WHITEPAPER

THE LONG AND THE SHORT OF IT – NEW MEASURES OF VOLATILITY

ANDREW CLARK
CHIEF INDEX STRATEGIST
THOMSON REUTERS INDICES
September 2009

THOMSON REUTERS

THE LONG AND THE SHORT OF IT – NEW MEASURES OF VOLATILITY

INTRODUCTION

This is the accompanying white paper for the Thomson Reuters Webcast: "The Long and the Short of It – Measuring and Forecasting Volatility" first presented on Sept. 23, 2009. This paper will describe the tools used to model and forecast volatility as in the webcast.

The two main topics of this paper are wavelets and nonlinear time series analysis (a subset of dynamical systems). For readers unfamiliar with either of these two areas, it suggested they browse the first two books listed in Reference section at the end of the paper either before or after reading this paper.

WAVELETS

Throughout this discussion we will be referring to just one wavelet transform – the Modified Discrete Wavelet Transform (MODWT hereafter). The description that follows is adapted from Gencay *et al*. [1]

Let x be an arbitrary length vector of observations. The length $(J+1)N$ vector of MODWT coefficients ● is obtained via

$$\varpi = \tilde{W} x$$

Where $\tilde{W}$ is an $(J+1)\ N \times N$ orthonormal matrix defining MODWT.

The vector of wavelet coefficients maybe organized into $J + 1$ vectors

$$\varpi = \left[\varpi_1, \varpi_2, \ldots, \varpi_j, \upsilon_j\right]^T$$

where ϖ_j is a length $N/2^j$ vector of wavelet coefficients associated with changes on a scale of length $●_j = 2^{j-1}$ and $●_j$ is a length $N = 2^j$ vector of scaling coefficients associated with averages on the scale of $2^j = 2●_j$.

The matrix $\tilde{W}$ is made up of $J+1$ submatrices, each one of them $N \times N$. MODWT utilizes the rescaled filters $(j = 1, \ldots, J)$

$$\tilde{h} = h_j / 2^j \text{ and } \tilde{g} = g_j 2^j$$

To construct the $N \times N$ dimensional submatrix $\tilde{W}_1$,

we circularly shift the rescaled wavelet filter vector $\tilde{h}_1$ by integer units to the right so that

$$\tilde{W} = \left[\tilde{h}_1^{(1)}, \tilde{h}_2^{(2)}, \tilde{h}_3^{(3)}, \ldots \tilde{h}_1^{(N-1)}, \tilde{h}_1^{(N-1)}, \tilde{h}_j\right]^T$$

The remaining submatrices $\tilde{W}_2, \ldots \tilde{W}_j$ are formed

similarly by replacing $\tilde{h}_1$ by $\tilde{h}_j$.

In practice, a pyramid algorithm is commonly used to compute MODWT. This paper will not review that algorithm. The interested reader is referred to Gencay *et al* [1] for details on the pyramid algorithm.

The multiresolution analysis (MRA) used in the webcast to decompose the volatility signal into its various time components starts with utilizing the MODWT via

$$x_t = \sum_{j=1}^{J+1} \tilde{d}_{j,t} \text{ where } t = 0, \ldots, N-1$$

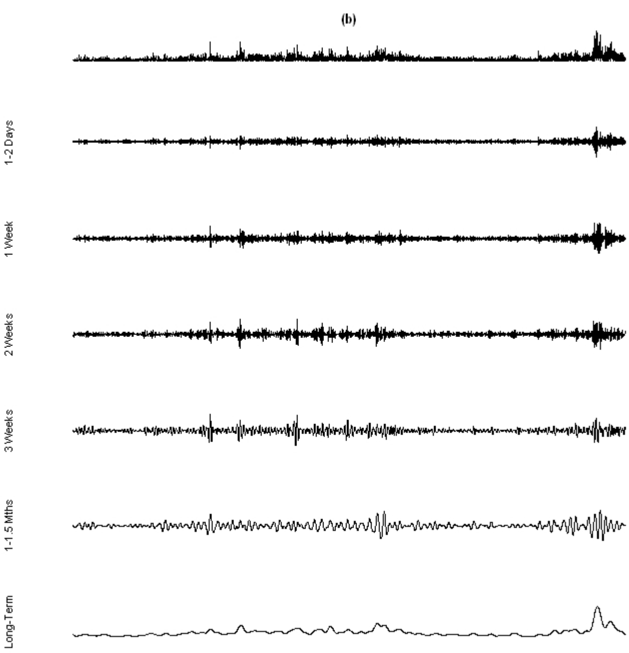
where $\tilde{d}_{j,t}$ is the tth element of $\tilde{d}_j = \tilde{W}_j^T \varpi_j$ for $j = 1, \ldots, J$.

THOMSON REUTERS

A key feature of the MRA using MODWT is that the wavelet details and the smooth are associated with zero-phase filters. This means the interesting features in the wavelet details and smooth can be perfectly aligned with the original time series. It also means the details are on the same scale as the original time series but decomposed to *J* details. So the smallest time interval (referred to subsequently as d1) covers 2^1 time periods. In the case of the webcast's volatility analysis, this smallest time interval equates to 2 days. Subsequent decompositions are assigned a time scale based on the appropriate value of 2^j. The final decomposition, referred to earlier as the smooth, was called long term volatility in the webcast.

Figure 1 shows the decomposition of the volatility of the TRX U.S. Equity Index

Figure 1: TRX U.S. Equity Index



As noted above, the wavelet details and the smooth – the 1-2 days through 1-1.5 month being the details and the long-term being the smooth – line up perfectly with the original series which is the first line graphed in Figure 1. Because of this, we can follow the flow of increased volatility at various time periods, in particular close to the very end of the series, a time period which corresponds to the volatility seen in 2008. Note that the changes in volatility are not equal across time scales, as is often assumed in standard economic and financial theories.

DYNAMICAL SYSTEMS

In non-linear time series analysis (an applied math subset of dynamical systems) the reconstruction of the vector space which is equivalent to the original state space of the time series data is the basis of almost all the methods this kind of analysis follows. Kantz and Schreiber [2], the authors of TISEAN, give an excellent presentation of these techniques which this paper will use.

In the webcast, we were looking for ways to exploit any determinism the data may have. This determinism comes about when the time evolution of a trajectory in the reconstructed space depends *only* on its current position in the new space. The uniqueness of the dynamics in the reconstructed space is not the only property we want, though. We also want dimensions, lyanpunov exponents and entropies. In order to guarantee that all the correct values are found for these variables, the structure of the tangent space, i.e. the linearization of the dynamics at any point in the state space, must be preserved by the reconstruction process. Thus an embedding of a compact smooth manifold into R^m is defined to be a map which is a one-to-one immersion on the manifold. Therefore the crucial point of this analysis is to show under what conditions the projection due to the scalar measurements and the subsequent reconstruction by the delay vectors form an embedding.

The problems with the embedding of scalar data into R^m are two-fold. First is the specification of the correct Jacobian which is full rank. If this specification can be done correctly, we will have a *D* dimensional manifold with *D* being the rank of the Jacobian. Unfortunately, this manifold will probably be curved. So the second problem is, since we would like to work with a global representation of the dynamics in some vector space, an embedding of the curved manifold in R^2 must be found.

The first vector space variable that is often estimated in nonlinear time series studies is the time delay or time lag. As there is no mathematical formalism in the literature for the time delay, this means there is no rigorous way of determining its optimal value.

THOMSON REUTERS

Because of this, Kantz and Schreiber take a pragmatic approach since all known methods yield values of the lag that are of the same order of magnitude. Their preferred method for the computation of the lag is the minimum of the mutual information. This minimum's rough corollary in classical statistics is the first value at which the autocorrelation of the data has its first minimum.

To compute mutual information, a histogram is created of a certain bin length or resolution. Denote by p_i the probability that the data assumes a value inside the *i*th bin of the histogram and let p_{ij}*(•)* be the probability that *s(t)* (the Shannon entropy) is in bin *i* and *s(t+•)* is in bin *j*. Then the mutual information for time delay • is

$$I_{\varepsilon}(\tau)=\sum_{i,j} p_{ij}(\tau)\ln p_{ij}(\tau)-2\sum_{i} p_i \ln p_i$$

The first minimum, then, when • is small, is the point at which $I_{\bullet}$*(•)* is large and whose next and subsequent points show declining values for $I_{\bullet}$*(•)*.

The next variable to be computed is the embedding dimension. A precise knowledge of this dimension allows us to exploit any determinism with minimal computational effort. An estimate of the embedding dimension is often obtained through a technique called *false nearest neighbors*, which is explained below.

The basic idea behind false nearest neighbors is to search for points in the data which are neighbors in embedding space, but which should not be neighbors since their future temporal evolution is too different. If the correct embedding dimension for the data is m_0 and one studies the data in a lower dimensional embedding such that $m< m_0$, the transition from m_0 to *m* is a projection thereby eliminating certain axes from R^m. These points whose coordinates which are eliminated by the projection must differ strongly and therefore are the false nearest neighbors in the *m* dimensional space.

To compute the false nearest neighbor, for each point in the time series take its closest neighbor in *m* dimensions and compute the ratio of the distance between these two points in *m+1* dimensions and *m*. If this ratio is larger than a threshold *r*, the neighbor is false. An important criterion for this threshold is that it has to be large enough to allow for the (possible) existence of exponential divergence due to deterministic chaos.

If we denote the standard deviation of the data by • and use the maximum norm, we can compute the nearest neighbor ratio in this way

$$X_{fnn}(r)=\frac{\sum_{n=1}^{N-m-1}\Theta\left(\frac{\left|s_n^{m+1}-s_{k(n)}^{m+1}\right|}{\left|s_n^{m}-s_{k(n)}^{m}\right|}-r\right)\Theta\left(\left(\frac{\sigma}{r}\right)-\left|s_n^{m}-s_{k(n)}^{m}\right|\right)}{\sum_{n=1}^{N-m-1}\Theta\left(\left(\frac{\sigma}{r}\right)-\left|s_n^{m}-s_{k(n)}^{m}\right|\right)}$$

where $s_{k(n)}^{(m)}$ is the closest neighbor to s_n in *m* dimensions. The first step function in the numerator is unity if the closest neighbor is false, i.e. if the distance increases by a factor more than *r* when increasing the embedding dimension by 1. The second step function suppresses all those pairs whose initial distance was already larger than *•/r*. Pairs whose initial distance is larger than *•/r* by definition cannot be false neighbors, since, on average, there is not enough space to depart farther than •. Hence these are invalid candidates for the method, which should not be counted, and is also reflected in the normalization.

The last topic in non-linear analysis will be how to the compute the maximal lyapunov exponent. Since a positive maximal lyanpunov exponent is a strong signature of chaos, it is of considerable interest to calculate its value. Several methods have been proposed over the years to compute this exponent, so in this paper we will use the method of Kantz and Schreiber [2] which tests directly for the exponential divergence of nearby trajectories, one of the sure bits of evidence in terms of determining a system's stability (maximal lyapunov <= 0) or instability (maximal lyapunov > 0 but < ∞).

Choose a point s_{n_0} of the time series in embedding space and select all neighbors with distance smaller than •. Compute the average of all the distances over the neighbors to the reference part of the trajectory as a function of the relative time. The logarithm of the average distance at time •*n* is some effective expansion rate over the time span •*n* containing all the deterministic fluctuations due to projections and dynamics. Repeating this is for many values of n_0, the fluctuations of the effective expansion rates will average out. Thus one computes:

$$S(n) = \frac{1}{N} \sum_{n=1}^{N} \ln\left(\frac{1}{\left|U\left(s_{n_0}\right)\right|} \sum_{s_n \in U\left(s_{n_0}\right)} \left| s_{n_0+\Delta n} - s_{n+\Delta n} \right| \right)$$

The reference points s_{n_0} are embedding vectors and $U\left(s_{n_0}\right)$ is the neighborhood of s_{n_0} with diameter •.

If the set of *S(n)* curves exhibits a robust linear increase, their slope is an estimate of the maximal lyapunov exponent • per time step. The word "robust" is stressed here because if the values of *S(n)* only increases in a linear fashion for selected values of *m* and • then one does not have a positive signature of exponential divergence.

This is ends our brief review of the wavelet and dynamical systems tools used in the webcast. What is not covered but was used for the webcast – noise reduction and prediction models – are not topics easily summarized in a paper like this. Interested readers are again referred to Kantz and Schreiber [2] and the many papers cited in their book for good sources on these topics. Interested readers are also directed to a variety of software listed in the Reference section that will help them with the use of wavelets and nonlinear time series analysis.

REFERENCES

Primary References

R. Gencay, F. Selcuk and B. Whitcher, *An Introduction to Wavelets and Other FilteringMethods in Finance and Economic,* Academic Press, 2002

H. Kantz, and T. Schreiber, *Nonlinear Time Series Analysis,* Cambridge University Press, 2005

Secondary Sources

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

M. M. Dacorogna, R. Gencay, U. Muller, R.B. Olsen, O.V. Pictet, *An Introduction to High Frequency Finance*, Academic Press, 2001

Software Used

R

- Waveslim – wavelet tool maintained by B. Whitcher
- tseriesChaos and RTisean – nonlinear time series analysis based on text by Kantz and Schreiber

Visual Recurrence Analysis – a nonlinear analysis tool available at http://www.visualization-2002.org/VRA_MAIN_PAGE_.html

FOR MORE INFORMATION

Thomson Reuters Indices
Andrew Clark
Chief Index Strategist
Tel: +1 303.357.0557
andrew.clark@thomsonreuters.com

THOMSON REUTERS