is data recorded sequentially over a specified time period. There are cases where some observations that were supposed to be collected are not obtained and this result in missing values. Being unable to account for missing observation may result in a severe misrepresentation of the phenomenon under study. Further, it can cause havoc in the estimation and forecasting of linear and nonlinear time series as in [3]. This problem can be overcome through missing value imputation.

Imputation of missing values has been done for several linear time series models. For non-linear time series models, imputation has been done for ARMA models with stable errors as in [24]. For other nonlinear models, such as bilinear time series models, there is no evidence to show that imputation of missing values has been explicitly done. Therefore this study derived estimates of missing values for the bilinear time series models with normally distributed innovations. The missing values were derived using optimal linear interpolation techniques based on minimizing the h-steps-ahead dispersion error. Other techniques for estimating missing values that were used included the non-parametric methods of artificial neural network as in [4], [31] and exponential smoothing.

Interest in this study was also on the quality of the imputed values at the level of the individual, an issue that has received relatively little attention as in [5]. The basic idea of an imputation approach, in general, is to substitute a plausible value for a missing observation and to carry out the desired analysis on the completed data as in [22]. Here, imputation can be considered to be an estimation or interpolation technique.

The imputation of the missing value technique developed may be adopted by data analysts to improve on time series modeling.

## 2. Literature Review

Most of the real-life time series encountered in practice are neither Gaussian nor linear in nature and are adequately described by nonlinear models. One of the most important nonlinear models used in practice is the bilinear time series models. The nonlinearity of bilinear models can be approached in two ways. The first approach is to create a model that consist of a blend of non-Gaussian and nonlinearity which has been widely discussed as in [31] where he considers the existence of bilinear models with

## Appendix XI Publication

International Journal of Statistics and Applications 2016, 6(3): 113-122
DOI: 10.5923/j.statistics.20160603.04

# Estimation of Missing Values for BL (p, 0, p, p) Time Series Models with Student-t Innovations

**Poti Abaja Owili**

Mathematics and Computer Science Department, Laikipia University, Nyahururu, Kenya

**Abstract** In this study optimal linear estimators of missing values for bilinear time series models BL (p, 0, p, p) whose innovations have a student-t distribution are derived by minimizing the h-steps-ahead dispersion error. Data used in the study was simulated using the R Statistical Software where 100 samples of size 500 each were generated for the bilinear model BL (1, 0, 1, 1). The time series data generated was numbered from 1 to 500. In each sample, three data positions 48, 293 and 496 were selected at random and the value at these points removed to create artificial missing values. For comparison purposes, two commonly used non-parametric techniques of artificial neural network (ANN) and exponential smoothing (EXP) estimates were also computed. The performance criteria used to ascertain the efficiency of these estimates were the mean squared error (MSE) and Mean Absolute Deviation (MAD). The study found that ANN estimates were the most efficient for estimating missing values of the bilinear time series with student-t innovations. The study recommends the use of ANN for estimating missing values in bilinear time series model with student errors.

**Keywords** ANN, Exponential smoothing, MSE, Performance criterion, Simulation

## 1. Introduction

Data analysts are frequently faced with the missing value problem. Missing values may occur for various reasons which may include poor record keeping, lost records, technical error, collecting data at irregular times, etc., ([1], [2]). In addition, a peculiar case can arise when one may be interested in determining the likely value of a variable of interest at a time that may not coincide with a particular measurement or observation [3]. These can result in one or several observations missing.

These missing values must be accounted for since missing values have negative effects on the modeling of the data [4]. There are many ways of handling missing values. The common approach is to use imputation techniques. This involves using a substitute value to replace the missing observation as in [5]. According to [6], imputation broadly comprises several techniques that have been developed to compute missing values. These techniques may employ strategies such as mean substitution and artificial neural networks approach. It may also involve the use of appropriate statistical prediction or forecasting models such as regression, time series models, and Markov chain and Monte Carlo methods.

Estimation of missing values for bilinear time series has been done for a particular order of the bilinear time series BL (1, 0, 2, 0) by [4]. They used estimating functions criterion to derive the estimates of missing values. Other studies have also been done to estimate missing values for pure bilinear time series when the innovation sequence has the GARCH distribution [7]. Still the same authors have estimated missing values for pure bilinear time series when the innovation sequence has the normal distribution [8]. [9] also estimated missing values for bilinear time series for the pure bilinear case when the innovation sequence has the student-t distribution. They found that the estimates of the missing values were equivalent to a one–step-ahead forecast. Further, [10] used the linear interpolation criterion to estimate missing values for the BL (p, 0, p, p) when the error term follows the normal distribution.

The distribution of interest in this study is the student-t distribution. This distribution is characterized by long tails and is suitable for modeling financial data which is known to be highly skewed. There is no evidence to show that optimal linear interpolation approach based on the dispersion error has been used to estimate missing values for bilinear BL (p, 0, p, p) with student-t distribution.

### 1.1. Identification of Bilinear Time Series Models

Given a time series data, the first step in the identification process of bilinear time series model is to test whether the data can be modeled either as a linear time series or belongs to the broader class of nonlinear time series models. This involves testing a null hypothesis that the data is linear. This can be done using one of the statistical tests of linearity ([11];

* Corresponding author:
abajapoti@gmail.com (Poti Abaja Owili)
Published online at http://journal.sapub.org/statistics