KABARAK UNIVERSITY
6TH ANNUAL INTERNATIONAL RESEARCH CONFERENCE

Efficiency of Nonparametric Estimators for Missing Observations of Bilinear Time Series with Gaussian Innovation

POTI OWILI ABAJA/ DANKT NASSIUMA / DR LUKE ORAWO

20/07/2016

Introduction / Background

A **time series** is defined as data recorded sequentially over a specified period.

ince the data are records taken overtime, missing observations in time series are very common. They may occur as a result of

ost records, deletion of outliers, calender effects and defective measuring instruments

eing unable to account for missing data has several limitations:

A severe miss-representation of the phenomenon under study

Introduction... cont.

- A prohibiting factor in the use of certain methodologies
  - Can cause havoc in the estimation and forecasting of linear and nonlinear data (Abraham and Thavenaswan,1991)
  - Imputation is a necessary part of preprocessing of time series data
  - Inputation is a procedure that is used to fill in missing values by using substitutes. In any incomplete dataset, the observed values
  - provide indirect evidence about the likely values of the unobserved ones. This evidence, when combined with certain

BILINEAR TIME SERIES MODEL

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

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

- Where θ, bij and are constants while is a purely random process and ɵ=1

Cont...

- It 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 (see Subba Rao and Gabr, 1984)
- It is a parsimonious and powerful nonlinear time series model. Researchers have achieved forecast improvement with simple nonlinear time series models.

Imputation of missing values

- Gupta and Lam (1996) recommends imputation approach as a means of solving the missing value problem. Imputation is defined as a procedure that is used to fill in missing values by using alternative values. It is a statistical technique that is used to estimate missing values in an irregular time series (Fung, 2006;).
- According to Abrahantes, et. al. (2011), imputation broadly comprises several techniques that have been developed to compute missing values.

Statement of the problem

- Most of the techniques used for estimating missing values have on been concerned with linear time series models. Imputations of missing values for bilinear time series models have not been adequately considered.
- An estimator for missing values was developed for only a particular order of the simple bilinear time series, BL (1 ,0 ,2 ,0). This method does not take into account the distribution of the innovation sequence of bilinear time series .

Cont… Statement of the problem

- The performance of nonparametric methods in imputation missing values is still unknown. Therefore this study sought to determine the efficiency nonparametric methods of artificial neural networks (ANN) and exponential smoothing (EXP) in estimating missing values for bilinear time series.

Study objectives

hese included

a) Estimate missing values for bilinear time series models using ANN

b) Estimate missing values for Bilinear time series using Exponential Smoothing

c) Compare the efficiency of the estimates obtained

Brief literature review

- 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, 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),Least absolute deviation imputation (LADimpute) (Cao and Poh, 2006).

Cont...

- Sometimes, the ANNs provide better alternatives than the other techniques for solving a variety of problems (Wenzel and Schröter, 2010; Pashova and Popova, 2011).
- It is also used in time series imputation where researches due to the reported benefits (Junninen, et al., 2004).
- Shukur and Lee (2015) 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.

Cont…

- Pachepsky and Yakov (2010) developed a model that incorporated artificial neural network for infilling missing values in time series meteorological data.
- 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.
- 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.

METHODOLOGY

- Data was obtained through simulation using computer codes written in R software.
- Three data points 48, 293 and 496 were selected at random and data at these positions removed to create ‘missing value(s)’ at these points.
- 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 to calculate the MAD and MSE as well as in obtaining estimates based on exponential smoothing.

Findings / Results

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

MISSING POSITION	MAD		MSE	
	ANN	EXP	ANN	EXP
48	0.84293	0.762066	1.223745	1.054167
293	0.900142	0.908151	1.259637	1.215906
496	0.93157	0.952156	1.363035	1.474815
Total	2.674642	2.622373	3.846417	3.744888
Mean	0.891547	0.874124	1.282139	1.248296

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

MISSING POSITION	MAD ANN	MAD EXP	MSE ANN	MSE EXP
48	1.135251	0.98159	2.620982	1.542025
293	0.870468	0.811869	1.602965	1.078504
496	0.862982	0.932996	1.214815	1.369218
Total	2.868701	2.726455	5.438762	3.989747
Mean	0.956234	0.908818	1.812921	1.329916

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

MISSING POSITION	MAD ANN	MAD EXP	MSE ANN	MSE EXP
48	0.946487	0.872985	1.57565	1.296111
293	0.871257	0.907073	1.246586	1.290295
496	0.948569	0.914346	1.46972	1.369107
Total	2.766313	2.694404	4.291956	3.955513
Mean	0.922104	0.898135	1.430652	1.318504

Findings

- Exponential smoothing estimates were more efficient than ANN estimates for normally distributed data
- Efficiency of the estimates did not improve with the sample size.

Conclusions

- For normally distributed data, exponential smoothing techniques gave more efficient estimates than ANN neural network estimates.
- The distribution of the data plays a role in the determination of the estimation technique

Recommendations

1. For normally distributed data, EXP estimates should be used instead of the ANN estimates.
2. Further research should be done to compare the efficiency of the nonparametric estimates when the innovation sequence follows other distributions such as stable infinite variance distribution