7-1-2006

Wavelet based SARIMA models for forecasting natural gas demand

S.M. Aqil Burney
Higher Education Commission, Government of Pakistan

Afzal Saleemi
Department of Computer Science, University of Karachi, Pakistan

Syed Akhter Raza

DISCUSSION

Wavelet based SARIMA models for Forecasting Natural Gas Demand

S.M. Aqil Burney
Major Afzal Saleemi
Syed Akhter Raza

ABSTRACT

Time series analysis has attracted a lot of attention of researchers in recent times due to availability of sophisticated computing facilities. In this paper wavelet transformation and seasonal ARIMA methodology have been used to analyze and forecast time series. First we analyze time series data for gas demand of Sui southern Gas Company (SSGC) of Pakistan and forecast with Box-Jenkins SARIMA models then we look at wavelet-based multiresolution analysis (MRA) and SARIMA models predictions using the compressed and de-noised signals. With the right choice of mother wavelets, this method is very successful in analyzing and forecasting time series. In the later part of this paper we compare forecast performance of the three models in consideration.

Key Terms: *Time Series Analysis, SARIMA, Wavelets, Multiresolution Analysis*

1. Introduction

M odeling and forecasting the time series data present a number of challenges to business community. The various financial derivatives used to hedge exposure, to the often abnormally high and rapidly mean-reverting or non- stationary time series require an understanding of the process and the ability to accurately estimate and forecast future trends in the series. In this study to achieve this goal, we decompose our original series using a robust smoother wavelet transform. Further, we reconstruct the original series using wavelet shrinkage procedure to obtain filtered compressed series of gas demand.

We examine the gas demand series at different time locations and levels of resolution to reveal and differentiate between signal and noise. Further, we cleanse the data into the more fundamental levels of frequency resolution. As it is from these levels that we base the reconstruction of our filtered series, we need to ensure that they are minimum contaminated by noise. Using the filtered data, we explore time series models as possible candidates for explaining the demand process and evaluate their forecasting ability. These models include the Seasonal ARIMA and wavelet based compressed and de-noised models. What we find is that, models from the Wavelet class using the filtered data produce the best forecasts as compared to traditional SARIMA model using the unfiltered or original data.

Wavelet analysis in time series has attracted the attention of lot of researcher in recent past in the field of data mining [7], mathematics [5], statistics [3] and computer sciences [6][7]. Various methods, such as linear regression, Fourier transform based methods

and ARIMA models have been used to analyze time series. All these models are based on the primary assumption of stationarity and require transformations of nonstationary data for analysis. Both linear and non-linear models have limitations when it comes to analyzing nonstationary data. The most frequently used Fourier transformation is able to reveal the frequencies present in a signal but it is not possible to say when they are present. This was the starting point for the wavelet transformation. Using the wavelet transformation a time-scale representation of a time series signal is possible. That means wavelet analysis is the breaking up of a signal into its shifted and scaled versions.

2. Wavelet MRA in Time Series Analysis

The wavelet multiresolution analysis (MRA) is the process of decomposition of a discrete signal into approximate and detailed signals at each time scale through a series of scaling functions $\phi_{j,k}(x)$ and wavelet functions $\psi_{j,k}(x)$, where $k \in Z$ and its reconstruction at the finest resolution[5]. These scaling and wavelet functions are obtained by dilating and translating the mother scaling function ϕ and mother wavelet function ψ as under:

$$\phi_{j,k}(x) = 2^{\frac{-j}{2}} \phi(2^{-j} x - k)$$

$$\psi_{j,k}(x) = 2^{\frac{-j}{2}} \psi(2^{-j} x - k)$$

$$f(x) = \sum_k c_k^{j0} \phi_{j0,k}(x) + \sum_{j>j0} \sum_k d_k^j \psi_{j,k}(x) \qquad (1)$$

here $f(x)$ is the function of approximations and details of discrete signal $x(t)$. The sum with coefficients c_k^{j0} and d_k^j shows the approximation at the coarser scale and details at all scales respectively. It means the wavelet function is scaled and shifted along the axis and therefore the signal, too. This time-scale joint decomposition structure makes wavelet analysis interesting for time series analysis [2] [3].

3. Applications of Wavelet in Time Series

Wavelets offer a variety of possible applications on time series. Compression and De-noising are two its most popular functions.

(i) Compression of Signal using Wavelets

Wavelet analysis is used for filtering and smoothing of data. A typical signal may have many large values spread out over time whereas its wavelet transformation may tend to have a few large coefficients that contain the maximum energy of the signal and many small coefficients, which are relatively unimportant. The DWT is defined in terms of a wavelet filter or mother wavelet and an associated filter known as scaling filter or father wavelet. These are band pass filters; hence the MRA and synthesis of a discrete signal x_t can be implemented using low pass and high pass filters [2][3]. The extraction of

135

estimated signal using inverse DWT from few selected coefficients greater in magnitude than a defined threshold is called compression or shrinkage of signal [3]. In wavelets maximum energy is concentrated in first few coefficients, hence the decomposed coefficients below a threshold are set to zero and a compressed signal is reconstructed on the basis of rest of the approximated and detailed coefficients. Our aim of this paper is to give a comparative study of the Box-Jenkins SARIMA models [4] with wavelet based SARIMA using compressed signal discussed in next section.

(ii) De-noising Using Wavelets

Many methods have been developed for smoothing a signal, in the hope that the noise can be suppressed and the significant patterns retained and revealed. These have ranged from simple moving averages or moving medians to methods of considerable mathematical complexity. Wavelets offer a smoothing approach, which is relatively simple to use, while adapting well and automatically to the form of the signal being smoothed. Here, we concentrate on one way in which smoothing using wavelets may be achieved i.e., by de-noising transformed signal using threshold. We used a de-noised signal for forecasting gas demand of SSGC and also compare the results.

4. Box-Jenkins SARIMA Modeling

The steps in developing the ARIMA model include identification, estimation and then diagnostic checks on the identified model [4]. If the diagnostics do not endorse the selected model, the process is started over again. The most useful tools in model identification are the sample autocorrelation function and the sample partial autocorrelation function. The autocorrelation function is used to identify the autoregressive process (AR), and the sample partial autocorrelation function is used to identify the moving average process (MA). Mathematically the pure ARIMA model is written as

$$Y_t = \mu + \frac{\theta(B)}{\varphi(B)} a_t \qquad (2)$$

Y_t = Response series a_t = Independent disturbance

$\phi(B)$ = Autoregressive operator $\theta(B)$ = Moving average operator

We applied many seasonal ARIMA models on SSGC gas demand series and found SARIMA $(0, 0, 1)(0, 1, 1)_{12}$ suitable[1].

5. Data and Results

The time series used in this study was collected from Sui Southern Gas Company during the period from July 1999 to June 2004[1]. The observations in the series

136

are the average of daily demand of gas for a month. Among the various Wavelets used in this study we found db4 is the best, details are not given due to brevity. Fig. 1 shows the original signal and its decomposition using the multiresolution analysis for db4 [5].
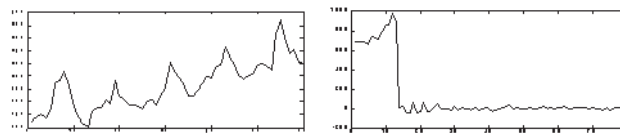

Fig. 1 Original and decomposed signals using db4

Analysts depending on the problem in hand generally choose the compression rate. We have compressed the original signal while using 16.46% of the transformed coefficients with compression rate of 0.167 (Fig. 2a). The compressed signal allows us to trade off a small amount of precision for very large gains in speed and retained 99.67% of total energy.
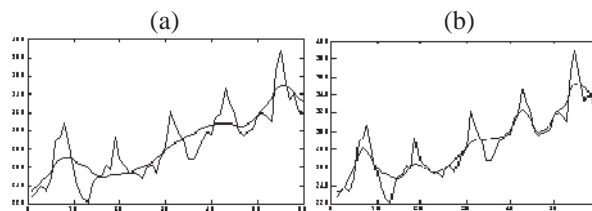

Fig. 2. (a): Original and compressed (b): Original and de-noised signal using db4

The comparison of the original and the filtered signal (Fig. 1) shows that the low frequency part is completely extracted. The compression process removed most of the noise from signal (Fig. 2a). De-noising is also a quite similar phenomenon to the wavelet filtering. After synthesizing the signal using approximation coefficients at coarser scale and detail coefficients at all scales, signal still contains the noise which can be extracted by threshold see [3]. Again we used a db4 wavelet and global threshold for transformation and de-noising a signal to get its smooth version. We removed the noise without compromising sharp details (Fig. 2b) of the signal, which is the strength of wavelet analysis and used de-noised signal for forecasting.
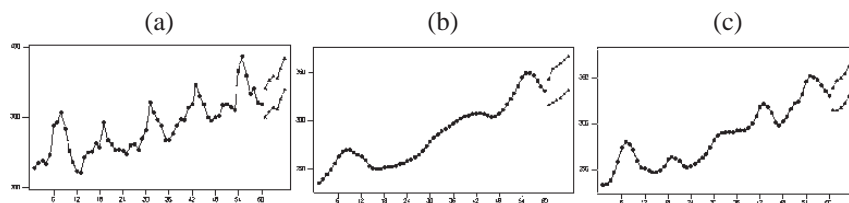

Fig. 3: Actual and forecast using SARIMA (a): On actual series (b): On compressed series by db4 (c): On de-noised series by db4 with 95% confidence intervals

137

6. Forecasting

We used the Box-Jenkins seasonal ARIMA model discussed in [1] for our comparative study of wavelet based SARIMA for both compressed and de-noised signals. Table 1 and Fig. 3 clearly show that forecasts using compressed and de-noised series have been significantly improved. Results show that SARIMA model on the de-noised signal is the best for series under study and gives minimum SSE, AIC, BIC and maximum value of adjusted R^2, however what is important is wavelet analysis with compressed model has given even better forecast than linear Box-Jenkins SARIMA Model which may prove very economical and speedy in future for large data bases applications. The forecasted values for next six months are given in Table 1.

Analysis	(a)	(b)	(c)
R-Squared	0.753	0.764	0.907
Adjusted R^2	0.742	0.754	0.903
S.E.	9.244	8.91789	4.989
SSE	3845	3578	1120
Mean dep. var	19.208	19.208	19.308
S.D dep. var	18.183	17.966	15.996
AIC	7.346	7.274	6.229
BIC	7.463	7.391	6.229

Period	(a)	(b)	(c)	Actual
61	320.054	329.542	326.765	333
62	330.078	335.878	330.469	348
63	336.333	338.411	333.816	362
64	333.314	341.300	338.229	363
65	347.118	344.837	345.879	353
66	361.307	349.070	356.803	431

Table 1: Comparative Results and Forecasts using
(a) Original (b) Compressed (c) De-noised series along with actual values

7. Conclusion

We compared forecasts using traditional seasonal ARIMA model with wavelet based SARIMA models for natural gas demand data and believe that moving towards a wavelet based forecasting system would significantly improve the forecasting efficiency for nonstationary time series available through the traditional stationary models. We also conclude that compression of original time series signal using wavelets also improve the speed and efficiency of forecast model as compared to direct application of traditional time series analysis techniques.

REFERENCES

Burney, S.M. Aqil, Syed Akhter Raza (2004). Time Series Forecasting With SARIMA Model, Conference Proceedings, International Conference on Software Engineering & Applications, Islamabad.

Misiti, M. Misiti, Y. Oppenheim, G. Poggi, J.-M. (2000). Wavelet Toolbox User's Guide, The Math Works, Inc.

138

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

Box, G.E.P., Jenkins, G.M. & Reinke, G.C. (2003). Time Series Analysis Forecasting and Control, 3rd Edition, Pearson Education Inc.

Daubechies, I. (1992). Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Stollnitz, E.J., DeRose, T.D. & Salesin, D.H. (1996). Wavelets for Computer Graphics: Theory and Applications, Morgan, San Francisco.

Tao Li, Qi Li, Shenghuo Zhu, Mitsunori Ogihara (2002). A Survey of Wavelet Applications in Data Mining ACM SIGKDD Explorations Newsletter, Volume 4, Issue 2, pages 49-68.

Corporate Responsibility means:

"The obligations to pursue those policies to make those decisions or to follow those lines of action which are desirable in terms of the objectives and values of our society."
Howard R. Bowen. Social Responsibilities of the Businessman, New York: Harper and Brothers.