International Journal of Applied Mathematics

Volume 27 No. 5 2014, 431-436
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v27i5.1

A SIMULATION STUDY ON ESTIMATING TIME SERIES ARIMA MODELS BY SPLINES

Roshanak Alimohammadi

Department of Mathematics
Alzahra University
Vanak, Tehran, IRAN

Abstract: Prediction and interpolation may be considered as two major purposes of time series analysis. Selecting a proper model in time scope as a member of ARIMA models is an important task and is required many steps to obtain the proper model. Some nonparametric regression methods such as splines have many applications in various fields. In this article, spline methods are applied to estimate time series models in a simulation study. In the simulation study, some data sets are generated of various ARIMA models. Then, the basic ARIMA model that is considered for generating data is fitted to each of the data sets as the proper model and the fitness of the models are investigated. Besides, smoothing spline method is applied for obtaining the proper pattern of the same data sets. Furthermore, fitness of these methods is compared by Sum of Square Errors (SSE) criterion to determine the more appropriate method and determining performance of smoothing spline.

AMS Subject Classification: 62M10
Key Words: simulation, smoothing spline, time series, ARIMA models, nonparametric regression

1. Introduction

One of the most commonly used methods in the literature of time series analysis

Received: July 7, 2014

is ARIMA models. An Autoregressive Integrated Moving Average (ARIMA) model is generally referred to as an ARIMA(p, d, q), where parameters p, d and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model, respectively. The model of time series Z_t can be written as:

$$\Phi_P(B)(Z_t) = (1 - B)^d \theta_q(B) a_t, \tag{1}$$

where $\Phi_P(B)$ is the autoregressive polynomial part of the backward operator B of order p, $\theta_q(B)$ is the moving average polynomial of order q, d is the degree of differentiating and a_t is a white noise time series. For more information, the reader can be referred to Box and Jenkins, [4].

2. Smoothing Spline

Basically, minimizing penalized sum of squares criterion is the base of smoothing spline. The criterion includes two components, goodness of fit and smoothness as follows:

$$S(g) = \sum_{i=1}^{n} (y_i - g(t_i))^2 + \lambda \int_a^b (g''(t))^2 dt, \tag{2}$$

where g is any twice-differentiable function on interval $[a, b]$, g'' is the second derivative of g, λ is smoothing parameter and n is the sample size.

In relation (2), the penalized least squares estimator $\hat{g}$ is defined as the minimizer of $S(g)$ over the class of all twice differentiable functions g. This minimizer function is called smoothing spline.

The proper value of smoothing parameter (λ) may be choosen by two major methods, namely Cross Validation (CV) and Generalized Cross Validation (GCV). To study more details, the author can be referred to Green and Silverman [1].

3. Simulation Study

In this section, a simulation study is carried out to assess fitness of smoothing splines to time series data and its performance is compared with the related ARIMA model, that is the models witch is applied to generate data in simulation. As some references, simulation study of time series and spline models are discussed by Cryer and Chan [2], Pffaf [3], and Wang [5].

In the simulation study, sample sizes $n = 50$, $n = 100$, $n = 250$, $n = 500$ and $n = 1000$ of different ARIMA models are considered to assess the fitness of spline method.

Tables 1, 2 and 3 demonstrate the fitness of ARIMA and smoothing spline (S.S) methods for each of the data sets generated from the ARIMA models.

model	order	n=50 S.S	n=50 ARIMA	n=100 S.S	n=100 ARIMA
AR	1	5.21	40.94	19.66	90.28
AR	2	0	39.52	17.1	84.42
AR	3	0	37.37	15.15	84.82
MA	1	59.76	165.63	125.84	332.61
MA	2	59.2	624.4	118.84	1141.26
MA	3	68.45	2024.41	148.23	3903.28
ARMA	(1,1)	44.4	38.57	96.37	90.32
ARMA	(2,2)	45.17	32.63	99.77	84.1
ARMA	(2,1)	8.56	39.65	114.5	83.56
ARMA	(1,2)	31.79	39.91	66.24	85.59
ARIMA	(1,1,1)	2.63	40.97	10.87	84.42
ARIMA	(2,1,2)	2.42	32.8	9.74	74.751
ARIMA	(2,1,1)	0	39.85	8.89	83.65
ARIMA	(1,1,2)	6.78	40.16	13.5	85.62

Table 1: SSE of S.S and ARIMA for $n = 50$ and $n = 100$

Now, test of difference between means of fitted spline ($\hat{y_i}$) and the amount of simulated data (y_i), is done as follows:

$$H_0 : \bar{\epsilon} = 0 \quad \text{versus} \quad H_1 : \bar{\epsilon} \neq 0, \tag{3}$$

where $\epsilon_i = y_i - \hat{y}_i$ and $\bar{\epsilon}$ is the mean of ϵ's in the simulated data sets. The test is carried out for all of the simulated data sets of the ARIMA models. Table 4 demonstrates p-values of hypothesis test (3) of for all of the considered models.

However, normality of $\bar{\epsilon}$ is acceptable for all of the sample sizes based on Central Limit Theorem.

In Table 4, p-value of test (3) is obtained. The amounts in Table 4 show that for all of the data sets, H0 is accepted and the tests are not meaningful. Therefore, equality of means of the data sets and the fitted spline is accepted for all of the cases. Then, despite of generating the main data from AR, MA,

model	order	n=50 S.S	n=50 ARIMA	n=100 S.S	n=100 ARIMA
AR	1	77.36	252.03	286.02	471.23
AR	2	89.44	247.31	563.07	466.94
AR	3	74.48	241.84	1052.89	465.21
MA	1	337.28	913.52	675.56	1916.7
MA	2	333.4	3245.51	691.13	7079.86
MA	3	421.01	11828.63	872.6	25943.87
ARMA	(1,1)	257.2	251.86	490.4	469.64
ARMA	(2,2)	261.4	246.61	491.33	463.42
ARMA	(2,1)	298.66	243.87	537.69	465.88
ARMA	(1,2)	178.09	243.71	367.5	464.8
ARIMA	(1,1,1)	67.8	247.79	214.09	471.41
ARIMA	(2,1,2)	60.72	241.6	203.18	458.98
ARIMA	(2,1,1)	78.96	243.89	317.72	466.84
ARIMA	(1,1,2)	55.5	243.72	147.87	465.65

Table 2: SSE of S.S and ARIMA for $n = 250$ and $n = 500$

model	order	S.S n=1000	ARIMA n=1000
AR	1	775.66	1002.2
AR	2	1120.94	1000.31
AR	3	2190.74	994.49
MA	1	1381.77	3899.47
MA	2	1423.15	14450.12
MA	3	1810.19	52029.51
ARMA	(1,1)	1025.98	1002.14
ARMA	(2,2)	1025.46	988.58
ARMA	(2,1)	1114.79	998.38
ARMA	(1,2)	749.32	999.68
ARIMA	(1,1,1)	668.48	1002.24
ARIMA	(2,1,2)	725.79	990.81
ARIMA	(2,1,1)	902.02	998.38
ARIMA	(1,1,2)	637.34	999.69

Table 3: SSE of S.S and ARIMA for $n = 1000$

model	order	n=50	n=100	n=250	n=500	n=1000
AR	1	0.82	0.73	0.79	0.74	1
	2	0.93	0.82	0.95	1	1
	3	0.95	0.76	0.96	0.9	0.78
MA	1	1	1	1	0.87	0.82
	2	1	1	1	0.83	0.85
	3	1	0.78	0.85	0.85	0.84
ARMA	(1,1)	1	1	1	1	1
	(2,2)	1	1	1	1	1
	(2,1)	0.98	1	1	1	1
	(1,2)	1	1	1	1	1
ARIMA	(1,1,1)	0.82	0.71	0.73	0.67	0.8
	(2,1,2)	0.87	0.74	0.77	0.64	0.85
	(2,1,1)	0.96	0.71	0.88	0.8	1
	(1,1,2)	0.88	0.78	0.69	0.58	1

Table 4: P-Values of test of hypothesis (3)

ARMA and the generally model ARIMA, smoothing spline has a better performance for most of the data sets. On the other hand, for some situations that the fitted splines are not better, the above results for test (3) show that there are not meaningful difference between means of the real data sets and the fitted splines.

4. Conclusion

In this paper, smoothing spline approach is applied for modeling of time series data in a simulation study. In this purpose, some data sets are generated from various ARIMA models. Then, performance of two fitted models is assessed, these methods include original ARIMA models and smoothing spline. The paper results show that spline has an acceptable performance for modeling of time series data in all of the considered situations. It is an interesting result because of data sets are generated from the specified ARIMA model, but smoothing spline has a good fitness for the data from ARIMA models. Finally, application of smoothing spline is suggested for modeling of time series data, because of simplicity and precision of the method.

Acknowledgments

This research is supported financially by Alzahra University.

References

[1] P.J. Green, B.W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London (1994).

[2] J.D. Cryer, K.S. Chan, *Time Series Analysis with Applications in R*, Springer, New York (2008).

[3] B. Pfaff, *Analysis of Integrated and Cointegrated Time Series with R*, Second Edition, Springer, New York (2008).

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

[5] Y. Wang, *Smoothing Spline, Applications and Methods*, Chapman and Hall, CRC Press (2011).