Summarizing Time Series: Learning Patterns in 'Volatile' Series

Saif Ahmad, Tugba Taskaya-Temizel, and Khurshid Ahmad

Department of Computing, University of Surrey, Guildford, GU2 7XH
Surrey, UK
{s.ahmad,t.taskaya,k.ahmad}@surrey.ac.uk
http://www.computing.surrey.ac.uk

Abstract. Most financial time series processes are nonstationary and their frequency characteristics are time-dependant. In this paper we present a time series summarization and prediction framework to analyse nonstationary, volatile and high-frequency time series data. Multiscale wavelet analysis is used to separate out the trend, cyclical fluctuations and autocorrelational effects. The framework can generate verbal signals to describe each effect. The summary output is used to reason about the future behaviour of the time series and to give a prediction. Experiments on the intra-day European currency spot exchange rates are described. The results are compared with a neural network prediction framework.

1 Introduction

Understanding and interpreting time varying phenomena are amongst the key challenges in various branches of science and technology. Autoregressive analysis of time series has been carried out over the last 50 years [2] with encouraging results. However, such techniques do not quite explain nonstationary phenomena [6], which may be characterised by long-range dependencies [22].

Techniques for time series analysis have ranged from machine learning approaches, which use artificial neural networks for prediction of stock time series [23], to genetic algorithms that learn to predict [10]. Data mining concepts with the aim to discover hidden patterns, similarity searches, and incremental mining have also been applied to time serial databases [11], [12] and [20]. Agrawal et al. introduce the concept of a shape definition language (SDL), which allows a variety of queries about the shapes, found in historical time sequences [1]. Natural language generation (NLG) based systems produce English language summaries of time varying phenomena by performing tasks such as microplanning and syntactic realisation on the analysed data [3] and [19].

Financial time series exhibit quite complicated patterns (for example, trends, abrupt changes, and volatility clustering), which appear, disappear, and re-appear over time [8]. Such series are often referred to as nonstationary, whereby a variable has no clear tendency to return to a fixed value or linear trend. Most of the time series

Z.R. Yang et al. (Eds.): IDEAL 2004, LNCS 3177, pp. 523–532, 2004.

analysis techniques discussed above, fail to address the transient nature of such time series. In short, ‘sharp variations’ [14] in a time series are excluded, and these variations are of interest to theoreticians and practitioners.

Wavelet analysis is a relatively new field in signal processing. Wavelets are mathematical functions that ‘cut’ up data into different frequency components, and then study each component with a resolution matched to its scale – a scale refers to a time horizon [7]. Wavelet filtering is particularly relevant to volatile and time-varying characteristics of real world time series and is not restrained by the assumption of stationarity [16]. The wavelet transform decomposes a process into different scales [15], which makes it useful in differentiating seasonalities, revealing structural breaks and volatility clusters, and identifying local and global dynamic properties of a process at these timescales [9]. Wavelet analysis has been shown to be especially productive in analysing, modeling, and predicting the behaviour of financial instruments as diverse as shares and exchange rates [4], [17] and [18].

In this paper, we propose an automatic time series analysis approach based on providing a summary of the data with respect to the ‘chief features’ of the data, which may be predictive of future events and behaviour. We extract important occurrences (for example turning points) from the data and look for their possible reappearance in the future. Our time series summarization framework adapts concepts from multiscale wavelet analysis to deal with nonstationary, nonperiodic, and volatile financial time series data. More specifically, our framework uses the discrete wavelet transform (DWT) and multiscale volatility analysis to summarize features like trend, cycle (seasonality), turning points and variance change in the original data and facilitates a prediction based on these features. We present our framework and report results on Intraday tick data for the British Pound (£) – US Dollar ($) exchange rate series and compare our results with a neural network prediction framework: the fit to the data is quite good (mean square error = 0.0000381) and compares well with predictions based on the use of neural networks where the mean square error is around 20 times higher.

2 Motivation

With the advent of the Internet and online data vendors, more and more financial time series data is being recorded, supplied and stored online. In financial markets, traders both ‘bid’, price at which they are prepared to buy and ‘ask’, price at which they will sell. This system of bid / ask pricing ensures the sell / purchase of instruments without any delay. A day’s trading (comprising almost 24 hours) for such data could generate 25,000 to 30,000 *ticks* per day per instrument.

The engineering of this high-frequency data, that is acquiring, preprocessing, and analysing the data is made more complex by nonstationarities in the data. The tick data arrives at irregular intervals and in order to use time series analysis techniques on such unordered data, some pre-processing is required. Data *compression* is one such pre-processing technique that aggregates the movement in the dataset over a certain period of time. The *compression* acts as a surrogate for the original: the maxima (High) and minima (Low) of the data over a fixed interval (typically, one minute) and the value at the start (Open) and the end (Close) of the minute acts as the surrogate for other data during the minute. Data *compression* essentially yields four new time series: Open, High, Low, and Close data values. For instance, the Intraday trading of an exchange rate instrument (£/$) comprising over 25,000 data points can be compressed into 1-minute slices resulting in 1440 data points per day.

In order to identify key patterns of behaviour in a non-linear time series x_t, particularly in its return value, $r_t = \log (x_t/x_{t-1})$, or its volatility, $v_t = |r_t|$, it is important to understand that there may be purely local changes in time domain, global changes in frequency domain, and there may be changes in the variance parameters. The discrete wavelet transformation (DWT) is one such technique for parameterizing this dynamic behaviour.

Neural network literature suggests that they can learn the behaviour of a time series and produce results that compare well with other nonlinear approaches to time series analysis – for example, nonlinear autoregressive (AR) analysis. Instead of statistically computing AR coefficients, a neural network learns the behaviour of the time series by a change in the weights of the interconnected neurons comprising the network. Once trained, the network can predict the future behaviour of the time series [5] and [21].

In this paper, we will compare the two parameterizing techniques, namely, wavelet analysis and neural networks.

3 Wavelet Analysis of Time Series: Annotation and Prediction

3.1 A Brief Note on the DWT

The Discrete Wavelet Transform represents a signal as a sum of approximations (As) and details (Ds) that are localized in time and frequency. The DWT is a discrete convolution process that can be expressed by the following formula:

$$w * x_t = \sum_{i=-\infty}^{\infty} w_i x_{t-i} \tag{1}$$

Here x_t is the original signal while w is the low- or high-pass filter corresponding to the *prototype* wavelet. In practice, the DWT is implemented via a pyramidal algorithm [13]. By achieving good time-frequency resolution the DWT is able to tame nonstationarities and volatilities to discover interesting local and global patterns in a signal. In Eq (1), w_i are a set of parameters that encapsulate the behavior of a nonlinear time series and as such wavelet analysis can be used to learn various behavioral patterns in a time series. The details (D_i) at various levels (i = 1 to L) of decomposition help capture local fluctuations over the whole period of a time series. The key frequency components at each decomposition level helps in summarizing the behavior of a time series up until the immediate past as also the future behavior. The highest-level approximation (A_L) captures the overall movement (trend) of the time series.

The decomposition of a series into the details and approximations will help, as we will show, in the annotation of the series: wavelet analysis can be used to identify variance change in addition to trend and seasonal components. Once identified, a set of pre-stored phrases can be attached to the series as a whole or to discrete points in the series – the annotation of volatilities, trends and seasonal variations can then be produced automatically. The details and approximations can be used to forecast the future seasonal variations and trends [17].

3.2 Algorithms for Summarizing and Predicating Patterns

We have developed two algorithms that can automatically summarize and predict a locally volatile time series (tick data) in terms of the 'chief features' of the series including turning points, trend, cycle, and variance change.

I. Compress the tick data to get Open (O), High (H), Low (L) and Close (**C**) value for a given compression period (for example, one minute or five minutes).

II. Calculate the level **L** of the DWT needed based on number of samples **N** in **C** of Step I,
L = floor [log (N)/log (2)].

III. Perform a level-**L** DWT on **C** based on results of Step I and Step II to get, $\mathbf{D_i}$, i = 1, . . ., L, and $\mathbf{A_L}$.

III-1. Compute **trend** by performing linear regression on $\mathbf{A_L}$.

III-2. Extract **cycle** (seasonality) by performing a Fourier power spectrum analysis on each $\mathbf{D_i}$ and choosing the $\mathbf{D_i}$ with maximum power as $\mathbf{D_S}$,

$$D_i(k) = (1/N) \sum_{t=0}^{N-1} D_i(t).e^{-j2\pi f_k t}, k = 0,1,\ldots, N-1$$

III-3. Extract **turning points** by choosing extremas of each $\mathbf{D_i}$.

IV. Locate a **single variance** change in the series by using the NCSS index on **C**,

$$\tilde{P}_k = \frac{\sum_{t=L_j-1}^{k} \tilde{w}_{j,t}^2}{\sum_{t=L_j-1}^{N-1} \tilde{w}_{j,t}^2}, \; k = L_j - 1, \ldots, N-2$$

where, w_j is the *level-j* DWT of the volatility series $\boldsymbol{v_t}$ of **C**.

V. Generate a graphical and verbal **summary** for results of Steps III-1 to III-3 and IV.

Fig. 1. Time series summarization algorithm.

In Fig. 1., we present the first algorithm – the *time series summarization algorithm*. It uses the DWT to process the raw time series and its first difference expressed through its volatility to extract numerical values corresponding to the 'chief features'. The system then generates a graphical and a verbal summary, describing the market dynamics at different scales (time horizons). The separation of a time series into its time scale components using the DWT facilitates forecasting by applying the appropriate procedure to each component – the aggregate forecast can then be obtained by recombining the component series.

The second algorithm, *time series prediction algorithm*, is presented in Fig. 2. After the time series has been summarized and separated into 'key' components, the trend and seasonal components are projected separately and then recombined to give an aggregate forecast. The seasonal component is symmetrically extended based on its distinct amplitude and period. The trend, which is linear in the level-L DWT approximation, is modeled by a first order polynomial function. The time series prediction algorithm does not predict the exact future values of a time series; it rather gives the overall market movement and major fluctuations for the forecast period specified. If we are able to extend all the wavelet components (A_L and every D_i) and recombine the individual extensions, we can get an exact forecast for the specified period. However, the dynamics of other wavelet components, for example the irregular fluctuations, are more complicated and need more study before they can be modeled or extended with a fair degree of confidence.

I. Summarize the tick data using the time series summarization algorithm of Fig. 1.

II. For a N-step ahead **forecast**, extend the **seasonal** component $\mathbf{D_S}$ *symmetrically* N points to the right to get $\mathbf{D_{S,\ forecast}}$.

III. For a N-step ahead **forecast**, extend the **trend** component $\mathbf{A_N}$ *linearly* N points to the right to get $\mathbf{A_{N,\ forecast}}$.

IV. Add the results of Steps II and III to get an **aggregate** N-step ahead **forecast**,
$Forecast = D_{S,\ forecast} + A_{N,\ forecast}$

Fig. 2. Time series prediction algorithm.

4 A Prototype System for Annotating and Predicting Time Series

We have developed a prototype system in Matlab®, a commercial mathematical package developed by *The MathWorks, Inc.* The system can automatically annotate and predict a time series. The prototype has sub-systems for compressing tick data, performing the DWT analysis, and the fast Fourier transform analysis on the compressed data. The compressed data is decomposed into trend and seasonal components for annotation purposes; annotation being performed by using a set of phrases that are selected by looking, for example, at the derivative of the seasonal and trend components. The prediction module of the prototype can project these components separately and recombine them to give an aggregate forecast.

4.1 An Experiment on the £/$ Tick Data

Consider the five minutes compressed tick data for the £/$ exchange rate on March 18, 2004 (Fig. 3). We use our summarization algorithm (Fig. 1) and prediction algorithm (Fig. 2) to process this data. The results are compared with a neural network prediction framework employing multi-layer perceptrons (MLPs); the DWT results look more promising than the MLP.

Fig. 3. Analyzed signal – the £/$ exchange rate on March 18, 2004 compressed to 5 minutes.

4.2 Annotation

The graphical summary produced by the system using the time series summarization algorithm is shown in Fig. 4.

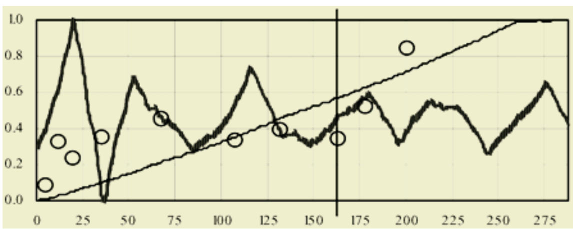

Fig. 4. Graphical summary of series of Fig. 3 showing trend, seasonality, turning points and variance change.

The open circles show turning points, the vertical line shows the variance change location, the thick fluctuating line shows the seasonal component and the thin slanting line shows the trend. The values of all components in Fig. 4 have been scaled between zero and one for display purposes.

The verbal summary of the extracted features produced by the system is shown in Table 1, which can be used to annotate a time series. The trend information suggests a long-term upward movement. However, a major inflexion point at t = 260 where the slope drops drastically by 94 percent suggests a downtrend. The cyclical component peaks with a period of 30 to 60, suggesting that this seasonal behaviour will continue in the near future.

After summarizing the time series and separating it into its time scale components using the DWT, we are now ready to project these components separately and recombine the projections to get an aggregate forecast for the next day.

Table 1. Verbal summary of series of Fig. 3.

Feature	Phrases	Details
Trend	*1st Phase*	$x_1^{Trend} = 6.36e - 5\,t + 1.81, t < 260$
	2nd Phase	$x_2^{Trend} = 3.65e - 6\,t + 1.83, 261 < t < 288$
Turning Points	*Downturns*	108, 132, 164, and 178
	Upturns	5, 12, 20 36, 68, and 201
Variance Change	*Location*	164
Cycle	*Period*	42
	Peaks at	21, 54, 117, 181, 215, and 278

4.3 Prediction

For prediction, we use the 'chief features' of the previous day (March 18, 2004), the trend and information about the dominant cycle (Table 1), to reproduce the elements of the series for the following day (March 19, 2004). The prediction results are shown in Fig. 5. The prediction (bottom curve) is in agreement with the actual time series (top curve), which shows a downturn. The cyclical fluctuations in the prediction curve do not seem to match too well and there is an observable divergence after observation number 125. However, as the market seems to pick up after observation number 250, the prediction curve also starts to pick up. The correlation between the predicted and actual time series is 62.4 % and the mean square error is 0.0000381.

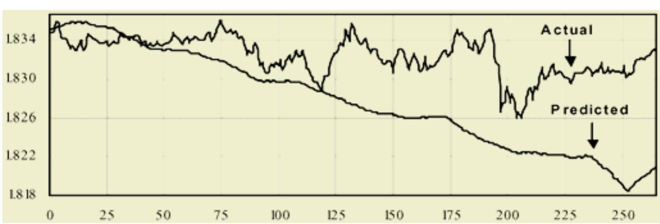

Fig. 5. Next day (March 19, 2004) forecast (bottom curve) of series of Fig. 3 along with the actual series (top curve) for the next day.

4.4 Predicting with Neural Networks: A Comparison

For evaluating our results, we use a multi layer perceptron (MLP) to perform nonlinear AR prediction on the five minutes compressed tick data for the £/$ exchange rate on March 18, 2004 (Fig. 3). This MLP has the following configuration: 11 input layers, 8 hidden layers and 1 output layer. The number of tapped delays in the input layer is 10. Backpropagation is used for training to predict the next value. Fig. 6 shows the results of the analysis.

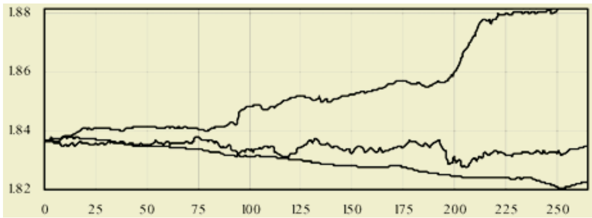

Fig. 6. Forecast comparison: from top, the first curve is the MLP prediction, second is the actual (test) data and the third is the DWT prediction.

There is clearly a greater divergence from the original for the MLP prediction, the root mean square error being about 17 times that of the DWT prediction (0.000673 vs. 0.0000381). The basic assumption in the MLP prediction approach is that of short-range dependency, where it is assumed the next value of a time series is dependant only on the previous value. However for many financial time series data, for example foreign exchange rates, the correlations between variables do not decay at a sufficiently fast rate and observations separated by great periods of time would still exhibit significant correlation. Such time series are said to be generated by long-memory or long-range dependent processes and require different approaches to modeling than the so-called short-memory processes (for example AR models). The wavelet analysis has been shown to approximately decorrelate time series with long memory structure [22]. This provides a sound technique for testing and modeling nonstationary features without knowing the exact nature of the correlation structure of a given time series. This could perhaps be the reason for a better fit to the data obtained using the DWT as compared to the MLP (Fig. 6).

Table 2 shows a comparison of the mean square error and correlation statistics for the two methods. The mean square error for the MLP prediction is 17 times higher than the DWT prediction. Moreover, the correlation between the actual and predicted for the DWT is very good (+ 62.4 %) as compared to a negative correlation (- 61.8 %) for the MLP.

Table 2. Comparison between DWT and MLP.

	Prediction	**Mean Square Error**	**Correlation**
DWT	Trend + Seasonality	0.0000381	+ 62.4 %
MLP	All Values	0.000673	- 61.8 %

5 Conclusions

In this paper we have presented a time series summarization, annotation, and prediction framework based on the multiscale wavelet analysis to deal with nonstationary, volatile and high frequency financial data. We have shown that the multiscale analysis can effectively deconstruct the total series into its constituent time scales: specific forecasting techniques can be applied to each timescale series to gain efficiency in forecast. The results of experiments performed on Intraday exchange data show promise and clearly point towards the advantages of the wavelet analysis over neural networks for summarizing and predicting highly volatile time series. However, continuously evolving and randomly shocked economic systems demand for a more rigorous and extended analysis, which is being planned.

The next and perhaps theoretically and empirically more challenging step is to consider and understand the dynamics of other time scale components generated by the wavelet analysis, for example the irregular fluctuations and the decorrelated white noise. Successful analysis of agents operating on several scales simultaneously and of modeling these components could result in more exact forecasts.

References

1. Agrawal, R., Psaila, G., Wimmers, E. L., and Zait, M., “Querying Shapes of Histories,” *Proceedings of the International Conference on Very Large Data Bases (VLDB)*, pp. 502-514, 1995.
2. Box, G. E. P, and Jenkins, G. M., “Time Series Analysis: Forecasting and Control,” *Time Series and Digital Processing*, Second Edition, Holden Day, San Francisco, 1976.
3. Boyd, S., “TREND: A System for Generating Intelligent Description of Time-Series Data”, *Proceedings of the IEEE ICIPS*, 1998.
4. Brock, W. A., “Whither Nonlinear?,” *Journal of Economic Dynamics and Control*, vol. 24, pp. 663-678, 2000.
5. Campolucci P., Piazza F., “On-Line Learning Algorithms for Locally Recurrent Neural Networks,” *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 253-271, 1999.
6. Chatfield, C., “The Analysis of Time Series: An Introduction,” Fifth Edition, Chapman and Hall/CRC, 1996.
7. Daubechies, I., “Ten Lectures on Wavelets,” SIAM, Philadelphia, 1992.
8. Gençay, R., Selçuk, F., Whitcher, B., “An Introduction to Wavelets and Other Filtering Methods in Finance and Economics,” Academic Press, 2002.
9. Gençay, R., Selcuk, F., Whitcher, B., “Differentiating Intraday Seasonalities Through Wavelet Multi-scaling,” *Physica A*, vol. 289, pp. 543-556, 2001.
10. Kaboudan, M. A., “Genetic Programming Prediction of Stock Prices,” *Computational Economics*, vol. 16, pp. 207-236, 2000.
11. Keogh, E., and Pazzani, M., “An Indexing Scheme for Fast Similarity Search in Large Time Series Databases,” *Proceedings of 11th Int’l Conference on Scientific and Statistical Database Management*, 1999.

12. Keogh, E., Chakrabarti, K., Pazzani, M., and Mehrotra, S., "Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases," *SIGMOD Record* (ACM Special Interest Group on Management of Data), pp. 151-162, 2001.
13. Mallat, S., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, 1989.
14. Mallat, S., and Zhong, S., "Characterization of Signals from Multiscale Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710- 732, 1992.
15. Meyer, Y., "Wavelets: Algorithms and Applications," *Society for Industrial and Applied Mathematics*, Philadelphia, 1993.
16. Qian, S., and Chen, D., "Understanding the Nature of Signals whose Power Spectra Change with Time," *IEEE Signal Processing Magazine*, 1999.
17. Ramsey, J. B., and Zhang, Z., "The Analysis of Foreign Exchange Data Using Waveform Dictionaries," *Journal of Empirical Finance*, vol. 4, pp. 341-372, 1997.
18. Renaud, O., Starck, J., L., and Murtagh, F., "Prediction Based on a Multiscale Decomposition," *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 1, No. 2, pp. 217-232, 2003.
19. Sripada, S. G., Reiter, E., Hunter, J., and Yu, J., "Generating English Summaries of Time Series Data Using the Gricean Maxims", *SIGKDD*, USA, 2003.
20. Walid, G. A., Mohamed, G. E., and Ahmed, K. E., "Incremental, Online, and Merge Mining of Partial Periodic Patterns in Time Series Databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 3, pp. 332-342, 2004.
21. Wan E., "Time Series Prediction Using a Neural Network with Embedded Tapped Delay-Lines in Predicting the Future and Understanding the Past," *SFI Studies in the Science of Complexity*, Addison-Wesley, Eds. A. Weigend , N. Gershenfeld, 1993.
22. Whitcher, B., Percival, D. B. and Guttorp, P., "Multiscale Detection and Location of Multiple Variance Changes in the Presence of Long Memory," *Journal of Statistical Computation and Simulation*, Vol. 68, pp. 65-88, 2000.
23. Zemke, S., "Bagging Imperfect Predictors," *Proceedings of Artificial Neural Networks in Engineering*, St. Louis, Missouri, pp. 1067-1072, 1999.