Journal of Applied Statistics, Vol. 30, No. 5, 2003, 537–553

Routledge
Taylor & Francis Group

Modelling and forecasting by wavelets, and the application to exchange rates

H. WONG[1], **WAI-CHEUNG IP**[1], **ZHONGJIE XIE**[2] **& XUELI LUI**[2],
[1]*The Hong Kong Polytechnic University, Hung Hom, Hong Kong and* [2]*Department of Probability and Statistics, Peking University, People's Republic of China*

ABSTRACT *This paper investigates the modelling and forecasting method for non-stationary time series. Using wavelets, the authors propose a modelling procedure that decomposes the series as the sum of three separate components, namely trend, harmonic and irregular components. The estimates suggested in this paper are all consistent. This method has been used for the modelling of US dollar against DM exchange rate data, and ten steps ahead (2 weeks) forecasting are compared with several other methods. Under the Average Percentage of forecasting Error (APE) criterion, the wavelet approach is the best one. The results suggest that forecasting based on wavelets is a viable alternative to existing methods.*

1 Introduction

Forecasting by time series analysis approach has been developed over several decades (see Grenander & Rosenblatt, 1957; Box & Jenkins, 1970; Makridakis & Wheelwright, 1993). The modelling and forecasting methods for stationary time series have been applied to many different fields and many successful results have been obtained in diverse areas. Readers may find some interesting case studies in the book by Xie (1993).

As for the forecasting method of non-stationary time series, we also find many useful methods for applications. Particularly in the classic work of Box & Jenkins (1970), several models—for example, ARIMA and the seasonal ARIMA model—were introduced and have been widely used by statisticians. They are now implemented in many popular software packages, such as SAS, SPSS, S-Plus and so on.

Correspondence: H. Wong, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, People's Republic of China.

ISSN 0266-4763 print; 1360-0532 online/03/050537-17
DOI: 10.1080/0266476032000053664

However, evidently not all non-stationary series may be transformed to stationary series by the differencing procedure, so alternatives need to be considered. Makridakis *et al.* (1982) discussed several forecasting methods such as Holt–Winters, econometric models, etc. These methods also perform reasonably well. However, from a statistical point of view, they have a lack of theoretical basis and rigorous mathematical argument. On the other hand, some theories reported in the statistical literature are relatively complicated to use in practice, even though their large sample behaviour is impressive (see Li, 1990).

In recent years, wavelet theory developed very rapidly and has shown very strong applicability in diverse fields. The main advantages for such a mathematical tool may be comprehended in several aspects. The first is the well-known localization property and 'Zoom effect' of the wavelet functions. Many of these functions may have finite support in the time domain and also possess very good cancellation in the frequency domain (or vice versa). Moreover, we know that the wavelet transform sometimes modifies the non-stationary of the time series, namely, after wavelet transformations, many non-stationary processes become stationary processes (see Masry, 1993). Finally, the emergence of many algorithms (e.g. Fast Wavelet Transform—Meyer, 1993), make the wavelet transformation easier and easier to apply in practical applications, thus making the wavelet approach a viable alternative to existing methods.

What is the relationship between wavelet theory and modelling of time series? How does one use the advantages of wavelet tools in forecasting? These are very interesting and challenging problems to the time series analysts, and have drawn the attention of many researchers. Several reports have been published in this field and very attractive ideas are proposed by these authors (see Dahlhaus *et al.*, 1995; Tsatsan *et al.*, 1993; Li & Xie, 1999; Dijkerman & Mazumdar, 1994).

In this paper, w attempt to bring together the existing literature on statistical estimation (Brillinger, 1996) and the ongoing work on hidden periodicities analysis (Li & Xie, 1997) to demonstrate how wavelet theory can be usefully applied to the modelling and forecasting of time series. Brillinger (1996) suggested the statistical estimation of a deterministic regression function in the presence of noise. Unlike some papers published before (see Donoho *et al.*, 1995), Brillinger considered the case where the residual series of the model is weakly stationary and not restricted to the case of iid series. The estimates suggested in his paper are all consistent in the L^2 sense. In this paper, we develop his theory further to obtain the strong consistency of the estimates of the regression function.

Recently, Li & Xie (1997) improved the theory of hidden periodicities analysis by wavelets analysis, the model suggested in the paper is

$$H(t) = \sum_{k=1}^{p} \alpha_k e^{i\lambda_k t} + \xi(t), \quad t = 0, \pm 1, \pm 2, \ldots \tag{1}$$

where $\{\alpha_1, \ldots, \alpha_p\}$ are unknown random variables, $\{p, \lambda_1, \ldots, \lambda_p\}$ are all unknown parameters and may be estimated from a finite set of samples, $H(t_1), H(t_2), \ldots, H(t_N)$. Under some mild conditions, the estimates are all consistent in the almost surely sense.

Following the well-known X-11 modelling (see Cleveland & Tiao, 1976) we assume the observation $x(t), t \in Z$, may be decomposed into three parts,

$$x(t) = T(t) + H(t) + \varepsilon(t), \quad t \in Z \tag{2}$$

where $T(t)$ is the trend component, $H(t)$ the harmonic component and $\varepsilon(t)$ the irregular component. In the following section, we briefly introduce the theory and method for detecting the trend component $T(t)$ by wavelet approach. Theorem 1 shows the strong consistency of the estimate $\hat{T}_N(t)$. Theorem 2, in section 3, gives the algorithm for estimating the hidden periodicities by wavelets.

In section 4, we apply our method to the forecasting of exchange rates (US Dollar against DM) for two weeks ahead prediction, which was based on 512 observation samples for constructing the model. Some of the published literature (see Mills, 1993) concluded that the best model for exchange rate forecasting is the random walk model. The authors of this paper do not agree with this conclusion. It is apparent that a very strong limitation for the random walk model is that it can only give a one-step ahead prediction, and a longer forecasting horizon is not entertained. On the other hand, in the final part of this paper, several kinds of forecasting error are compared; readers may find that even in the forecasting error respect, the random walk model is not the best one.

2 Trend component estimation by wavelets

2.1 Basic model

Suppose that the samples $y(t)$, $t = 0, \pm 1, \ldots, \pm T$, follow the model

$$y(t) = T(t) + \eta(t) \tag{3}$$

$$T(t) = f\left(\frac{t}{T}\right), \quad t = 0, \pm 1, \ldots, \pm T \tag{4}$$

where $f(t)$ is a real function, $f(t) \equiv 0$, $t \notin [-1, 1]$, and $\eta(t)$ is stationary noise.

Let $\{V_j\}_{j \in Z}$ be an MRA (Multi-resolution Analysis) of $L^2(R)$, $\{\phi(t-k),\ k \in Z\}$ forms an orthonormal basis in V_0, where $\phi(t)$ is the scale function. Let $\psi(t)$ be the mother wavelet determined by $\phi(t)$ and for any given integer l, $\{\phi_{l,k}(t),\ \psi_{j,k}(t), j \geqslant l, k \in Z\}$ provides a complete orthonormal basis in $L^2(R)$.

Put

$$\hat{\alpha}_{l,k} = \frac{1}{T} \sum_{t=0}^{T} \phi_{l,k}\left(\frac{t}{T}\right) y(t) \tag{5}$$

$$\hat{\beta}_{j,k} = \frac{1}{T} \sum_{t=0}^{T} \psi_{j,k}\left(\frac{t}{T}\right) y(t) \tag{6}$$

Then we have the following empirical wavelet estimation for the function $f(t)$

$$\hat{f}_T(t) = \sum_k \hat{\alpha}_{l,k} \phi_{l,k}(t) + \sum_{j \geqslant l} \sum_k \hat{\beta}_{j,k} \psi_{j,k}(t) \tag{7}$$

or

$$\hat{f}_T(t) = \sum_j \sum_k \hat{\beta}_{j,k} \psi_{j,k}(t) \tag{8}$$

In practical applications, it can be assumed that there exists an integer $\mathcal{J}$ (sufficiently large), such that $f(t) \in V_{\mathcal{J}}$. In such a case, equations (7) and (8) may be written as

$$\hat{f}_T^{(t)} = \sum_k \hat{\alpha}_{\mathcal{J},k} \phi_{\mathcal{J},k}(t) + \sum_{j=1}^{\mathcal{J}-1} \sum_k \hat{\beta}_{j,k} \psi_{j,k}(t) \tag{9}$$

for $1 < \mathcal{J}$, or

$$\hat{f}_T^{(t)} = \sum_{j=-\infty}^{\mathcal{J}-1} \sum_k \hat{\beta}_{j,k} \psi_{j,k}(t) \tag{10}$$

$$= \sum_k \hat{\alpha}_{\mathcal{J},k} \phi_{\mathcal{J},k}(t) \tag{11}$$

In order to prove Theorem 1, several mathematical assumptions are necessary.

2.2 Statistical properties of the estimates

Assumption 1

$f(t)$ is a real function with bounded variation on $[-1, 1]$ and vanishes outside the interval.

Assumption 2

The scale function $\phi(t)$ satisfies

(a) $\phi(t)$ has finite support on $[0, 2N-1]$, $N > 0$. (12)

(b) $\phi(t)$ has bounded variation on $[0, 2N-1]$.

(c) $\phi(t) \in S_r = \{g(x) : |g^{(k)}(x)| \leqslant c_{kp}/(1+|x|)^p,\ k, p = 0, 1, \ldots, r\}$ (13)

Assumption 3

The cumulant functions of the zero mean stationary series $\eta(t)$, $t \in Z$, satisfy

$$K_m = \sum_{u_1} \sum_{u_2} \ldots \sum_{u_{m-1}} |C_m(u_1, u_2, \ldots, u_{m-1})| < +\infty, m \geqslant 2 \tag{14}$$

also

$$\sum_u |u| \cdot |C_{\eta\eta}(u)| < +\infty \tag{15}$$

and

$$f_{\eta\eta}(0) > 0 \tag{16}$$

where

$$C_m(u_1, u_2, \ldots, u_{m-1}) = Cum\{\eta(t+u_1), \eta(t+u_2), \ldots, \eta(t+u_{m-1}), \eta(t)\} \tag{17}$$

and $f_{\eta\eta}(\lambda)$ is the spectral density function of $\eta(t)$.

Under Assumptions 1 and 3 and parts (a) and (b) of Assumption 2, Brillinger (1996) proved the L^2-consistency of the $\{\hat{\alpha}_{l,k}\}$, $\{\hat{\beta}_{j,k}\}$ and $\hat{f}_T(t)$.

Now, we propose Theorem 1, which shows that, if the scale function $\phi(t) \in S_r$, then $\hat{f}_T(t)$ converges to $f(t)$ in an almost sure sense.

Theorem 1

Suppose that $y(t)$ follows model (3), and $\{y(t),\ t=0, \pm 1, \ldots, \pm T\}$ are sample points. Under Assumptions 1 and 2, when $T \to \infty$,

$$\hat{f}_T(t) \to f(t) \text{ a.s.} \tag{18}$$

where $\hat{f}_T(t)$ is defined in equations (7) and (8).

The proof of Theorem 1 may be found in the Appendix.

2.3 *Time varying filtering by wavelet functions*

In practical applications, equations (7)–(9) are not the best formulas for calculation. The following derivation leads us to establish a relationship between the wavelet estimation of $f(t)$ and the time varying filtration.

First, we replace the empirical coefficients of wavelet transformation (5) into (9), then

$$\begin{aligned} \hat{f}(t) &= \sum_k \hat{\alpha}_{J,k} \phi_{J,k}(t) \\ &= \sum_k \left(\frac{1}{T} \sum_{u=-T}^{T} \phi_{J,k}\left(\frac{u}{T}\right) y(u) \right) \phi_{J,k}(t) \\ &= \frac{2^J}{T} \sum_{u=-T}^{T} y(u) \left(\sum_k \phi(2^J t - k) \phi\left(2^J \frac{u}{T} - k\right) \right) \end{aligned} \tag{19}$$

where

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k),\ \forall j,\ k \in Z \tag{20}$$

Put

$$g(t, u) = \frac{2^J}{T} \sum_k \phi(2^J t - k) \phi\left(2^J \frac{u}{T} - k\right) \tag{21}$$

then equation (19) may be rewritten as

$$\hat{f}(t) = \sum_{u=-T}^{T} g(t, u) y(u) \tag{22}$$

Equation (22) shows that the estimation of $f(t)$ is essentially the output of a time varying filter, the weighted coefficients; or, in engineering terminology, the impulse response function of the filter is defined by the scale function.

Accordingly, the estimation of the trend component $T(t)$ in equation (4) may be represented as

$$\hat{T}(t) = \frac{1}{T} \sum_{u=-T}^{T} w(t, u) y(u) = \hat{f}\left(\frac{t}{T}\right) \tag{23}$$

where

$$w(t,u)=\frac{2^J}{T}\sum_k \phi\left(2^J\frac{t}{T}-k\right)\phi\left(2^J\frac{u}{T}-k\right) \tag{24}$$

Suppose that $2^J < T$ for sufficiently large T, then put

$$D = T/2^J \tag{25}$$

where D is a positive integer

$$\theta(l)=\phi(l/D) \tag{26}$$

then

$$\begin{aligned} w(t,u) &= \frac{1}{D}\sum_k \phi\left(\frac{1}{D}(t-Dk)\right)\phi\left(\frac{1}{D}(u-Dk)\right) \\ &= \frac{1}{D}\sum_k \theta(t-Dk)\theta(u-Dk) \\ &= \frac{1}{D}\sum_l \theta(l)\theta(l+(u-t)) \quad (\text{put } t-Dk=1) \\ &= \frac{1}{D}\sum_l \theta(l)\theta(l-(t-u)) \triangleq w(t-u) \end{aligned} \tag{27}$$

Hence, equation (23) is

$$\hat{T}(t)=\hat{f}\left(\frac{t}{T}\right)=\sum_{u=-T}^{T} w(t-u)y(u) \tag{28}$$

which shows that the trend component may be estimated by a filter, and the impulse response function of the filter is

$$w(x)=\frac{1}{D}\sum_l \theta(l)\theta(l-x) \tag{29}$$

3 Hidden periodicities analysis by wavelets

Hidden periodicities analysis has been investigated by many researchers for more than five decades (see Priestly, 1981). The model is assumed to be

$$y(t)=\sum_{l=1}^{q}\alpha_l\cos(\lambda_l t+\theta_l)+\xi(t), \quad t\in Z \tag{30}$$

where the first part of equation (30) is called the harmonic component, $\xi(t)$ the noise, $\{\theta_l\}$ are random phases and usually assumed to be iid uniformly distributed on $[-t, t]$. Starting from finite observation samples $\{y(1), y(2), \ldots, y(N)\}$, how to estimate the unknown parameters q, $\{\lambda_l, \alpha_l\}_1^q$ is the main problem in hidden periodicities analysis. In the usual case, researchers have assumed that q is known a priori, and $\xi(t)$ is an iid series or Gaussian random variables (see Priestley, 1981). However it is worth noting that, in many practical situations, the number of the harmonic component q is usually unknown and $\xi(t)$ deviates from a Gaussian or iid series.

In the following, we assume

$$y(t) = \sum_{l=1}^{q} \alpha_l e^{i\lambda_l t} + \xi(t) \tag{31}$$

where

$$-\pi < \lambda_1 < \lambda_2 < \ldots < \lambda_q < \pi \tag{32}$$

$$0 < \lambda \leqslant |\lambda_m - \lambda_n|, \ 1 \leqslant m, n \leqslant q, n \neq m \tag{33}$$

q is an unknown integer, $\{\alpha_l\}$ are complex random variables, satisfying

(1) $\text{Var}(\alpha_l) < +\infty, \ l = 1, 2, \ldots, q$ (34)

(2) $\{\alpha_l\}$ are uncorrelated random variables

$$P\{\alpha_l \neq 0\} = 1, \ l = 1, 2, \ldots, q \tag{35}$$

(3) $0 < \alpha \leqslant |\alpha_l|^2, \ l = 1, 2, \ldots, q$ (36)

where α is a known number.

The noise $\xi(t)$ in equation (31) satisfies the following conditions.

(1) $\xi(t)$ is a linear process

$$\xi(t) = \sum_{l=0}^{\infty} d_l \varepsilon(t - l) \tag{37}$$

where $\{d_l\}$ satisfies

$$\sum_{l=0}^{\infty} \sqrt{l}\,|d_l| < +\infty \tag{38}$$

(2) $\varepsilon(t)$ is an ergodic stationary series with

$$E\varepsilon(t) = 0, \ \text{Var}(\varepsilon(t)) = \sigma^2 < +\infty \tag{39}$$

and $\varepsilon(t)$ is uncorrelated with $\{\alpha_l\}$.

Let $\psi(t)$ be a mother wavelet, its Fourier transform $\hat{\psi}(\omega)$ is compactly supported on a finite interval $[-M, M]$ and satisfies

$$\int_{-M}^{M} \hat{\psi}(\omega) d\omega \neq 0; \ \int_{-M}^{M} |\hat{\psi}(\omega)| \, d\omega < +\infty \tag{40}$$

Define the periodogram $I_T(\lambda)$ and its wavelet transform as

$$I_T(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} y(t) e^{-i\lambda t} \right|^2, \quad -\pi \leqslant \lambda \leqslant \pi \tag{41}$$

$$\bar{W}_{j,k} = 2^{-j/2} \int_{-\pi}^{\pi} \psi_{j,k}^{\text{per}}(\lambda) I_T(\lambda) \mathrm{d}\lambda \tag{42}$$

where

$$\psi_{j,k}^{\text{per}}(x) = \sum_{n} (2\pi)^{-1/2} \psi_{j,k}\left(\frac{x+\pi}{2\pi} + n\right) \tag{43}$$

Then, we have the following theorem.

Theorem 2

Suppose that the observation model is the same as equation (31) and the conditions (32)–(40) are fulfilled, the sample size T and the resolution level j satisfy

$$\frac{\lim\limits_{\substack{j\to\infty \\ T\to\infty}} 2^j}{T} = 0 \quad \frac{\lim\limits_{\substack{j\to\infty \\ T\to\infty}} 2^j}{\sqrt{T \log T}} = +\infty \tag{44}$$

Then, for sufficiently large j the following results hold:

(1) For all $k \in I(\lambda_1, 2^{-2j})$

$$\bar{W}_{j,k} = \frac{|\alpha_1|^2}{2\pi} \int_{-M}^{M} \hat{\psi}(\omega) \mathrm{d}\omega + o(1), \quad \text{for } k = \{0, 1, 2, \ldots, 2^j - 1\} \tag{45}$$

where

$$I(\lambda_1, 2^{-2j}) = \left\{ k : \left| \frac{k}{2^j} 2\pi - \pi - \lambda_1 \right| < 2^{-2j} \quad \text{or} \right. \tag{46}$$

$$\left. \left| \frac{k}{2^j} 2\pi - \pi - \lambda_1 \right| > 2\pi - 2^{-2j},\, k \in \{0, 1, \ldots, 2^j - 1\} \right\}$$

(2) For all $k \in A_T$,

$$\bar{W}_{j,k} = o(1), \tag{47}$$

where

$$A_T = \bigcap_{l=1}^{q} \left\{ k : 2^{-j/2} \leqslant \left| \frac{k}{2^j} 2\pi - \pi - \lambda_1 \right| \leqslant 2\pi - 2^{-j/2},\, k \in \{0, 1, \ldots, 2^j - 1\} \right\} \tag{48}$$

The proof of Theorem 2 may be found in Li & Xie (1997).

This theorem provides a procedure to detect the hidden periodicities.

Step 1. Let $\mathcal{J} = \{0, 1, 2, \ldots, 2^j - 1\}$, for an appropriate integer j_0, calculate

$$\{\bar{W}_{j,k},\, k \in \mathcal{J}\},\, j = j_0, j_0 + 1, \ldots \tag{49}$$

Step 2. Let

$$k(j) = \arg\left\{\max_{k\in\mathcal{J}} (\bar{W}_{j,k})\right\} \tag{50}$$

$$\mathrm{MW}(j) = \max_{k\in\mathcal{J}} (\bar{W}_{j,k}) \tag{51}$$

(a) If $\mathrm{MW}(j) \sim$ Constant for $j = j_0, j_0+1, j_0+2, \ldots$, then put

$$\hat{\lambda}_1 = 2\pi\left(\frac{k(j)}{2^j} - \frac{1}{2}\right), \text{ for sufficiently large } j \tag{52}$$

$\hat{\lambda}_1$ is an estimate of a hidden frequency.

(b) If $\mathrm{MW}(j) \to 0$, for $j = j_0, j_0+1, j_0+2, \ldots$, then there is no hidden frequency in the observation data.

Step 3. If $\hat{\lambda}_1$ is an estimate of hidden frequency, put

$$\hat{\alpha}_1 = \frac{1}{2\pi T}\sum_{t=1}^{T} y(t)\mathrm{e}^{-i\hat{\lambda}_1 t} \tag{53}$$

$$y_1(t) = y(t) - \hat{\alpha}_1 \mathrm{e}^{it\hat{\lambda}_1},\ t = 1, 2, \ldots, T \tag{54}$$

then repeat the calculation from steps 1 to 3.

Step 4. When $\mathrm{MW}(j) \to 0$, for $j = j_0, j_0+1, \ldots$, then stop.
Suppose $\{\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_{\hat{q}}\}$ have been detected, then

$$\hat{q} = q;\ \hat{\lambda}_l = \lambda_l,\ l = 1, 2, \ldots, q, \text{ a.s.} \tag{55}$$

Under mild mathematical conditions, He (1987) proved that $\{\hat{\alpha}_j\}$ defined in equation (53) are strongly consistent estimates of the amplitudes of model (30).

For example,

$$q = 2,\ y(t) = 2.5(\mathrm{e}^{-i0.3t} + \mathrm{e}^{i0.3t}) + \varepsilon(t),\ t = 1, 2, \ldots, 512 \tag{56}$$

where $\varepsilon(t) \sim N(0, 0.2^2)$, then we have the following results.

j	5	6	7	8	9
$\mathrm{MW}(j)$	0.84	0.86	0.85	0.72	0.73
$k(j)$	8, 24	14, 50	27, 50	53, 103	104, 408
$\hat{f}(j)$	$\{-0.25, 0.25\}$	$\{-0.28, 0.28\}$	$\{-0.29, 0.29\}$	$\{-0.29, 0.29\}$	$\{-0.3, 0.3\}$
$\mathrm{av}(A_T)$	0.27	0.27	0.05	0.01	0.004

From the results, we see that when j increases, the maximum values of $\bar{W}_{j,k}$ have very similar magnitudes, and outside the neighbourhood of the frequency, the average value, denoted as $\mathrm{av}(A_T)$ decreases rapidly. Accordingly, we may be assured that there are at least two hidden frequencies

$$\begin{cases} \hat{f} = \{-0.3, 0.3\} \\ \hat{\alpha} = 2.505 \end{cases} \tag{57}$$

Evidently, such estimates of the parameters in equation (56) are good enough for applications. For $q > 2$, readers may refer to the simulation results in Li & Zie (1997). In summary, the above sections show that we can model the trend and harmonic component of a time series by wavelets.

4 Modelling and forecasting for exchange rate data

4.1 Modelling of exchange rate data

Figure 1 is the US dollar against DM exchange rate data from 1 August 1989 to 31 July 1991.

It is apparent that the exchange rate series is not a stationary series (see Mills, 1993). We may consider that the process can be decomposed as

$$x(t) \equiv T(t) + \eta(t), \; t \in Z \tag{58}$$

where $T(t)$ is the trend component, $\eta(t)$ is a stationary series.

Following the statistical detection method introduced in Section 2, it is necessary to find a suitable wavelet $\psi(t)$ and scale function $\phi(t)$ which satisfy the conditions (a), (b) and (c) of assumption 2 (see equation (13)). We introduce the scale function $\phi(t)$ suggested by Daubechies (1992). $\phi(t)$ possesses finite support on $[0, 2N-1]$ in the time domain, where we select $N = T$. Figure 2 illustrates the scale function $\phi(t)$.

For the trend component $T(t)$, we use equation (28), where

$$w(x) = \frac{1}{D} \sum_{l} \theta(l)\theta(l-x) \tag{59}$$

$$D = \frac{T}{2^{\mathcal{J}}} \tag{60}$$

$$T = 512, \mathcal{J} = 6$$

The trend component detected by our wavelet method is shown in Fig. 3. The comparison of the trend component against the original observation is illustrated in Fig. 4.

From Fig. 4, we see that the trend component detected by our method is quite satisfactory in following the variation of the exchange rate data.

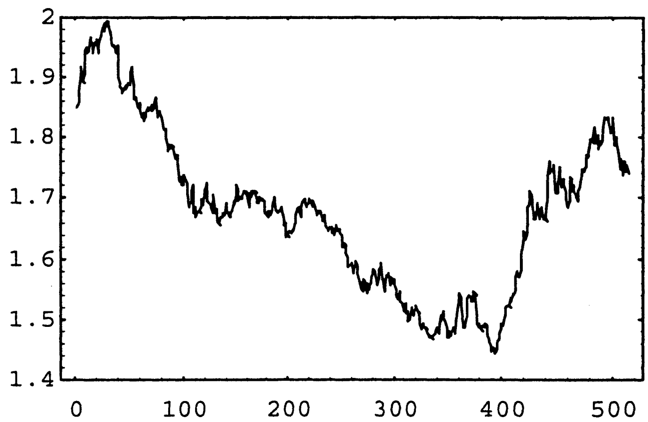

FIG. 1. US dollar versus DM exchange rate. (1 August 1989 to 31 July 1991)

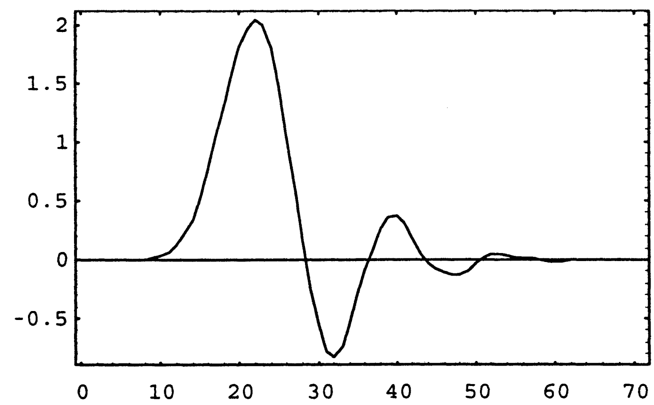

FIG. 2. Daubechies scale function $N = 7$.

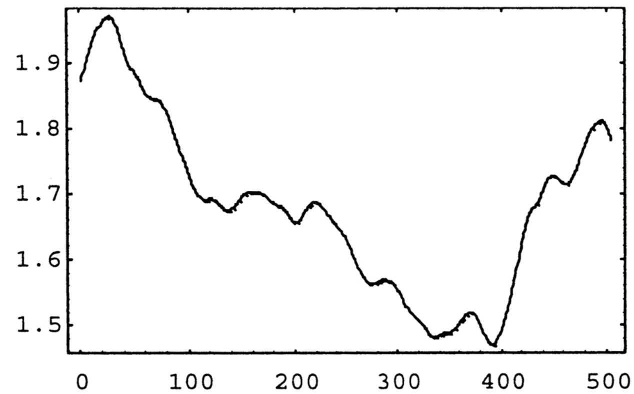

FIG. 3. Trend component of the exchange rate data of Fig. 1.

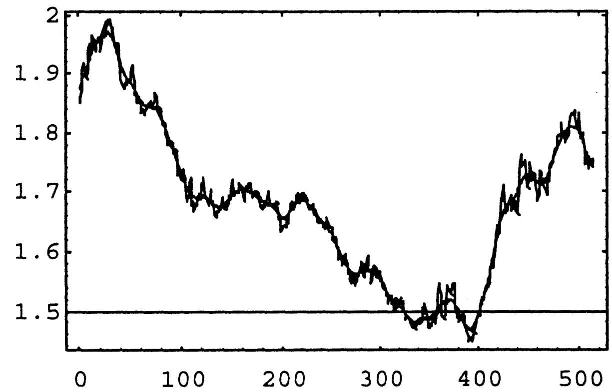

FIG. 4. The trend curve and the original data of exchange rate.

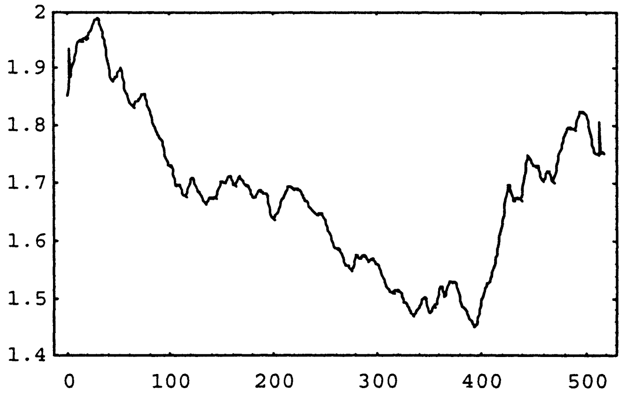

FIG. 5. Trend component detected by X-11.

Figure 5 shows the trend component detected by the well-known X-11 method (see Cleveland & Tiao, 1976).

Comparing Fig. 3 with Fig. 5, we see that the trend component from the wavelet approach is smoother than that from the X-11 method, particularly at the two end-points. Thus, it is reasonable to conclude that, in detecting the trend component of the exchange rate data, the wavelet method is better than that of the X-11 method.

Now consider the stochastic component $\eta(t)$ of model (58). We want to know whether the harmonic component $H(t)$ is involved in $\eta(t)$. Hence, we use the algorithm introduced in Section 3 for detecting the $H(t)$.

For convenience in calculating the wavelet coefficients of equation (42) we propose the following approach.

Suppose that $\{x_1, x_2, \ldots, x_T\}$ are real observation samples of equation (30), γ_k denotes the sample auto-covariance

$$\gamma_k = \frac{1}{T} \sum_{l=1}^{T-|k|} x_l x_{l+|k|}, \ k = 0, \pm 1, \ldots, \pm (T-1) \tag{61}$$

then the periodogram $I_T(\lambda)$ (see equation (42)) may be represented as

$$I_T(\lambda) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} \gamma_k e^{-ik\lambda} \tag{62}$$

(see Priestly, 1981). Hence the wavelet coefficient of the periodogram (see equation (42))

$$\begin{aligned} 2^{j/2} \bar{W}_{j,k} &= \int_{-\pi}^{\pi} \psi_{j,k}^{\text{per}}(\lambda) I_T(\lambda) \mathrm{d}\lambda \\ &= (2\pi)^{-1/2} \sum_n 2^{j/2} \int_{-\pi}^{\pi} \psi\left(2^j\left(\frac{\lambda+\pi}{2\pi} + n\right) - k\right) I_T(\lambda) \mathrm{d}\lambda \qquad (63) \\ &= (2\pi)^{-1/2} \sum_n 2^{j/2} \left(\frac{1}{2\pi} \sum_{s=-(T-1)}^{T-1} \gamma_s \int_{-\pi}^{\pi} \psi\left(2^j\left(\frac{\lambda+\pi}{2\pi} + n\right) - k\right) e^{-is\lambda} \mathrm{d}\lambda \right. \end{aligned}$$

$$\bar{W}_{j,k} = 2^{-j} \sum_{s=-(T-1)}^{T-1} (-1)^s \gamma_s \times \hat{\psi}\left(2\pi \frac{s}{2^j}\right) \mathrm{e}^{-i(2\pi k 2^{-j})s}, \quad k = \{0, 1, 2, \ldots, 2^j - 1\} \tag{64}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of the wavelet $\psi(t)$.

In the following, we select the Marr wavelet

$$\psi(t) = \mathrm{e}^{-t^2/2} - \frac{1}{2}\mathrm{e}^{-t^2/8} \tag{65}$$

so that

$$\hat{\psi}(\omega) = (\mathrm{e}^{-\omega^2/2} - \mathrm{e}^{-2\omega^2})\sqrt{\pi} \tag{66}$$

$$\bar{W}_{j,k} = 2 \times 2^{-j} \sum_{s=1}^{T-1} \gamma_s (-1)^s \hat{\psi}\left(2\pi \frac{s}{2^j}\right) \cos(2\pi k \times 2^{-j} s) \tag{67}$$

since $\hat{\psi}(0) = 0$, $\gamma_s = \gamma_{-s}$.

Then for the residual data, i.e. data obtained after the removal of the trend component, we have the following results:

j	5	6	7	8	9	10	12
MW(j)	3.57×10^{-6}	1.41×10^{-6}	5.9×10^{-7}	4.9×10^{-7}	2.9×10^{-7}	1.96×10^{-7}	4.8×10^{-8}

Since the max($W_{j,k}$) decreases to zero when $j \to +\infty$, we conclude that there is no evident harmonic component in the residual.

Then, by Maximum Entropy (ME) spectral estimation (see Burg, 1972) we also obtain a similar conclusion (see Fig. 6), i.e. no evident hidden frequency exists in the residual. Here, we want to emphasize that, after removing the trend by Wavelet Filtration from the exchange rate data, there is no harmonic component in the residual. However, this does not mean that if we detect the trend by other methods, the same conclusion is true.

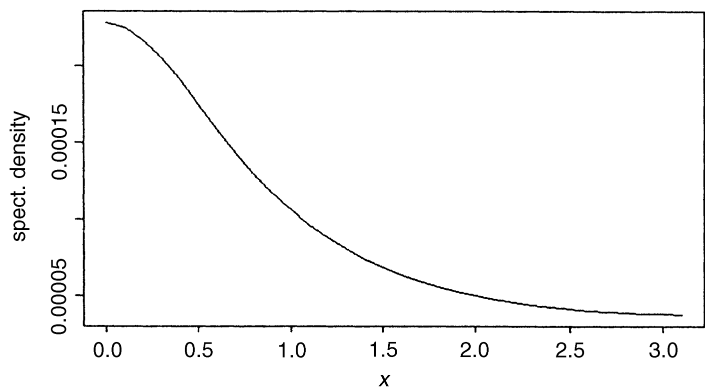

FIG. 6. Spectral density estimated by ME + BIC.

TABLE 1. Different methods of predictions

Step	Real	Wavelet	State-space	Box–Jenkins	X-11
1	1.7729	1.7824	1.7933	1.7916	1.769
2	1.7465	1.7743	1.7910	1.7915	1.7811
3	1.7636	1.7666	1.7896	1.7914	1.7919
4	1.7576	1.7590	1.7905	1.7912	1.7683
5	1.7382	1.7511	1.7919	1.7911	1.7613
6	1.7594	1.7432	1.7910	1.7910	1.7641
7	1.7433	1.7351	1.7899	1.7909	1.7786
8	1.7453	1.7266	1.7899	1.7908	1.7910
9	1.7534	1.7179	1.7908	1.7907	1.7829
10	1.7422	1.7089	1.7905	1.7906	1.7844
APE		9.5×10^{-3}	2.2×10^{-2}	2.2×10^{-2}	1.48×10^{-2}

5 Extrapolation

In the general case, when $T(t)$ and $H(t)$ have been detected, then

$$\xi(t) = x(t) - T(t) - H(t) \tag{68}$$

which may be considered as a stationary series. Under the ME criterion and order selection by BIC (see Xie, 1993), we have the following results.

Order p	0	1	2	3	4	5
BIC	-8.8416	-10.3912	-10.1021	-9.9212	-9.8491	-9.7940

Thus, we fitted an AR(1) model for the residual $\xi(t)$:

$$\xi(t) - (0.6311)\xi(t-1) = \varepsilon(t) \tag{69}$$

where $\{\varepsilon(t)\}$ is a white noise series,

$$E\varepsilon(t) = 0, \operatorname{Var}(\varepsilon(t)) = 5.477 \times 10^{-3} \tag{70}$$

In order to make forward forecasting, it is necessary to fit a function for the final part of the trend $\hat{T}(t)$. Under the MSE criterion we have

$$\hat{T}(t) = 1.80342 + 0.00304576t - 3.33 \times 10^{-4}t^2 + 2.403 \times 10^{-6}t^3 \tag{71}$$

Table 1 shows several ten steps (2 weeks) ahead predictions of US dollar against the DM by different methods (see Makridakis *et al.*, 1982; Xie, 1993), where the average percentage of predicted error (APE) is defined as

$$APE = \frac{1}{10}\sum_{k=1}^{10}\frac{|\text{Forecast}(k) - \text{Real}(k)|}{\text{Real}(k)} \tag{72}$$

It will also be interesting to compare the various methods on their one-step-ahead prediction errors. In addition to the methods in Table 1, we add the random walk model and have the following results.

	Wavelet	State-space	Box–Jenkins	X-11	Random walk
APE	5.35×10^{-3}	1.15×10^{-2}	1.05×10^{-2}	2.25×10^{-3}	1.06×10^{-2}

6 Conclusions

In this paper we have proposed a modelling and forecasting procedure for non-stationary series by the wavelet approach. The numerical procedure is fairly easy to apply for practical data. All the proposed estimators in this paper for detecting trends and hidden frequencies are strongly consistent. The ten steps ahead predictions for the exchange rate data show that the APE of the wavelet approach is the minimum compared with a number of well-established methods.

Acknowledgement

This research work was supported by the Hong Kong Research Grants Council and partly supported by the NSFC, No. 1017/005 of China.

REFERENCES

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

Brillinger, D. R. (1996) Some uses of cumulants in wavelet analysis, *Nonparametric Statistics*, 6, pp. 93–114.

Burg, J. P. (1972) The relationship between maximum entropy spectra and maximum likelihood spectra, *Geophysics*, 37, pp. 375–376.

Cleveland, W. S. & Tiao, G. (1976) Decomposition of seasonal time series: a model for the Census *X*-11 program, *Journal of the American Statistical Association*, 71, pp. 581–587.

Dahlhaus, R., Neumann, M. H. & Von Sachs, R. (1995) Nonlinear wavelet estimation of time-varying autoregressive process. Preprint No. 159, Weierstrass—Institute fur Angewandte Analysis and Stochastic, ISSN 0946-8633, Berlin.

Daubechies, I. (1992) *Ten Lectures on Wavelets* (Philadelphia, SIAM).

Dijkerman, R. W. & Mazumdar, R. R. (1994) Wavelet representations of stochastic processes and multiresolution stochastic model, *IEEE Transactions on Signal Processing*, 42(7), pp. 1640–1652.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995) Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society*, B, 57, pp. 301–369.

Grenander, U. & Rosenblatt, M. (1957) *Statistical Analysis of Stationary Time Series* (New York, Wiley).

He, S. (1987) On estimating the hidden periodicities in linear models, *Acta Mathematicae Applicatae Sinica*, 3, pp. 168–179.

Li, G. (1990) Estimation of parameters in ARIMA models, *Acta Mathematicae Applicatae Sinica*, 6(1), pp. 23–34.

Li, Y. & Xie, Z. (1997) The wavelet detection of hidden periodicities in time series, *Statistics and Probability Letters*, 35, pp. 9–23.

Li, Y. & Xie, Z. (1999) The wavelet identification of thresholds and time delay of threshold autoregressive models, *Statistica Sinica*, 9, pp. 153–166.

Makridakis, S., Andersen, A., Carbrone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. & Winkler, R. (1982) The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting*, 1, pp. 111–153.

Makridakis, S. & Wheelwright, S. C. (1993) *Forecasting: Methods and Applications*, 2nd edn (Canada, Wiley).

Masry, E. (1993) The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion, *IEEE Transactions on Information Theory*, 39, pp. 260–264.

Meyer, Y. (1993) *Wavelet Algorithms and Applications* (Philadelphia, SIAM).

Mills, T. C. (1993) *The Economic Modelling of Financial Time Series* (Cambridge University Press).

Priestley, M. B. (1981) *Spectral Analysis of Time Series* (Academic Press).

Tsatsan, M. K. & Giannakis, G. B. (1993) Time-varying system identification and model validation using wavelets, *IEEE Signal processing*, 41(12), pp. 3512–3523.

Walter, G. G. (1994) *Wavelets and Other Orthogonal Systems with Applications* (Boca Raton, Florida, CRC Press).

Xie, Z. (1993) *Case Studies in Time Series Analysis* (Singapore, World Scientific).

Appendix

Proof of Theorem 1

Under the Assumptions 1, 3 and (a), (b) of Assumption 2, Brillinger (1996) proved the following results:

(1) $E(\hat{\alpha}_{l,k} - \alpha_{l,k}) = O(2^{1/2} \times T^{-1})$

$E(\hat{\beta}_{l,k} - \beta_{l,k}) = O(2^{1/2} \times T^{-1})$

where

$$\alpha_{l,k} = \int_R \phi_{l,k}(t) f(t) \mathrm{d}t,\ l,\ k \in Z \tag{A1}$$

$$\beta_{j,k} = \int_R \psi_{l,k}(t) f(t) \mathrm{d}t,\ j,\ k \in Z \tag{A2}$$

are the wavelet coefficients of $f(t)$, $\hat{\alpha}_{l,k}$, $\hat{\beta}_{l,k}$ are empirical wavelet coefficients defined in equations (5) and (6).

(2)

$$\mathrm{Cov}(\hat{\alpha}_{l,k}, \hat{\alpha}_{l,k'}) = 2\pi f_{\eta\eta}(0) \frac{1}{T} \int_{-1}^{1} \phi_{l,k}(t)\phi_{l,k'}(t)\mathrm{d}t + O(2^{l} \times T^{-2}) \tag{A3}$$

$$\mathrm{Cov}(\hat{\alpha}_{l,k}, \hat{\beta}_{j,k'}) = 2\pi f_{\eta\eta}(0) \frac{1}{T} \int_{-1}^{1} \phi_{l,k}(t)\psi_{j,k'}(t)\mathrm{d}t + O(2^{(l+j)/2} \times T^{-2}) \tag{A4}$$

$$\mathrm{Cov}(\hat{\beta}_{j,k}, \hat{\beta}_{j',k'}) = 2\pi f_{\eta\eta}(0) \frac{1}{T} \int_{-1}^{1} \psi_{j,k}(t)\psi_{j',k'}(t)\mathrm{d}t + O(2^{(j+j')/2} \times T^{-2}) \tag{A5}$$

Based on these results, the consistency of the estimate $\hat{f}_T(t)$ (see equations (7)–(9) in the L^2 sense has been proved by Brillinger (1996).

(3)

$$E\hat{f}_T(t) = \sum_k \alpha^U_{J,k}\phi_{J,k}(t) + O(2^J \times T^{-1}) \tag{A6}$$

$$\mathrm{Cov}(\hat{f}_T(t), \hat{f}_T(s)) \approx \frac{2\pi}{T} f_{\eta\eta}(0) \sum_k \phi_{J,k}(t)\phi_{J,k}(s) \tag{A7}$$

for sufficiently large T, the joint cumulants of order m are

$$O(2^{(m-1)J} T^{-m+1}) \tag{A8}$$

Accordingly, the fourth-, third-, second- and first-order cumulants of $\hat{f}_T(t)$ are $O(2^{3J} \times T^{-3})$, $O(2^{2J} \times T^{-2})$, $O(2^{J} \times T^{-1})$ and $O(1)$ respectively.

It is well-known that the relationships between moments and cumulants are

$$\begin{aligned} m_1 &= c_1 \\ m_2 &= c_2 + c_1^2 \\ m_3 &= c_3 + 3c_1c_2 + c_1^3 \\ m_4 &= c_4 + 3c_2^2 + 4c_1c_3 + 6c_1^2c_2 + c_1^4 \end{aligned} \tag{A9}$$

and so on. Hence

$$E(x - Ex)^4 = c_4 + 3c_2^2 \tag{A10}$$

i.e.

$$E|\hat{f}_T(t) - E\hat{f}_T(t)|^4 = O(2^{3J} \times T^{-3}) + O(2^{2J} \times T^{-2}) \tag{A11}$$

By the well-known Markov inequality we have

$$P\{|\hat{f}_T(t) - E\hat{f}_T(t)| \geqslant \varepsilon\} \leqslant c \times 2^{2J} \times T^{-2}/\varepsilon^4 \tag{A12}$$

for any $\varepsilon > 0$, so that

$$\sum_{T=1}^{\infty} P\{|\hat{f}_T(t) - E\hat{f}_T(t)| \geqslant \varepsilon\} \leqslant \sum_{T=1}^{\infty} \frac{c2^{2J}}{\varepsilon^4 T^2} < +\infty \tag{A13}$$

this leads to

$$\hat{f}_T(t) \to E\hat{f}_T(t), \text{ a.s.} \tag{A14}$$

namely,

$$\hat{f}_T(t) = f_J(t) + O\frac{2^J}{T} \tag{A15}$$

where

$$f_J(t) = \sum_k \alpha_{J,k} \phi_{J,k}(t) \tag{A16}$$

Then, using a Theorem in Walter (1994, Chapter 2) when $\phi(t) \in S_r$, r is a positive integer, and when $J \to \infty$,

$$f_J(t) \to f(t) \text{ a.e.} \tag{A17}$$

that shows $\hat{f}_T(t) \to f(t)$ a.s. (A18)

for sufficiently large T, J and $2^J/T \to 0$.