مجلّه‌ی پژوهش‌های آماری ایران
J. Statist. Res. Iran 7 (2010): 213–222
دوره‌ی ۷، شماره‌ی ۲، پاییز و زمستان ۱۳۸۹، صص ۲۱۳–۲۲۲
پژوهشکده‌ی آمار

Using Wavelets and Splines to Forecast Non-Stationary Time Series

Mina Aminghafari* and Shokoufeh Roosta

Amirkabir University of Technology

Abstract. This paper deals with a short term forecasting non-stationary time series using wavelets and splines. Wavelets can decompose the series as the sum of two low and high frequency components. Aminghafari and Poggi (2007) proposed to predict high frequency component by wavelets and extrapolate low frequency component by local polynomial fitting. We propose to forecast non-stationary process using splines based on this procedure. This method is applied to forecast simulated data and electricity load consumption of two regions. Result of the study show, the proposed method performance is better than the local polynomial fitting.

Keywords. Wavelet transform; splines; forecasting; non-stationary time series.

MSC 2010: 62M20, 62M10, 65T60, 65D07.

1 Introduction

Classical approaches to forecast non-stationary time series need to perform the preprocessing, such as differentiating and filtering to transform a non-stationary process to stationary one. Selection of an appropriate transform is not always easy.

In the recent years, application of wavelets developed very rapidly in different domains such as statistics and image processing due to their well properties. Considering local property of wavelets, they are well suited to

* Corresponding author

handle non-stationary time series. A non-stationary process can be converted to stationary one, using a regular wavelet.

Several authors considered wavelets in time series domain, see e.g. Dahlhaus et al. (1995), Percival and Walden (2000). Although many approaches of time series use wavelets, there is less attention to forecasting time series. The first approach which uses wavelets for prediction is based on wavelet spectrum as a local version of Fourier spectrum (Fryzlewicz et al., 2003). On one hand, this approach especially designed for the class of locally stationary wavelet process and on the other hand, their parameters selection algorithm seems not to be convergence. Renaud et al. (2003) used another approach which estimates the prediction parameters by direct regression of the stationary process on Haar non-decimated wavelet coefficients depending on its past values. Nevertheless, this approach for stationary process and classical models differ from parameterization viewpoint. Aminghafari and Poggi (2007), extended this approach using more regular wavelets and extrapolated the low frequency component (using local polynomial fitting) to forecast non-stationary signal.

We assume the observation Y_t, can be decomposed to two parts as follows: $Y_t = f(t) + X_t$ where X_t is a stochastic process and $f(t)$ is a deterministic smooth function. If the level of decomposition is suitably chosen, the deterministic part of signals (trend or smooth part) can be estimated by polynomial fitting on the approximation coefficients. Also purely stochastic part can be predicted by detail coefficients (Aminghafari and Poggi, 2007).

On the other hand, local polynomial fitting is a poor technique to extrapolate a signal. In this paper, we propose to bring together the existing method and extrapolation with splines. After a short review on wavelets, we describe the former approach briefly in Section 3. In Section 4, we describe spline extrapolation and in Section 5, we compare our approach with classical one and the proposed approach by Aminghafari and Poggi (2007) through simulated and real data.

2 Non-decimated Wavelet Transform

Wavelet transform is a powerful tool in the non-parametric statistics and signal processing. This transform is local. It can analyse the data in both time and frequency aspect simultaneously.

The usual wavelet transform which is often used in the estimation context is discrete wavelet transform (DWT). Although, this transform has many good properties (such as performing an orthonormal basis) but it is not appropriate for time series purposes. This transform is not a shift invariant transform. It means when we observe a new observation, all wavelet coefficients have to be recomputed. To overcome this drawback, in this article, we use non-decimated wavelet transform, which is shift invariant. The wavelet coefficients at a specified level can be computed from previous one using a recursive formula. The non-decimated wavelet coefficients of $X = (X_1, \ldots, X_n)$ are defined for level $j \geqslant 1$ as follows:

$$c_{j,t} = \sum_{l=0}^{L-1} h_l c_{j-1,(t-2^{j-1}l)_+} \tag{1}$$

$$w_{j,t} = \sum_{l=0}^{L-1} g_l c_{j-1,(t-2^{j-1}l)_+} \tag{2}$$

where and h and g are the low-pass and high-pass filters associated to selected wavelet, normalized in $\mathbb{L}^1$, L is the filter length and $(t-2^{j-1}l)_+$ is positive part of $(t-2^{j-1}l)$. Let us note that the signal X can be considered as initial value of this algorithm i.e. $c_0 = X$. The approximation coefficients $(\{c_{j,t}\}_{1\leqslant j\leqslant J,t})$ and detail coefficients $(\{w_{j,t}\}_{1\leqslant j\leqslant J,t})$ have low frequency and high frequency nature respectively. Let us note that J is decomposition level and t is time. For more details on non-decimated wavelet (NDW) transform, the reader can refer to Percival and Walden (2000), Chapter 5.

3 The Basic Method

In this section, we describe briefly the proposed method for forecasting non-stationary time series in Aminghafari and Poggi (2007). For non-stationary time series, consider an observed time series of the form $Y_t = X_t + f(t)$, where X_t is a purely stochastic time series and $f(t)$ is a deterministic component. From the reconstruction equation, we have $X = A_J + \sum_{j=1}^{J} D_j$, (see Percival and Walden (2000), p. 173), then we can write:

$$Y_t = (A_J(X))_t + \left(\sum_{j=1}^{J} D_j(X)\right)_t + f(t), \qquad 1 \leqslant t \leqslant N \tag{3}$$

where $A_J(X)$ and $D_j(X)$ are respectively suitably chosen reconstructed version of approximation and detail coefficients of level j, which obtained by

non-decimated wavelet (NDW) coefficients of X. If we choose a convenient wavelet i.e. using a wavelet which has enough zero moments, the details are supposed to be free of f. Hence, the term $D_t = (\sum_{j=1}^{J} D_j(X))_t$ can be predicted using a regression between observations and their past wavelet coefficients $\{w_{j,k}\}_{k \leqslant t-s}$.
To predict Y_{N+1}, when we have observed $Y_1, \ldots, Y_N$, the prediction equation is considered as follows:

$$\hat{Y}_{N+1} = \hat{D}_{N+1} + \hat{Z}_{N+1} \tag{4}$$

where $\hat{D}_{N+1}$ represents the prediction of high frequency stationary components i.e. D_t and $\hat{Z}_{N+1}$ is extrapolation of the rest i.e. extrapolation of $Z_t = Y_t - \widehat{(\sum_{j=1}^{J} D_j(X))_t} = Y_t - \widehat{D_t}$ which is an estimation of $(A_J(X))_t + f(t)$. The quantity $\hat{D}_{N+1}$ can be written as follows:

$$\hat{D}_{N+1} = P_N^T \alpha$$

where P_N contains wavelet coefficients until time N and α is the vector of parameters to be estimated by minimizing the empirical mean square prediction error (MSPE) defined as follows:

$$\begin{aligned} P_t &= [w_{1,t}, \ldots, w_{1,t-2r_1}, \ldots, w_{J,t}, \ldots, w_{J,t-2^J r_J}]^T, \quad t = 1, \ldots, N \\ \alpha &= [a_{1,1}, \ldots, a_{1,r_1}, \ldots, a_{J,1}, \ldots, a_{J,r_J}]^T \end{aligned} \tag{5}$$

Let us note that α can be estimated by minimizing another criterion such as mean absolute prediction error. In the presence of outliers in the data this criterion is preferred to be used. Since in our references MSPE is used, we use this one. Signal Z_t can be extrapolated by various deterministic or stochastic ways. For example, using polynomial fitting (Aminghafari and Poggi, 2007), using Kernel smoothing (Aminghafari and Poggi, 2011) or by splines which is proposed in this paper. In the next section, we describe the extrapolation technique using cubic splines. When Z_{N+1} is extrapolated, it replaced in equation (4), then the final prediction of $\hat{Y}_{N+1}$ will be obtained where N is the number of observations.

4 Extrapolation with Spline

Splines are useful tools for interpolation and extrapolation of signal. Splines are piece-wise polynomial, hence, they have local nature in the analysis of signal. For more details about splines, see Hastie et al. (2001) Chapter 5.

Let us denote the extrapolation of a signal Z_t at time $t+1$ by $\hat{Z}_{t+1}$ using a cubic spline. $\hat{Z}_{t+1}$ can be written as follows:

$$\hat{Z}_{t+1} = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3 + \sum_{k=1}^{K} \beta_{4+k} (t - \xi_k)_+^3 \tag{6}$$

where ξ_k's are knots, K is the number of knots and $(t-\xi_k)_+$ is positive part of $(t - \xi_k)$. For the given knots, we estimate β's by minimizing empirical mean square prediction error. To extrapolate or interpolate using splines, one needs to select the number of knots and their position. In the following remark, we describe the method that we use for this purpose.

Remark 1. To select optimal knots, we use an iterative algorithm based on the Gauss Newton method. For details, see Gallant and Fuller (1973) and Björck (1996). To this end we use the `optknt` function of MATLAB which is based on algorithm described in de Boor (1997). This function needs the initial values. For selecting the initial knots we propose to divide the data into intervals of equal lengths. We call this interval a knot interval. For each interval, we choose two knots, which are the relative maximum and minimum number of the data in that interval. For example, if we have 1000 data as a training set and we choose an interval of length 200, i.e. we have 10 knots overall.

In this approach, one needs to determine the length of interval. We use a similar approach to Aminghafari (2006), a cross-validation technique. Suppose that we observed n observations. We consider $n-m$ observation as training set and the last m sets as test set. We fit cubic spline to training set considering the proposed interval length and extrapolate the test set one by one. We select the length which has the best error criterion on the test set.

5 Experimentation

Now, we explain the framework of our experimentation. We compare following methods in this paper:

- First method: forecasting based on autoregressive model denoted by AR in the tables;

Table 1. Prediction performances of different methods for prediction of a sinusoidal data

Method	Parameter	$\bar{R}$
AR	$p = 14$	1.15
Wavelet+Polynomial	window=10	**1.05**
Wavelet+Spline	knot interval=200	1.11

- Second method: forecasting based on wavelet and polynomial (proposed method by Aminghafari and Poggi, 2007) denoted by Wavelet + Polynomial in the tables;

- Third method: forecasting based on wavelet and spline (our proposed method) denoted by Wavelet + Polynomial in the tables.

Let us recall that we use cubic spline to extrapolate Z_{t+1}. We use Daubechies wavelet with 2 vanishing moments, $db2$ wavelet and reserved filter. This wavelet is orthogonal to polynomial of order 2. We use also the root of an unbiased estimator of mean square prediction error as the criterion for comparing these methods as follows:

$$R = \sqrt{\frac{1}{N-n}\sum_{t=n+1}^{N}(X_t - \hat{X}_t)^2} \tag{7}$$

where N represents the number of total data, $N - n$ represents the number of data to be predicted based on each method, X_t is the real value and $\hat{X}_t$ is its prediction.

5.1 Simulated Data: A Sinusoidal Function Contaminated by Gaussian Noise

We simulate 30 realizations of time series of $N = 1000$ from the model of a sinusoidal function contaminated by Gaussian noise with zero mean and variance equal to 1. In this data, we choose onward, the first 904 data as training set and the last 96 data points as test set. We perform the one step ahead prediction methods on test set. For this data, we define $\bar{R}$ as the mean of R on 30 simulated realizations. The results are given in Table 1 wherein p is the order of AR model.

The best prediction method for this data is obtained for polynomial and wavelet method. Our proposed method has better result than AR prediction.

Table 2. Prediction performance of our method for different knot intervals for FEC data

knot interval	$\boldsymbol{R}$
100	417.31
125	411.21
150	409.90
175	409.95
200	**409.28**

Table 3. Prediction performances of different methods for prediction of FEC data

Method	**Parameter**	$\boldsymbol{R}$
AR	$p = 31$	971.48
Wavelet+Polynomial	window=10	414.20
Wavelet+Spline	knot interval=200	**410.14**

5.2 France Electricity Consumption Data

To examine the performance of our method, we consider France electricity consumption (FEC) data. The consumption is measured from first August 1985 to 4th July 1992, every half an hour. We divide the data in two parts: we take the first 1344 data as past data and then we predict 96 points. In Table 2, the performance of our method is computed for different knot intervals.

The proposed method select the best knot interval equal to 200. In Table 3, R has been computed for different methods. Our proposed method has the best performance. The second method i.e. wavelet and polynomial has performance near to our method. As expected, fitting AR model to this data leads to an inadequate result.

5.3 Tehran Electricity Consumption Data

This data set is collected on the first region of Tehran electricity consumption (TEC) every hour from 22 Mars 2009 to 22 September 2009. We consider the last 1440 data and divide the data in two parts: we take the first 1344 data as past data and we predict 96 points. In Table 4, the performance of our method is computed for different knot intervals.

Table 4. Prediction performance of our method for different knot intervals for TEC data

knot interval	***R***
100	164.86
125	163.40
150	163.09
175	162.68
200	**161.43**

As shown in Table 4, the approach propose to select knot interval equal 200 for which we have the best performance. In Table 5, R is given for different methods. The best performance is obtained for our proposed method. The performance of wavelet and polynomial is near to performance our method. AR has the worst performance. Figure 2 shows TEC data and its prediction based on proposed method for the knot interval equal to length 200.

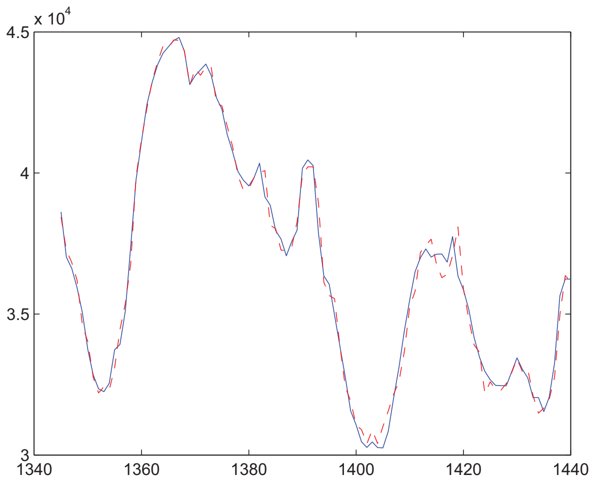

Figure 1. FEC data to be predicted (line) and its prediction using our method (line dots)

Table 5. Prediction performances of different methods for TEC data

Method	Parameter	R
AR	$p = 27$	190.11
Wavelet+Polynomial	window=10	165.05
Wavelet+Spline	knot interval=200	**162.43**

6 Conclusion

As expected, our proposed method leads to better result on real data than the method based on polynomials and wavelets. For simulated data, these two methods have the similar results. The proposed method seems interesting and needs more study. The proposed method does not need any preprocessing as applied in the classical method and does not need to know if the data is stationary or not. In the classical method, we need to perform some transformation to change the data to stationary one. If this transformation does not selected correctly, the results can be inadequate. Two directions for future work will be considered. In this paper, we use cubic spline. First, it seems we can apply other splines such as B-splines. Second perspective can be studying on knots selection. We expect applying a better strategy to choose knots leads to a better result.

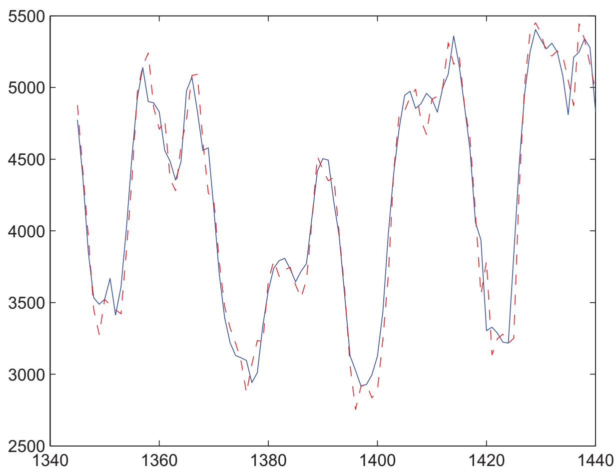

Figure 2. TEC data to be predicted (line) and its prediction using proposed method (line dots)

References

Aminghafari, M. (2006). *Methodes d'ondellettes en statistique des signaux temporels uni et multivaries*. PhD thesis, Paris 11 university, Orsay, France.

Aminghafari, M. and Poggi, J.-M. (2007). Forecasting time series using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, **5**, 709-724.

Aminghafari, M. and Poggi, J.-M. (2011). Non-stationary time series forecasting using wavelets and kernel smoothing, To appear, *Communications in Statistics: Theory and Methods*.

Björck, A. (1996). *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia.

Dahlhaus, R., Neumann, M.H. and Von Sachs, R. (1999). Nonlinear wavelet estimation of time varying autoregresive processes. *Bernoulli*, **5**, 873-906.

de Boor, C. (1977). Computational aspects of optimal recovery, in *Optimal Estimation in Approximation Theory*, Micchelli, C.A. and Rivlin T.J. eds., 69-91, Plenum Publ., New York.

Fryzlewicz, P., Van Bellegem S., and Von Sachs R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, **55**, 737-764.

Gallant, A.R. and Fuller, W.A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association* , **68**, 144-147.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag, New York.

Percival, D.B. and Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis*, Cambridge University Press, London.

Renaud, O., Starck, J.L. and Murtagh, F. (2003). Forecasting based on a multiscale decomposition. *International Journal of Wavelets Multiresolution and Information Processing*, **1**, 217-232.

Mina Aminghafari
Faculty of Mathematics and
Computer Science,
Amirkabir University of Technology,
Tehran, P.O. Box: 15875-4413,
Iran.
email: *aminghafari@aut.ac.ir*

Shokoufeh Roosta
Faculty of Mathematics and
Computer Science,
Amirkabir University of Technology,
Tehran, P.O. Box: 15875-4413,
Iran.