Analysis of Subdivision Wavelet for Time Series Model

Limei Zhang[1] Bo Wang[2] Xuebiao Wang[3] Junde Yang[4] Jiuqi Li[2]

1. School of Science, 2. The Ocean Engineering College, 4. School of Mechanical Engineering, Dalian Ocean University, Dalian Liaoning China 116023 zlm@dlfu.edu.cn;

3.College of Math and Quantitative Economics of Dongbei University of Finance & Economics

Abstract

HP filter used for separating trend and disturbance from time series data and given by Hodrick and Prescott has been widely used in research in quantitative economics in recent years. Economists identify with Hp filter method when dispose the seasonal data and have differences on the other sort of data. So a new filter method came from subdivision wavelet is given in the paper which is used instead of HP filter. According to the data of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006, the trend and the disturbance of the two time series are separated apart by the decomposition rule of subdivision wavelet. Then the reconstruction of the two time series is executed respectively by subdivision and subdivision wavelet. Taking note of the features of the initial data and the reconstruction data of subdivision wavelet of the two time series, the auto-regressive moving average (ARMA) model is used for the data of marine fishing production modelling and the Logistic growth curve is used for the aquaculture production in marine waters modelling. The results of modelling demonstrate the validity of the methods presented in the paper.

Keywords: HP filter; subdivision wavelet; Logistic model; ARMA model

1 Introduction

Time-series analysis is important for a wide range of disciplines transcending both the physical and social sciences for proactive policy decisions. Statistical models have sound theoretical basis and have been successfully used in a number of problem domains in time series forecasting. Due to power and flexibility, Box-Jenkins ARMA model has gained enormous popularity in many areas and research practice for the last three decades. In the course of modeling, the variation of time series data is described by extrapolation. Measurement of smoothness of time series data is needed before modeling. For the non-smooth time series some transform such as difference for the data is necessary.

HP filter presented by Hodrick and Prescott[1] are widely used for separating trend and disturbance from time series data in the research in quantitative economics in recent years. HP filter is devised in accordance with the principal of mean shift for symmetrical data. The trend part g_t can be separated from the initial time series y_t and g_t are the solution of the following formula

$$Min\{\sum_{t=1}^{T}(y_t - g_t)^2 + \lambda\sum_{t=1}^{T}|(g_t - g_{t-1})(g_t - g_{t-2})|\}. \quad (1)$$

Where the first item of the polynomial in the big bracket represents the measurement of undulation of the initial data y_t comparing with the trend part g_t and the second item of the polynomial gives the extent of smoothness of the trend part. The positive number λ is defined as smooth parameter and it is used to regulate the proportion between the two parts. The important problem of HP filter method is the choice of the smooth parameter λ. Different smooth parameter λ leads to different period and smoothness. Economists identify with Hp filter method when dispose the seasonal data and have differences on the other sort of data.

The goal of the paper is to implement a new filter method which is used instead of HP filter for time series analysis. In the method the trend and the disturbance can be separated from initial time series by a kind of subdivision wavelet, which is widely used in computer aided geometric design. The results of the data of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006 disposed by the new method demonstrate the validity of our scheme. By using preliminary software testing of Eviews and Matlab according to time series theory, ARMA and Logistic models are suitable for modeling the two time series data.

2 Subdivision and subdivision wavelet

The wavelet transform has emerged as an exciting tool for signal and image processing including detection, de-noising, filtering and synthesis. Subdivision refers to a class of modeling schemes that define an object through recursive refinement starting from an initial control polygon (or net). Similar to B-splines, the final curve (or surface) is defined by the vertices of the initial control polygon (or net) and the selective subdivision algorithm. The basic idea of subdivision [2] is to define a smooth curve (or surface) as the limit curve (or surface) of a subdivision process in which an initial control polygon (or net) is repeatedly refined with new control polygon (or net) inserted. Subdivision scheme's good similarity among different subdivision layers makes the subdivision wavelet has better approximation between the trend signal and the initial signal than the ordinary wavelets. This is the reason why we use subdivision wavelet (the second generation wavelet) as the filter in the course of data disposal in this paper.

Suppose that P^k represents the column vector of the points on the curve. k is the number of the layers of the data in multi-representation. In some circumstance the column vector P^{k+1} with more points is needed for refining the curve. The process is called subdivision and the refinement of subdivision is as follow.

$$P^{k+1} = S^{k+1} P^k . \tag{2}$$

Where S^{k+1} is called matrix of subdivision.

Subdivision wavelet can be derived from the inverse calculation of subdivision. The trend P^k and the disturbance Q^k can be decomposed by P^{k+1} in terms of the first and the second equation in formula (3) respectively.

$$P^k = A^{k+1} P^{k+1}, Q^k = B^{k+1} P^{k+1} . \tag{3}$$

Meanwhile to reconstruct P^{k+1} via the trend P^k and the disturbance Q^k, the following formula can be used. That is

$$P^{k+1} = S^{k+1} P^k + T^{k+1} Q^k \tag{4}$$

When reconstruct P^{k+1} the $k+1$ layer column vector, the reconstruction refinement of subdivision wavelet i.e. formula (4) is better than the refinement of subdivision i.e. formula (3) because of the disturbance being treated in formula (4). Suppose matrix of subdivision S^{k+1} is known, the core of conformation of subdivision wavelet is finding matrices T^{k+1}, A^{k+1} and B^{k+1} satisfying the following equation

$$\begin{pmatrix} A^{k+1} \\ B^{k+1} \end{pmatrix} \begin{pmatrix} S^{k+1} & T^{k+1} \end{pmatrix} = \begin{pmatrix} I & O \\ O & I \end{pmatrix}$$, where I represents the identity matrix. (5)

The coefficients of subdivision wavelet in [3] are used for decomposition and reconstruction in the paper because of its batter approximation quality. For a column vector with 10 closed points, the coefficients of subdivision wavelet are enchased in the following matrices i.e.

$$S^{k+1} = \frac{1}{16} \begin{pmatrix} -1 & 13 & 5 & -1 & 0 \\ -1 & 5 & 13 & -1 & 0 \\ 0 & -1 & 13 & 5 & -1 \\ 0 & -1 & 5 & 13 & -1 \\ -1 & 0 & -1 & 13 & 5 \\ -1 & 0 & -1 & 5 & 13 \\ 5 & -1 & 0 & -1 & 13 \\ 13 & -1 & 0 & -1 & 5 \\ 13 & 5 & -1 & 0 & -1 \\ 5 & 13 & -1 & 0 & -1 \end{pmatrix} \qquad T^{k+1} = \frac{1}{16} \begin{pmatrix} 11 & -3 & 1 & 0 & -1 \\ 3 & -11 & 1 & 0 & -1 \\ -1 & 11 & -3 & 1 & 0 \\ -1 & 3 & -11 & 1 & 0 \\ 0 & -1 & 11 & -3 & 1 \\ 0 & -1 & 3 & -11 & 1 \\ 1 & 0 & -1 & 11 & -3 \\ 1 & 0 & -1 & 3 & -11 \\ -3 & 1 & 0 & -1 & 11 \\ -11 & 1 & 0 & -1 & 3 \end{pmatrix}$$

$$A^{k+1}=\frac{1}{16}\begin{pmatrix} 11 & -3 & -1 & 1 & 0 & 0 & 1 & -1 & -3 & 11 \\ -3 & 11 & 11 & -3 & -1 & 1 & 0 & 0 & 1 & -1 \\ 1 & -1 & -3 & 11 & 11 & -3 & -1 & 1 & 0 & 01 \\ 0 & 0 & 1 & -1 & -3 & 11 & 11 & -3 & -1 & 1 \\ -1 & 1 & 0 & 0 & 1 & -1 & -3 & 11 & 11 & -3 \end{pmatrix} \qquad B^{k+1}=\frac{1}{16}\begin{pmatrix} 13 & -5 & -1 & 1 & 0 & 0 & -1 & 1 & 5 & -13 \\ 5 & -13 & 13 & -5 & -1 & 1 & 0 & 0 & -1 & 1 \\ -1 & 1 & 5 & -13 & 13 & -5 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 5 & -13 & 13 & -5 & -1 & 1 \\ -1 & 1 & 0 & 0 & -1 & 1 & 5 & -13 & 13 & -5 \end{pmatrix}$$

3 Filtering time series with subdivision wavelet

In the section the time series of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006 are decomposed and reconstructed by subdivision wavelet presented in the above section. Some of source data are of the paper [4] and the other source data come from China Fishery Statistical Yearbooks. The dot graphs of the data of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006 are shown in fig.1. The graph of marine fishing production rises with wave until the top at the year 1999 or 2000 then goes down slowly. Meanwhile the graph of aquaculture production in marine waters rises slowly at first and fleetly in the final. The year 2006 the aquaculture production in marine waters of China has increased by about 100 times as compared with that of the year1954 which is 15.37 ten thousand tons. The two time series have not only the continuity but also some stochastic factors. So the trend and the disturbance should be separated from the data of marine fishing production and aquaculture production in marine waters.

According to the data, the trend and the disturbance of the two time series data are separated apart via the decomposition rule of subdivision wavelet. Then the reconstruction of the two time series is executed respectively by subdivision and subdivision wavelet. The concrete results are in fig.2-fig.4. Fig.2. shows the trend and the disturbance of the data of marine fishing production after acted by subdivision wavelet via formula (3) one time. Fig.3. gives the comparison of the initial data of marine fishing and its subdivision result via formula (2) with the initial data of marine fishing and its reconstruction result by subdivision wavelet via formula (4). The relative error between the subdivision result and the initial data is 0.0309. Meanwhile the relative error between the reconstruction result by subdivision wavelet and the initial data is 0.0293. So the subdivision wavelet is a finer tool for data smoothing. Corresponding results of the data of aquaculture production in marine waters of China are in fig.4. From fig.1 to fig.4, in the abscissa 1 represents the year 1954 and 52 represents the year 2006.

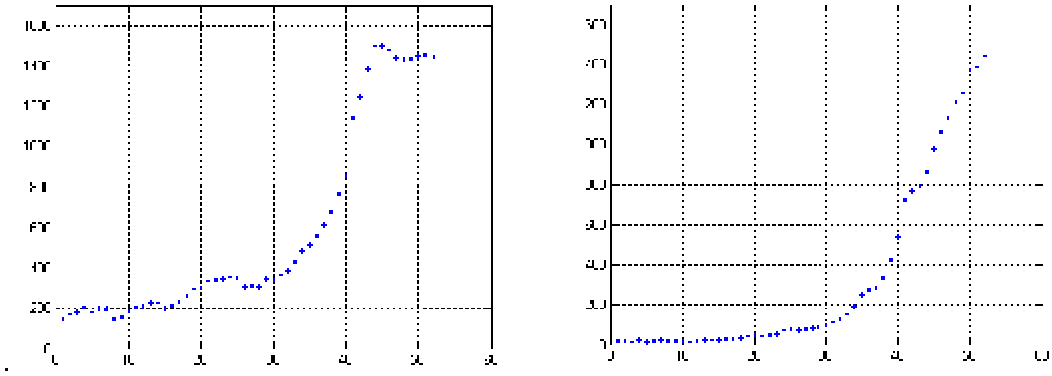

Fig.1. the dot graphs of marine fishing production (on the left) and aquaculture production in marine waters (on the right) of China from 1954 to 2006.

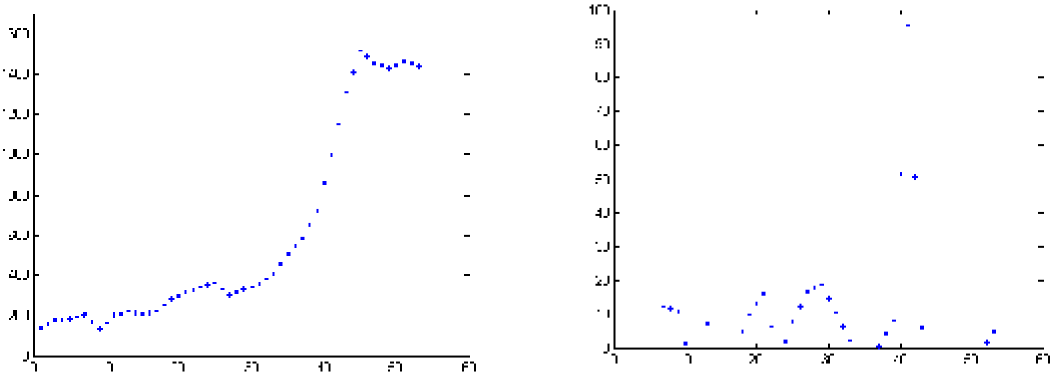

Fig.2. the trend (on the left) and the disturbance (on the right) of marine fishing production separated by subdivision wavelet.

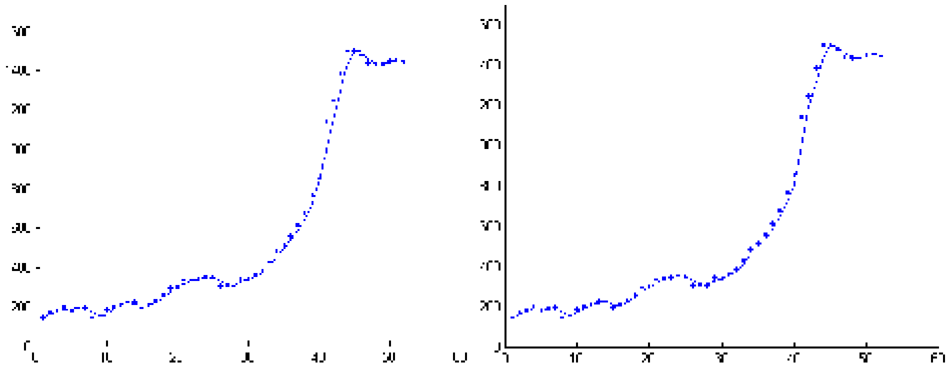

Fig.3. the initial data of marine fishing production and its subdivision result (on the left) and the initial data of marine fishing production and its reconstruction result by subdivision wavelet (on the right)

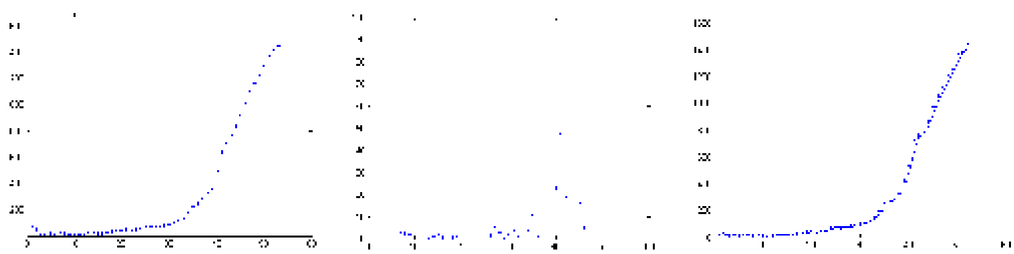

Fig.4. the trend (on the left), the disturbance (on the middle) of the data of aquaculture production in marine waters; the initial graph of aquaculture production in marine waters and its reconstruction result by subdivision wavelet (on the right)

4 Time series modeling

The reconstruction data of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006 by subdivision wavelet reveal the unitary properties so it is good for data multi-analysis. In the following course of disposal we use the reconstruction data of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006 as the basal data.

The reconstruction data of the marine fishing reveal the feature of auto-regressive moving average (ARMA) model [5] after analyzed by the software of Eviews.

ARMA(p.q) is presented by Box and Jenkins in 1970's and its general rule accords with

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \tag{6}$$

Where x_t represents the time series data, ϕ_i is of parameters of auto-regressive part, u_i represents the moving averages parts and θ_i is of the corresponding parameters of moving averages. The numbers p and q are the orders of auto-regressive part and the moving average part respectively.

To establish the ARMA model, testing the smoothness of the series according to the graph and testing unit root are necessary. The form and order of the model can be initially identified by the autocorrelation function and partial autocorrelation function figures. We use Akaike Information Criterion (AIC) and Schwartz criteria (SC) to determine the order of the model. That is, as far as possible select the one whose values of AIC and SC are the smaller. To eliminate autocorrelation the values of D.W. in the non-auto-correlation area must be selected. According to the autocorrelation function and partial autocorrelation function figures of the series, we can see that the Q-Stat value all pass Q-Test, and the form and order of the model can be initially identified. In order to establish a more appropriate model, we test a number of models. After comparing the data we select the AR (2) model. Using Eviews we get the following concrete model:

$$x_t = 1.8400 x_{t-1} - 0.9025 x_{t-2} + u_t$$

$R^2 = 0.9870$ $\textit{Adjusted } R^2 = 0.9867$; $AIC = 8.2990$, $SC = 8.3786$.

Fig.5. shows the comparison of the actual, fitted and residual values of the difference of the marine fishing production.

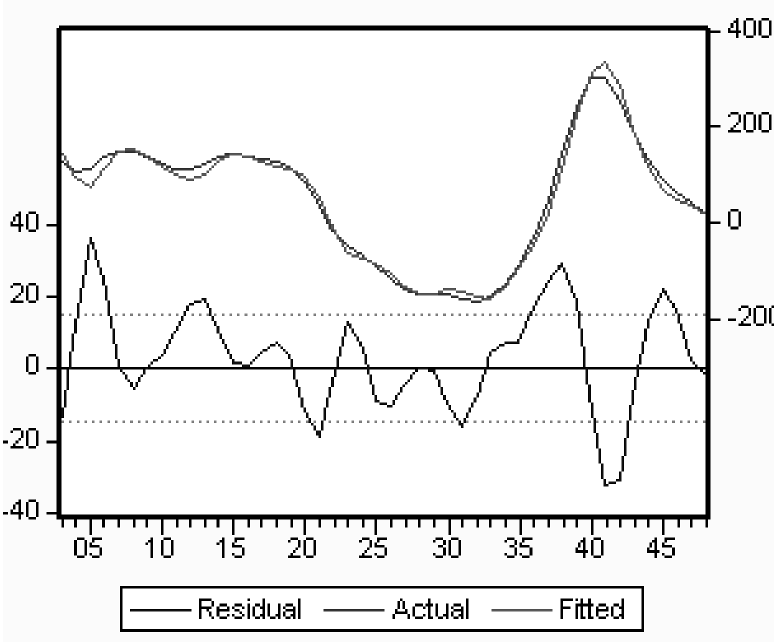

Fig.5.the comparison of the actual, fitted and residual values of the difference of the marine fishing production.

The reconstruction data of the aquaculture production in marine waters reveal the feature of Logistic regressive curve. Logistic curve model is given by P.F.Verhulst in 1845 and its effect is rediscovered by R.Pear and L.T.Reed in 1920's.The rule of Logistic growth model [6] is

$$y = \frac{L}{1 + ae^{-bt}}. \tag{7}$$

Where L represents the maximum of the curve, b represents the growth factor and a represents the reduction factor in some sense. The determination of the maximum L is the core in series modeling. According to the maximum of the real value and the rule of the minimum of the residual difference the range from 1700 to 1800 of L is determined. The calculation is executed by the soft of Matlab. The concrete formulae are $\frac{1700}{1+59.3469e^{-0.1201n}}$ when L equals1700 and $\frac{1800}{1+6.2077e^{-0.1633n}}$ when L equals 1800. The number n represents the ordinal number of the year and n equals one at the year 1954. When L equals 1800 the relative error between real data and the model data is 2.6672e-007.

The difference between the initial data and the fitted data by Logistic curve of the aquaculture production in marine waters reveals the feature of the white noise. Fig.6. shows the comparison of the reconstruction data by subdivision wavelet, the fitting curve by Logistic growth curve added with whit noise and the initial data of the aquaculture production in marine waters.

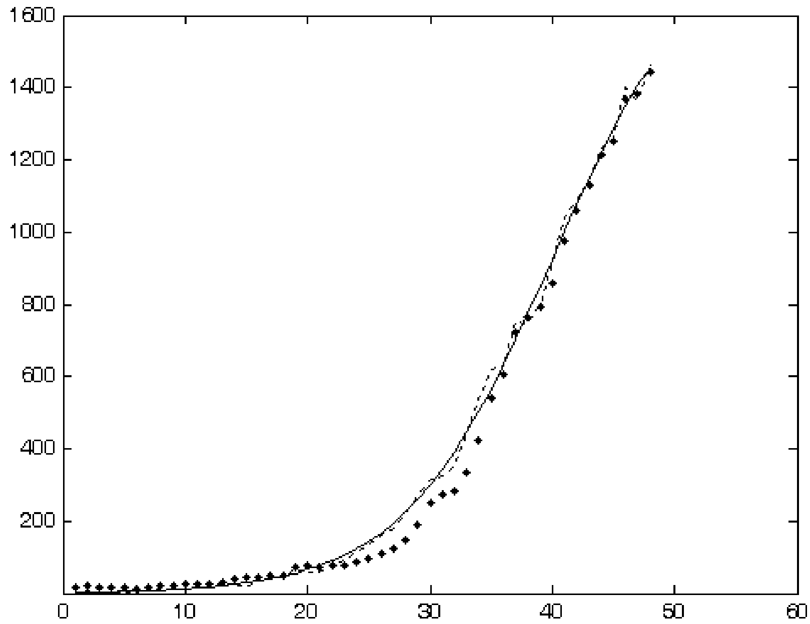

Fig.6. the comparison of the reconstruction data by subdivision wavelet (real line), the fitting curve by Logistic growth curve added with whit noise (dashed) and the initial data (quadrate dot line) of the aquaculture production in marine waters.

5 Conclusion

A new filter method came from subdivision wavelet is given in the paper which is used instead of HP filter when processing time series model. The data of marine fishing production and aquaculture production in marine waters of China from 1954 to 2006 are used for confirmation of the validity of the filters. The auto-regressive moving average (ARMA) model and the Logistic growth curve are used for modeling the data of marine fishing production and the aquaculture production in marine waters. More information about the advantage of approximating subdivision method and subdivision wavelet can be found in the paper [7] and [3]. The results of modelling demonstrate the validity of the methods presented in the paper. Further more, making the precision of data fitting progress is meaningful in the future.

References

[1] Hodrick R J,Prescott E C, Postwar U S: Business cycles: an empirical investigation, *Journal of Money*, *Credit and Banking*, **1**(1997), 1-16.

[2] Dyn N.: Subdivision scheme in computer aided geometric design, *Advances in Numerical, Clarendon Press*, **2**(1992), 36-104.

[3] Limei Zhnag, Zhongxuan Luo and Youxue Shi: A new subdivision wavelet and its application, *Numerical Mathematics A Journal of Chinese Universities,* **4**(2005), 309-316.

[4] The comparison table of production of marine waters for new-old criterion every

year, *China Fisheries,* (1998), 9-60.

[5] Xiaotong Zhang: Applied quantitative economics, China Machine Press, **3**(2009), 310-389.

[6] Lianfeng Yang: A Study on the Law of Chinese Internet-user Growth Based on Logistic Curve, *Journal of Xiamen University of Technology,* **4**(2008), 86-89.

[7] Limei Zhang, Zhongxuan Luo: An approximating C^3 stationary subdivision scheme, *Journal information & computational science,* **1**(2004), 157-161

Received June 14, 2010

Accepted November 25, 2010

Academic backgrounds:

09/1981 – 07/1985	Faculty of Mathematics in Jinzhou Normal College, Liaoning, China
09/1987 – 07/1990	Master's course of Graduate Department of Mathematics in Qufu Normal University, Shandong, China
09/2000 – 07/2004	Doctor's course of Graduate Department of Applied Mathematics Dalian University Of Technology, LIaoning, China

Research field: Mathematics Base in Fisheries Resource, Computer Aided Geometric Design.