

---

# 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

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. This is the bilinear model of interest in the study.

### 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

$$X_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + \sum_{i=1}^P \sum_{j=1}^Q b_{ij} x_{t-i} e_{t-j} + e_t$$

where  $\theta$ ,  $\phi$  and  $b_{ij}$  are constants while  $e_t$  is a purely random process which is normally distributed and  $\theta_o = 1$ . For the bilinear time series model BL(p,o,p,p), we have

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^p \sum_{j=1}^p b_{ij} x_{t-i} + e_t, \quad e_t \sim N(0, \sigma^2).$$

Bilinear time series models may have sudden burst of large negative and positive values that vary in form and amplitude depending on the model parameters and thus it may be plausible for modeling nonlinear processes as in [25]. A bilinear model is a member of the general class of nonlinear time series models called ‘State dependent models’ formed by adding the bilinear term to the conventional ARMA model as in [30].

It is a parsimonious and powerful nonlinear time series model. Researchers have achieved forecast improvement with simple nonlinear time series models. Reference [21] 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. Reference [9] also reported a forecast improvement with bilinear models in forecasting stock prices. The statistical properties of such models have been analyzed in detail as in [10], [11], [25], and [17] etc., while an economic application is presented as in [14].

The first step in identification of bilinear time series model is to determine whether a given data is nonlinear or not. Once the data is found to be nonlinear, it is important to fit an appropriate time series model to the data. Reference [39] pointed out that the second order properties of BL (p, p, 0, 1) are similar to those of ARMA (p, 1) and hence it is necessary to study higher order cumulants to distinguish them from nonlinear models. 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 as in [24].

For some super diagonal and diagonal bilinear time series, the third order moments do not vanish and the pattern of

nonzero moments can be used to discriminate between the bilinear models and white noise and also between different bilinear models. Looking at the 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. References [10] and [37] have shown that in most cases, second order autocorrelation will be zero for these models which makes it difficult to distinguish them from complete white noise.

Reference [24] showed that with a large bilinear coefficient  $b_{ij}$ , a bilinear model can have sudden large amplitude bursts and is suitable for some kind of seismological data such as earthquakes and underground nuclear explosions. The variant of the bilinear process is time dependent. This feature enables bilinear process to be used also for financial data as in [21]. Empirical studies have been done on estimating missing values for different time series data. Reference [26] used interpolation and mean imputation techniques to replace simulated missing values from annual hourly monitoring air pollution data.

Reference [29] developed alternative techniques suitable for a limited set of stable case with index  $\alpha \in (1, 2]$ . This was later extended to the ARMA stable process with index  $\alpha \in (0, 2]$  as in [24]. He developed an algorithm applicable to general linear and nonlinear processes by using the state space formulation and applied it in the estimation of missing values.

### 2.2. Missing Value Imputation for Nonlinear Time Series Models

Reference [35] derived a recursive estimation procedure based on optimal estimating function and applied it to estimate model parameters to the case where there are missing observations as well as handle time-varying parameters for a given nonlinear multi-parameter model. More specifically, to estimate missing observations, [3] formulated a nonlinear time series model which encompasses several standard nonlinear models of time series as special cases. It also offers two methods for estimating missing observations based on prediction algorithm and fixed point smoothing algorithm as well as estimating functions equations. Recursive estimation of missing observations in an autoregressive conditional heteroscedasticity (ARCH) model and the estimation of missing observations in a linear time series model are shown as special cases. However, little was done on bilinear time series models.

Reference [28] 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.

The work on estimation of missing values has also been extended to vector time series. A classic example is the studies done as in [20] who worked on estimation of missing

values in possibly partially non stationary vector time series. He extended the method as in [19] 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. No pattern of missing data is assumed. Future and past values are special cases of missing data that can be estimated in the same way. The method does not use the Kalman filter iterations and hence avoids initialization problems. The estimation of missing values is provided by the normal equations of an appropriate regression problem.

**2.3. Nonparametric Methods for Estimating Missing Values**

Nonparametric methods have also been proposed for missing data. Reference [26] 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 missingness, the kernel conditional mean estimator to impute the missing values is proposed as in [5]. Reference [13] studied the estimation of average treatment effects using non-parametrically estimated propensity scores. Time series smoothers estimate the level of a time series at time t as its conditional expectation given present, past and future observations, with the smoothed value depending on the estimated time series model as in [16].

Reference [25] derived a recursive form of the exponentially smoothed estimated for a nonlinear model with irregularly observed data and discussed its asymptotic properties. They made numerical comparison between the resulting estimates and other smoothed estimates. Reference [1] have used neural networks and genetic algorithms to approximate missing data in a database. A genetic algorithm is used to estimate the missing value by optimizing an objective function. Many approaches have been developed to recover missing values, such as k-nearest neighbor as in [38], Bayesian PCA (BPCA) as in [27], least square imputation as in [12], local least squares imputation (LLSimpute) as in [15] and least absolute deviation imputation (LADimpute) as in [5].

It can be seen from the literature that there are a variety of methods used for estimating missing values for different time series models. What is however lacking in the literature is an explicit method for estimating missing values for bilinear time series models. The study therefore purposed to address this gap. It estimated missing values for bilinear time series which have different probability distributions.

**2.4. Estimation of Missing Values Using Linear Interpolation Method**

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  $\{x_t, x_{t-1}, x_{t-2}, \dots, x_1\}$  i.e.,  $S_m^* = sp\{x_t : 1 \leq n, t \neq m\}$  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  $S_1 = sp\{x_{m-1}, x_{m-2}, \dots\}$  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  $S_*$ , such that  $S_m^* = S_1 \oplus S_*$ . A possible subspace is  $S_* = sp\{x_i - \hat{x}_i : i \geq m+1\}$ , where  $\hat{x}_i$  is based on the values  $\{x_{m-1}, x_{m-2}, \dots\}$ . The existence of the subspaces  $S_1$  and  $S_*$  is shown in the following 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 best 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.,

$$x_m^* = P_{S_m^*}^{x_m} = P_{S_1}^{x_m} + P_{S_*}^{x_m} = \hat{x}_m + P_{S_*}^{x_m}$$

$$\text{But } P_{S_*}^{x_m} = \left\{ \sum_{k=m+1}^n a_k (x_k - \hat{x}_k) : disp(x_m - P_{S_*}^{x_m}) \right\}$$

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

$$disp x_m = E(x_m - x_m^*)^2$$

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

By minimizing the dispersion with respect to the coefficients the optimal linear estimate is

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

### 3. Methodology

Three methodologies were used in this study, each corresponding with the estimation method used. These methods included optimal linear interpolation, artificial neural network and exponential smoothing. However, the time series data used and performance measures applied were the same for all the methods.

For optimal linear interpolation, the estimates of the missing values for bilinear time series models with normal innovations were derived by minimizing the dispersion error. The estimates of missing values using non parametric methods of ANN and exponential smoothing were also obtained.

#### 3.1. Data Collection

Data was obtained through simulation using computer codes written in R-software. These codes are shown in the Appendix. The time series were simulated from different simple bilinear models which have normal. The seed in the R program code was changed to obtain a new sample for the general bilinear model BL (1,0,1,1). For each program code, a set of 100 samples of size 500 were generated. These were to be used in the analysis.

#### 3.2. Missing Data Points and Data Analysis

Three data points 48, 293 and 496 were selected at random from a sample of size 500. Observations at these points were removed to create a ‘missing value(s)’ at these points to be estimated. Data analysis was done using statistical and computer software which included Excel, TSM and R and Matlab7.

#### 3.3. Performance Measures

The MAD (Mean Absolute Deviation) and MSE (Mean Squared Error) were used. These were obtained as follows

$$MAD = \sum |e_t| / n \quad (3)$$

and

$$MSE = \sum |e_t|^2 / n \quad (4)$$

## 4. Results

### 4.1. Derivation of the Optimal Linear Estimates of Missing Values

Estimates of missing values for pure bilinear time series models whose innovations have a Gaussian innovation were derived by minimizing the h-steps-ahead dispersion error. Two assumptions were made. The first one was that that the series 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.

### 4.2. Bilinear Time Series BL (1,0,1,1) with Normally Distributed Innovations

The stationary bilinear time series model BL(1,0,1,1) with normally distributed innovations order BL(1,0,1,1) is given by

$$x_t = \phi_1 x_{t-1} + b_{11} x_{t-1} e_{t-1} + e_t \quad (5)$$

where

$$e_t \sim N(0,1)$$

#### Theorem 4.1

The best 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 (5), the stationary BL (1,0,1,1) can be expresses as

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

The h-steps ahead forecast is given by

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

and the h-steps ahead forecast error is given by

$$x_{t+h} - \hat{x}_{t+h} = \sum_{i=1}^{h-1} \left( \prod_{j=1}^i (\phi_1 + b_{11} e_{t+h-j}) \right) e_{t+h-i} + e_{t+h} \quad (6)$$

Substituting (6) in (1) we have

$$\begin{aligned} \text{disp } x_m &= E(\hat{e}_m^2) \\ -2E \left\{ \hat{e}_m \bullet \left( \sum_{k=m+1}^n a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^i (\phi_1 + b_{11} e_{k-j}) \right) e_{k-i} \right) \right. \\ &\quad \left. + \sum_{k=m+1}^n a_k e_k \right\} \\ + E \left( \sum_{k=m+1}^n a_k \sum_{i=1}^{k-m} \left( \prod_{j=1}^i (\phi_1 + b_{11} e_{k-j}) \right) e_{k-i} \right)^2 \\ &\quad + \sum_{k=m+1}^n a_k e_k \end{aligned} \tag{7}$$

Simplifying each of the terms on the rhs of (7), we obtain

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

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

$$\begin{aligned} E 2 \left( \sum_{k=m+a}^n a_k \sum_{i=1}^{k-m} (\phi_1 + b_{11} e_{k-j}) \right) e_{k-j} \sum_{k=m+1}^n a_k e_k &= 0 \\ E \left( \sum_{k=m+1}^n a_k e_k \right)^2 &= \sum_{k=m+1}^n a_k^2 \sigma^2 \end{aligned}$$

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

Hence equation (8) becomes

$$\begin{aligned} \text{disp } x_m &= \sigma^2 - 2 \sum_{k=m+1}^n a_k \phi_1^{k-m} \sigma^{22} \\ &\quad + \sum_{k=m+1}^n a_k^2 (\phi_1^2 \sigma^2 + 3\sigma^4 b_{11}^2) + \sum_{k=m+1}^n a_k^2 \sigma^2 \end{aligned} \tag{8}$$

Differentiating (8) with respect to the coefficient  $a_k$ , we get

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

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

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

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

**4.3. 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

$$\begin{aligned}
 x_t &= \sum_{i=1}^p \varphi_i x_{t-i} + \sum_{i=1}^p \sum_{j=1}^p b_{ij} x_{t-i} e_{t-j} \\
 &+ e_t, e_t \sim N(0,1)
 \end{aligned} \tag{9}$$

The missing value estimate is based on the following theorem 4.2.

*Theorem 4.2*

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

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

*Proof*

Through recursive substitution of (9), the stationary bilinear time series model BL (p, 0,p,p) is expressed as

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

The h-steps ahead forecast is given by

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

or

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

and the forecast is

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

The forecast error is

$$\begin{aligned}
 x_k - \hat{x}_k &= \sum_{i=1}^p \sum_{i=1}^{k-m} \prod_{j=1}^i \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{k-j} \right\} e_{t-sj} \\
 &+ \sum_{s < r}^p \sum_r^p \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{k-j} \right\} \left\{ \phi_r + \sum_{j=1}^p b_{rj} e_{k-j} \right\} e_{k-s-r} \\
 &+ \sum_{r > s}^p \sum_s^p \left\{ \phi_r + \sum_{j=1}^p b_{rj} e_{k-j} \right\} \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{k-j} \right\} e_{k-s-r} + e_t \\
 &+ \sum_{i=0}^{\infty} \prod_{j=1}^i (\varphi_2 + b_{21} e_{k-2j+1} + b_{22} e_{k-2j}) \cdot (\theta_2 e_{k-2i-2} + \theta_1 e_{k-2i-1} + e_{k-2i}) \\
 &+ \{\theta_2 e_{k-2} + \theta_1 e_{k-1} + e_k\}
 \end{aligned} \tag{10}$$

Substituting in (10) in (1), we obtain

$$\begin{aligned}
 disp x_m &= E(x_m - \hat{x}_m)^2 - 2E(x_m - \hat{x}_m) \\
 &\bullet \sum_{k=m+1}^n a_k (x_k - \hat{x}_k) \\
 &+ E \left\{ \sum_{k=m+1}^n a_k (x_k - \hat{x}_k) \right\}^2
 \end{aligned} \tag{11}$$

The terms on the right hand side of (11) are simplified as follows

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

$$-2\hat{e}_m \bullet E \sum_{k=m+1}^n a_k \sum_{i=1}^p \sum_{j=1}^{\infty} \prod_{j=1}^i \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{t+h-j} \right\} e_{t-sj}$$

$$+ \sum_{s < r}^p \sum_{r}^p \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{t+h-j} \right\} \bullet \left\{ \phi_r + \sum_{j=1}^p b_{sj} e_{t+h-j} \right\} + e_{t+h-s-r}$$

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

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

$$= -2 \sum_{k=m+1}^n \phi^{(k-m)/s} \sigma^2$$

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

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

Third term

$$E \sum_{i=1}^p \sum_{i=1}^{k-m} \prod_{j=1}^i \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{k-j} \right\} e_{t-sj} +$$

$$E \left[ \begin{aligned} &\sum_{s < r}^p \sum_{r}^p \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{k-j} \right\} \left\{ \phi_r + \sum_{j=1}^p b_{sj} e_{k-j} \right\} e_{k-s-r} \\ &+ \sum_{r > s}^p \sum_{s}^p \left\{ \phi_r + \sum_{j=1}^p b_{rj} e_{k-j} \right\} \left\{ \phi_s + \sum_{j=1}^p b_{sj} e_{k-j} \right\} e_{k-s-r} \\ &+ e_t \end{aligned} \right]$$

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

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

$$+ \sum_{k=m+1}^n a_k^2 \left\{ \sum_{s=1}^p \phi_s^2 \sigma^2 + \sum_{j=1, s \neq j}^p b_{sj}^2 \sigma^2 + e_{k-j} + \sum_{j=1}^p 3b_{jj}^2 \sigma^4 \right\} \quad (12)$$

differentiating (12) with respect to  $a_k$  and setting to zero we obtain

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

Solving 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

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

$$= \sum_{i=1}^p \hat{\phi}_i x_{t-i} + \sum_{i=1}^p \sum_{j=1}^p \hat{b}_{ij} x_{t-i} \hat{e}_{t-j}$$

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

#### 4.4. Simulation Results

In this section, the results of the estimates obtained from the optimal linear estimate, artificial neural networks and exponential smoothing are presented. Simulation results are given in table1. These graphs are characterized by sharp outburst as clearly evident in BL (1,0,1,1). Sharp outburst is one of the characteristics of nonlinearity in bilinear models.

Table 1. Efficiency Measures obtained for Normal-BL(1,0,1,1).

MISSING POSITION	MAD			MSE		
	OLE	ANN	EXP	OLE	ANN	EXP
48.000	0.793	1.135	0.982	1.043	2.621	1.542
293.000	0.760	0.870	0.812	0.906	1.603	1.079
496.000	0.803	0.863	0.933	0.976	1.215	1.369
Total	2.356	2.869	2.726	2.925	5.439	3.990
Mean	0.785	0.956	0.909	0.975	1.813	1.330

From table 1, it is clear that the OLE estimates of missing values were the most efficient (mean MAD=0.785 for the different missing data point positions. This was followed by EXP smoothing estimates (mean MAD=0.908818). The estimates based on ANN were the least efficient.

## 5. Conclusion

In this study we have derived the estimates of missing values for bilinear time series models BL(p,0,p,p) whose innovations are normally distributed by minimizing the dispersion error. The study found the optimal linear estimate of the missing value is a function of both the data observations before and after the missing value position. Further, the study also found that optimal linear estimates were the most efficient estimates for the bilinear model BL(1,0,1,1) with normally distributed errors. The study recommends that for bilinear time series data with normal innovations which have missing values, OLE estimates be used in estimating the missing values.

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.

## References

- [1] Abdalla, M.; Marwalla, T.(2005). The use of Genetic Algorithms and neural networks to approximate missing data. *Computing and Informatics* vol. 24, 5571-589.
- [2] Abraham, B. (1981). Missing observations in time series. *Comm. Statist. A-Theory Methods*.
- [3] Abraham, B.; Thavaneswaeran, A. (1991). A Nonlinear Time Series and Estimation of missing observations. *Ann. Inst. Statist. Math.* Vol. 43, 493-504.
- [4] Bishop, C. M.(1995). *Neural Networks for pattern recognition*. Oxford: Oxford University Press.
- [5] Cao, Y; Poh K, L and Wen Juan Cui, W, J.(2008). A non-parametric regression approach for missing value imputation in microarray. *Intelligent Information Systems*. pages 25–34.
- [6] Cheng and D. M. Titterington D M(1994): *Neural networks: review from a statistical perspective*.
- [7] Cheng, P.(1994). Nonparametric estimation of mean of functionals with data missing at random. *Journal of the American statistical association*, 89, 81-87.
- [8] 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.
- [9] De Gooijer, J.C.(1989) Testing Nonlinearities in World Stock Market Prices, *Economics Letters* v31, 31-35
- [10] Granger, C. W; Anderson, A. P.(1978). An Introduction to Bilinear Time Series model. *Vandenhoeck and Ruprecht: Guttingen*.
- [11] Hannan, E J. (1982). "A Note on Bilinear Time Series Models", *Stochastic Processes and their Applications*, vol. 12, p. 221-24.
- [12] Hellem, B T (2004). Lsimpute accurate estimation of missing values in micro Array data with least squares method. *Nucleic Acids*, 32, e34.
- [13] Hirano, K; Imbens, G, W.; Ridder (2002). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71, 1161 - 1189.
- [14] Howitt, P. (1988), "Business Cycles with Costly Search and Recruiting", *Quarterly Journal of Economics*, vol.103 (1), p. 147-65.
- [15] Kim, J K and Fuller. W. (2004). Fractional hot deck imputations. *Biometrika* 91, 559-578.
- [16] 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.
- [17] Liu J. and Brockwell P. J. 1988. "On the general bilinear time series model." *Journal of Applied probability*, 25, 553–564.
- [18] Liu, J. (1989). A simple condition for the existence of some stationary bilinear time series.
- [19] Ljung, G. M. (1989). A note on the Estimation of missing Values in Time Series. *Communications in statistics simulation* 18(2), 459-465.
- [20] 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.
- [21] Maravall, A. (1983), "An application of nonlinear time series forecasting", *Journal of Business Economic Statistics*, 1, 66-74.
- [22] Mcknight, E, P; McKnight, M, K; Sidani, S.; Figueredo, A.(2007). *Missing data*. Guilford New York.
- [23] 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
- [24] Nassiuma, D. K.(1994). Symmetric stable sequence with missing observations. *J.T.S.A.* volume 15, page 317.
- [25] 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.
- [26] 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. Springer verlag. New York.
- [27] Oba S, Sato MA, Takemasa I, et al. A Bayesian missing value estimation method for gene expressioprofile data. *Bioinformatics* 2003; 19(16):2088-2096.
- [28] 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.
- [29] Pena, D., & Tiao, G. C. (1991). A Note on Likelihood Estimation of Missing Values in perspective. *Multivariate Behavioral Research*, 33, 545–571.
- [30] Pourahmadi, M. (1989) Estimation and interpolation of missing values of a stationary time series. *Journal of Time Series Analysis* 10(2), 149–69.
- [31] Priestley, M.B. (1980). State dependent models: A general approach to time series analysis. *profile data. Bioinformatics* 2003; 19(16):2088-2096.

- [32] Ripley, B. (1996). Pattern recognition and neural networks. Cambridge: Cambridge University Press.
- [33] Smith, K.W and Aretxabaleta, A.L (2007). Expectation-maximization analysis of spatial time series. *Nonlinear Process Geophys* 14(1):73-77.
- [34] Subba Rao, T. and Gabr, M.M. (1980). A test for non-linearity of stationary time series. *Time Series Analysis*, 1, 145-158.
- [35] Subba, R.T.; Gabr, M.M.(1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture notes in statistics, 24. New York. Springer.
- [36] Thavaneswaran, A.; 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.
- [37] Tong, H. (1983). *Threshold Models in Non-Linear Time Series analysis*. Springer Verlag, Berlin.
- [38] Troyanskaya, O, Cantor, M, Sherlock G, Brown, P, Hastie, T, Tibshirani R, Botstein D and Russ B. Altman1 (2001). Missing value estimation methods for DNA microarrays *BIOINFORMATICS* Vol. 17 no. 6 2001 Pages 520-525.
- [39] 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.