KABARAK UNIVERSITY
6TH ANNUAL INTERNATIONAL RESEARCH CONFERENCE

ESTIMATION OF MISSING VALUES FOR BILINEAR TIME SERIES MODELS WITH GARCH INNOVATIONS USING NONPARAMETRIC METHODS

POTI OWILI ABAJA

20/07/2016

Introduction / Background

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

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

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

Imputation is a necessary part of preprocessing of time series data

Introduction… cont.

mputation is a procedure that is used to fill in missing values by using substitutes. In any incomplete dataset, the observed values provide indirect

Being unable to account for missing data has several limitations:

- A severe miss-representation of the phenomenon under study
- 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)

.....Cont...

This study focussed on imputation of missing values for bilinear time series models, a class of nonlinear models. We imputed missing values for bilinear time series models with GARCH innovations using nonparametric methods

Bilinear time series models with GARCH errors

- 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…

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

- In this case the error term e follows the GARCH distribution
- 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).

Statement of the problem

- Bilinear time series models are an important class of the nonlinear time series that has been used to model financial data. These data are usually prone to occurrence of missing values.
- Different methods have been used to estimate missing values in time series analysis. These methods can be classified as parametric or nonparametric methods. The efficiency of the estimators based on the nonparametric methods has not been determined in the literature.

Cont… Statement of the problem

- Further, the imputation of missing values for bilinear time series models when the innovations have a GARCH distribution has not been considered. Therefore this study to obtained estimates of missing values for bilinear time series using two nonparametric methods of artificial neural networks and exponential smoothing.

Study objectives

hese included

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

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

c) Compare the efficiency of the estimates obtained

Brief literature review

- (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).
- It has been found that the exponential smoothing with a constant α=0.2 may produce better forecasts than those based on ARIMA models (Background Facts on Economic Statistics, 2013).

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 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 GARCH innovations

OSITION		ANN	EXP	ANN	EXP
		0.992834	0.82809	1.741109	1.293227
3		0.788114	0.719507	1.20582	1.023331
6		0.64545	0.680498	0.8832	0.863463
tal		2.426398	2.228095	3.83013	3.180021
ean		0.808799	0.742698	1.27671	1.060007

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

ISSING	MAD			MSE		
OSITION		ANN	EXP		ANN	EXP
		1.146296	1.260539		2.264935	2.609242
3		0.978307	1.110822		1.783063	2.251697
6		0.887704	1.045276		1.778419	2.052729
tal		3.012307	3.416637		5.826417	6.913668
ean		1.004102	1.138879		1.942139	2.304556

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

ISSING	MSE			MSE		
OSITION		ANN	EXP		ANN	EXP
		0.978307	1.110822		1.783063	2.251697
3		0.887704	1.045276		1.778419	2.052729
6		1.003744	1.300622		2.326084	3.63698
tal		2.869755	3.45672		5.887566	7.941406
ean		0.956585	1.15224		1.962522	2.647135

Findings

- the ANN estimates of missing values were the more efficient than fficient than EXP estimates

- This was not as clear cut as in the case of normal distribution since there was a model where exp proved more efficient
- It is also evident that efficiency of the estimates had a mixed correlation as the position of the missing data increased.

Conclusions

1. For Bilinear time series models with GARCH innovations Artificial neural networks gave more efficient estimates than EXP estimates.
2. 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