HIDDEN MARKOV INDEPENDENT COMPONENT ANALYSIS AS A MEASURE OF COUPLING IN MULTIVARIATE FINANCIAL TIME SERIES

Nauman Shah, Stephen J. Roberts

Pattern Analysis and Machine Learning Group, Department of Engineering Science
Oxford-Man Institute of Quantitative Finance
University of Oxford
nauman@robots.ox.ac.uk, sjrob@robots.ox.ac.uk

ABSTRACT

Modelling the dynamics of financial markets has been an area of active research in recent years. This paper presents a time series analysis model which can be used to infer patterns within financial data, in order to better understand the dynamics of financial markets. The focus of the paper is on finding causal and time-scale relationships between financial time series. Wavelets are used to extract useful time-scale information from financial data at different frequencies and mutual information between time series is used as the canonical measure of coupling. A Hidden Markov Independent Component Analysis (HMICA) model is used to infer a series of hidden states and it is shown that these hidden states are indicative of changes in mutual information between time series at various different time scales.

Keywords: ICA, Wavelets, Hidden Markov ICA, Financial Time Series Analysis

1 INTRODUCTION

One of the most prominent outcomes of research carried out in the field of financial engineering has been the development and rapid growth of algorithmic trading strategies. Due to the vast scale of the global financial markets and their constant evolution, research in this sector presents real challenges and opportunities. Algorithmic trading, also known as black-box or technical trading, is an automated trading platform, which relies on complex mathematical and statistical algorithms to make online trading decisions. Since the introduction of electronic trading in 1971, the proportion of trades that can be attributed to algorithmic trading has steadily increased. Algorithmic trading now accounts for over 60% of all trades taking place on the London Stock Exchange [1]. Many of the algorithmic trading engines currently in use act as a black-box to harness market inefficiencies in order to generate consistent positive returns. Till recently, algorithmic trading was usually done on an hourly basis. However, due to the easy and relatively cheap availability of high frequency market data, some of the latest algorithmic trading engines trade on a second by second or even tick by tick basis. The time series analysis model presented in this paper can potentially be used in the development of an algorithmic trading strategy.

2 TIME-FREQUENCY REPRESENTATION OF FINANCIAL TIME SERIES

There are a variety of methods currently in use to determine the time-frequency representation of a financial time series. The Fourier Transform, Empirical Mode Decomposition and Wavelet Analysis are all popular time-frequency analysis techniques. The method used in this paper is the Continuous Wavelet Transform, primarily because it allows for the inclusion of a prior distribution as its basis. Thus, prior knowledge can easily be incorporated into a trading model. Studies focusing on analysis of financial time series using a time-scale approach show some very promising results [5]. Therefore, there is potential for significant new developments in this field, especially considering the fact that almost all asset classes including Foreign Exchange (FX) and Equities exhibit time-scale behaviour.

2.1 Continuous Wavelet Transform

Wavelets are functions which can be used to represent a signal in a form which can be more easily analysed and comprehended. Wavelets allow for the localised analysis of signal components, which makes them especially interesting for use in dealing with FX data. The Continuous Wavelet Transform (CWT) is a powerful data analysis method which can be used to analyse the properties of a financial time series at different frequencies. The knowledge gained about the trends of a time series at different time scales can be used to develop trading models which take advantage of recurring patterns at various different frequencies. The CWT of a function $x(t)$ is given by:

$$T(u,b) = \frac{1}{\sqrt{u}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{u}\right) dt \qquad (1)$$

where u is the dilation parameter, also known as the scale, and b is the localisation parameter. The function $\psi(t)$, from which different dilated and translated versions are derived is called the mother wavelet.

The Morlet wavelet is used in all the analysis presented later in this paper. Morlet is a non-orthogonal wavelet which has both a real and a complex part. Such wavelets are also referred to as the analytical wavelets. Due to the complex part, Morlet wavelets can be used to separate both the phase and amplitude parts of a signal. A Morlet wavelet is represented by:

$$\psi(t) = \pi^{-\frac{1}{4}} \exp\left(i2\pi f_0 t\right) \exp\left(-\frac{t^2}{2}\right) \qquad (2)$$

3 MUTUAL INFORMATION

The Mutual Information (MI) of two signals is the canonical measure of coupling between the signals. For uncoupled signals MI is zero, whereas for coupled signals it has a positive value. The MI between two time varying signals, $\mathbf{x_1} = x_1(t)$ and $\mathbf{x_2} = x_2(t)$, is given by the Kullback-Leibler (KL) divergence [2]:

$$I[\mathbf{x_1}, \mathbf{x_2}] = KL(p(\mathbf{x_1}, \mathbf{x_2}) \,||\, p(\mathbf{x_1})p(\mathbf{x_2})) \qquad (3)$$

The Kullback-Leibler (KL) divergence between two dependent probability density functions, $p(\mathbf{x_1})$ and $p(\mathbf{x_2})$, is given by:

$$KL[p(\mathbf{x_1}) \,||\, p(\mathbf{x_2})] = \int_t p(\mathbf{x_1}) \log\left(\frac{p(\mathbf{x_1})}{p(\mathbf{x_2})}\right) dt \qquad (4)$$

Therefore, the MI is represented by:

$$I[\mathbf{x_1}, \mathbf{x_2}] = -\int\int p(\mathbf{x_1}, \mathbf{x_2}) \log\left(\frac{p(\mathbf{x_1})p(\mathbf{x_1})}{p(\mathbf{x_1}, \mathbf{x_2})}\right) d\mathbf{x_1} d\mathbf{x_2} \qquad (5)$$

4 The Wavelet-MI Algorithm

This section presents the Wavelet-MI model, which uses the wavelet coefficients to calculate the mutual information between two financial time series at any given time scale. For N time series, $\mathbf{x}_t = [x_1(t), x_2(t), ..., x_N(t)]$, analysed using the CWT at scales of 1 to u, the wavelet coefficients can be combined into a single matrix, $C_t = [c_1(t), c_2(t), ..., c_u(t)]$. Thus, the set of signals, $\mathbf{x}_t$, can be represented in terms of the wavelet coefficients, C_t, and the mother wavelet ψ_t:

$$\mathbf{x}_t = C_t \psi_t \qquad (6)$$

The CWT of a signal is given by equation 1. Equation 5 presents the MI equation. These two equations can be combined as shown in equation 7. The Wavelet-MI algorithm computes the mutual information,

$I[\mathbf{c_1}, \mathbf{c_2}]$, between the recovered wavelet coefficients, $\mathbf{c_1} = c_1(t), \mathbf{c_2} = c_2(t)$, of a multivariate time series, $\mathbf{x_1} = x_1(t), \mathbf{x_2} = x_2(t)$. Figure 1 shows the variation in MI across different time scales for various currency pairs. It is evident that MI and thus coupling between two currency pairs generally increases across time scales.

$$I[\mathbf{c_1}, \mathbf{c_2}] = -\int\int p(\mathbf{c_1}, \mathbf{c_2}) \log\left(\frac{p(\mathbf{c_1})p(\mathbf{c_1})}{p(\mathbf{c_1}, \mathbf{c_2})}\right) d\mathbf{c_1} d\mathbf{c_2} \qquad (7)$$
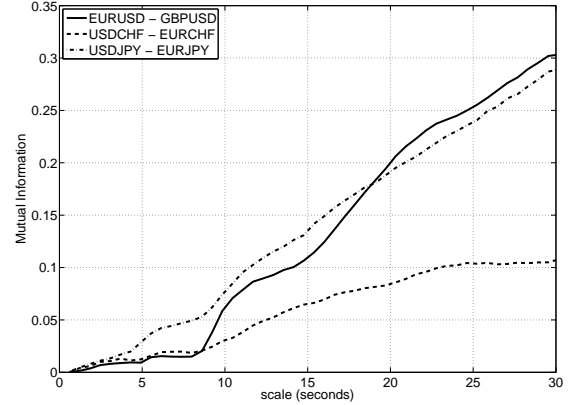

Figure 1: Mutual Information of three currency pairs, EURUSD-GBPUSD, USDCHF-EURCHF and USDJPY-EURJPY, at various time scales

In this paper, the method used for computing MI is based on a data-interpolation technique known as the Parzen-window density estimation, as discussed in [4].

5 HIDDEN MARKOV INDEPENDENT COMPONENT ANALYSIS

5.1 Markov Model

A Markov process is a statistical process in which future probabilities are determined by only its most recent values. Using the product rule, the joint probability of a variable x can be written as [2]:

$$p(x_1, ..., x_N) = \prod_{n=1}^{N} p(x_n \mid x_1, ..., x_{n-1}) \qquad (8)$$

A first-order Markov model assumes that all the conditional probabilities of equation 8 are dependent on only the most recent observation and independent of all others. Thus, a first order Markov model can be represented by:

$$p(x_1, ..., x_N) = p(x_1) \prod_{n=2}^{N} p(x_n \mid x_{n-1}) \qquad (9)$$

which can be simplified to:

$$p(x_n \mid x_1, ..., x_{n-1}) = p(x_n \mid x_{n-1}) \qquad (10)$$

which is a simplified form of a first-order Markov model.

5.2 Hidden Markov Model

A Hidden Markov Model (HMM) is a statistical model consisting of a set of observations which are produced by an unobservable set of latent Markov model states. It is widely used within the speech recognition sector. Due to its numerous advantages in inferring the hidden states of a dynamic system, it is increasingly being used in the financial sector as well. The aim of using a HMM is to infer the hidden states from a set of observations. Mathematically, the model can be represented by [2]:

$$p(X \mid Z, \theta) = p(z_1 \mid \pi)[\prod_{n=2}^{N} p(z_n \mid z_{n-1}, A)] \prod_{m=1}^{N} p(x_m \mid z_m, \mathbf{B})$$

(11)

where $X = x_1, ..., x_N$ is the observation set, $Z = z_1, ..., z_N$ is the set of latent variables, $\theta = \pi, A, \mathbf{B}$ represents the set of parameters governing the model. The HMM is represented using Markov chains in Figure 2, showing the hidden layer with states $\mathbf{z}_t$ and the observed layer with observation set $\mathbf{x}_t$. Also shown in the figure is the state transition probability $P(z(t+1) \mid z(t))$ and the emission model probability $P(x(t) \mid z(t))$
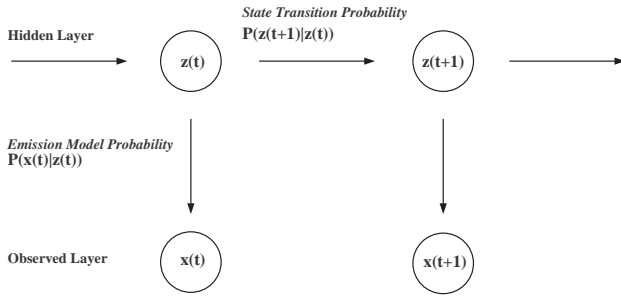

Figure 2: Hidden Markov Model graphical representation

5.3 Hidden Markov ICA Model

Independent Component Analysis (ICA) is a form of blind source separation model. The Hidden Markov ICA (HMICA) model is a Hidden Markov model with an ICA observation model. This section presents an overview of the HMICA model, using the set of equations detailed in [6].

Let $\theta = \pi, A, \mathbf{B}$ represent the parameters of a HMM, where $\mathbf{B}$ represents the parameters of the ICA observation model, A is the state transition matrix with entries a_{ij}, and π represents an initial state probability matrix. A HMM parameterised by some vector $\hat{\theta} = \hat{\pi}, \hat{A}, \hat{\mathbf{B}}$ can be trained using an Expectation-Maximisation (EM) algorithm as shown in [3].

Using analysis shown in [6], it can be proved that for an observation sequence $\mathbf{x}_t$ and hidden state sequence $\mathbf{z}_t$, the observation model parameters, $\mathbf{B}$, can be written in terms of an auxiliary function $\mathbf{Q}$, given by:

$$\mathbf{Q}(\mathbf{B}, \hat{\mathbf{B}}) = \sum_{k} \sum_{t} \gamma_k[t] \log p_{\hat{\theta}}(\mathbf{x}_t \mid \mathbf{z}_t)$$

(12)

where $\gamma_k[t]$ is the probability of being in state k. The log likelihood of the ICA observation model, with unmixing matrix $\mathbf{W}$ and M sources can be written as [6]:

$$\log p(\mathbf{x}_t) = \log |det(\mathbf{W})| + \sum_{i=1}^{M} \log p(a_i[t])$$

(13)

Substituting the ICA log likelihood, equation 13, into the HMM auxiliary function, equation 12, gives:

$$Q_k = \log |det(\mathbf{W}_k)| + \frac{1}{\gamma_k} \sum_{t} \gamma_k[t] \sum_{i} \log p(a_i[t])$$

(14)

The auxiliary function summed over all states k, becomes:

$$\mathbf{Q} = \sum_{k} Q_k$$

(15)

The HMICA model finds the unmixing matrix $\mathbf{W}_k$ for state k, by minimizing the cost function given by equation 15 over all underlying parameters.

6 RESULTS

This section presents the results obtained when the Wavelet-MI model presented in section 4 and the HMICA model presented in section 5 are simulated in Matlab. Figure 3 presents the Viterbi diagrams and the mutual information plots obtained by using FX data at various different time scales.

From the plots it is evident that there are significantly long periods of state stability. There is also some evidence of existence of recurring patterns, which can prove extremely useful in building a trading strategy. The HMICA code also gives the state transition matrix as an output. The state transition matrix gives the probability of change of state from state i to state j, i.e.:

$$P_{i,j} = P(z(t+1) = j \mid z(t) = i)$$

(16)

The state transition probability matrix, $P_{i,j}$, can also be written as:

$$P_{i,j} = \begin{pmatrix} p(0 \mid 0) & p(1 \mid 0) \\ p(0 \mid 1) & p(1 \mid 1) \end{pmatrix}$$

(17)

where $p_{j,i} = p(j \mid i)$ is the transition probability from state i to state j.

The state transition probability matrices, P_{scale}, for the USDJPY-EURJPY pair at various different time scales are given below. The Mutual Information diagrams and Viterbi plots for the USDJPY-EURJPY currency pair at various scales are presented in Figure 3.

$$P_5 = \begin{pmatrix} 0.9756 & 0.0244 \\ 0.0207 & 0.9793 \end{pmatrix}$$

(18)

$$P_{7.5} = \begin{pmatrix} 0.9761 & 0.0239 \\ 0.0162 & 0.9838 \end{pmatrix}$$

(19)

$$P_{10} = \begin{pmatrix} 0.9860 & 0.0140 \\ 0.0479 & 0.9521 \end{pmatrix}$$

(20)

(a) scale = 5 seconds

(b) scale = 7.5 seconds
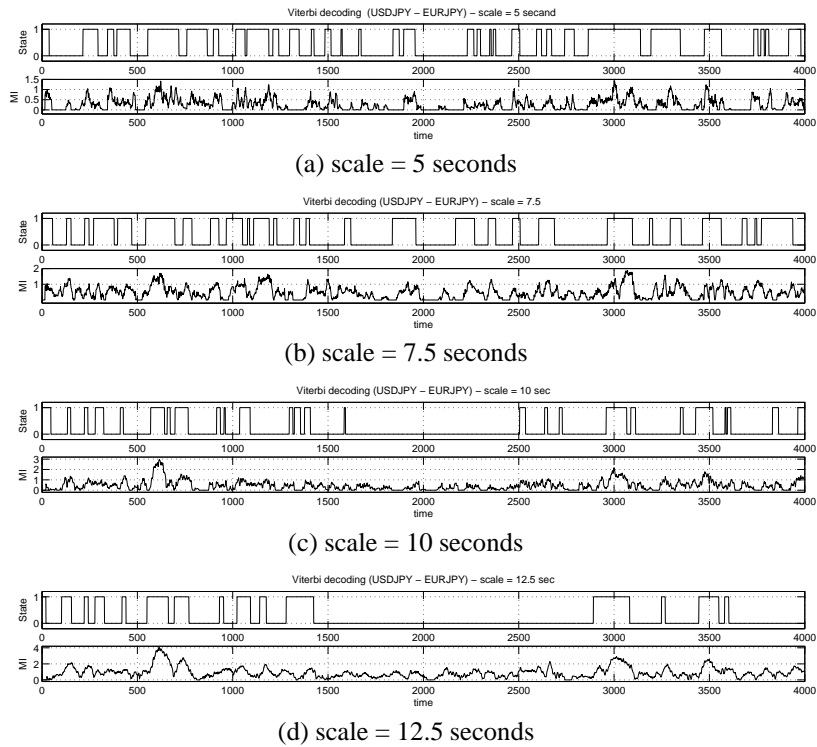

(c) scale = 10 seconds

(d) scale = 12.5 seconds

Figure 3: Viterbi diagrams showing state transitions in the Hidden layer for USDJPY-EURJPY at different time scales. Also shown are the Mutual Information (MI) plots of the currency pairs.

$$P_{12.5} = \begin{pmatrix} 0.9915 & 0.0085 \\ 0.0273 & 0.9727 \end{pmatrix} \qquad (21)$$

It is interesting to note that for significant portions of time, the length of time for which the state stays constant is over 100 samples (50 seconds) long. These periods of state stability are hence well-suited for placing a trade order. The state transition probability matrix, P_{ij}, can be used to make predictions about future states. Simulations conducted with Equities data using the models presented in this paper also give encouraging results.

7 CONCLUSIONS

This paper presents a statistical model for analysing the dynamics of multivariate financial time series. The CWT is presented as a useful tool for the analysis of financial data sets at various different frequencies. HMICA is used to extract the hidden states from multivariate financial time series. The hidden states stay constant for significant periods of time which is potentially useful for building efficient trading models. It is also shown that the hidden states are indicative of changes in mutual information between two FX returns time series.

References

[1] International Banking Systems Journal/Supplements/Trading Platforms Supplement. *International Banking Systems Journal*, June 2007.

[2] C.M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[3] AP Dempster, NM Laird, and DB Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[4] F. Long and C. Ding. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8): 1226–1238, 2005.

[5] P. Oswiecimka, J. Kwapien, S. Drozdz, and R. Rak. Investigating Multifractality of Stock Market Fluctuations Using Wavelet and Detrending Fluctuation Methods. *Acta Physica Polonica B*, 36(8):2447, 2005.

[6] W. Penny, R. Everson, and S.J. Roberts. Hidden Markov Independent Components Analysis. *Advances in Independent Component Analysis. Springer*, pages 3–22, 2000.