Simultaneous prediction of multiple financial time series using Supervised Learning and Chaos theory

Andrew. N. Edmonds	Prophecy Systems Ltd	34b Market Place, Olney Bucks, UK.
Diana Burkhardt	Luton University	Park Sq., Luton Beds, UK
Osei Adjei	Luton University	Park Sq., Luton Beds, UK

Abstract

"Embedded" time series are often used with Neural networks or other Supervised Learning Algorithms to generate predictions. Recent work in chaos theory has pointed to methods of determining the optimal embedding parameters for individual time series. The hypothesis is explored that these methods also hold when multiple time series are used together to generate a prediction, and that the optima for the individual series combined are the optimum for the group. A novel prediction explanation mechanism is described. Examples will be taken from foreign exchange time series, and the analyses will be performed using *The Prophet*, a time series prediction program. [Edmonds1]

Optimal embedding for a single time series

The process of "Embedding" as a means of making apparent the inner dynamics of chaotic time series was first described by David Ruelle in 1980 [Ruelle]. Floris Takens [Takens] in 1981 showed that, given a proper embedding, future values of a time series could be predicted to any given accuracy with a smooth function. Takens gave a clue to the correct embedding dimension for a given time series, but no clues as to the makeup of the smooth function.

The process of embedding a single time series is as follows: The scalar series is converted to a series of vectors.

$\mathbf{X}_t = x_t, x_{d+t}, x_{2d+t}, x_{3d+t}, \ldots x_{nd+t}$ Where d is the *Separation*, and n the *embedding dimension*.

In practice, in order to perform predictions for a given time series, first the separation is calculated, then the embedding dimension (this is so that the separation calculated can be used in the optimal dimension calculations). Finally the embedded vectors are used as a training set to some example of a supervised learning algorithm, along with some future value of the time series representing the point to be predicted. (in the above nomenclature I use x_{nd+t+1}.)

The training set thus contains pairs of embedded vectors drawn from the past history of the series on the stimulus or input side, and subsequent values of the series on the target or output side.

A Multivariate version of the above can be obtained by concatenating the stimulus, and separately the target, of several suitably embedded time series. The Supervised Learning Algorithm, commonly implemented as a neural net, learns to associate the input and output values, and when interrogated returns an interpolation between the respective targets of stimulus points that are close in the embedding space.

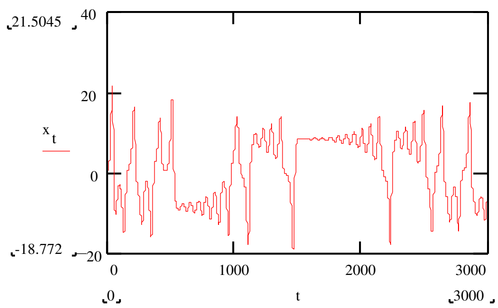


Fig. 1 Lorenz time series

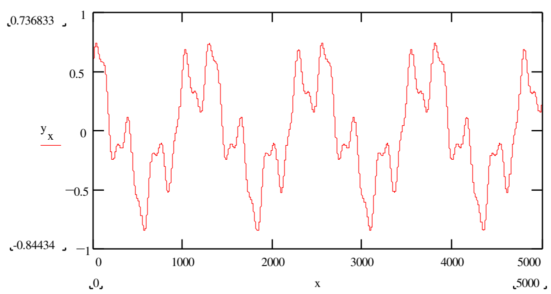
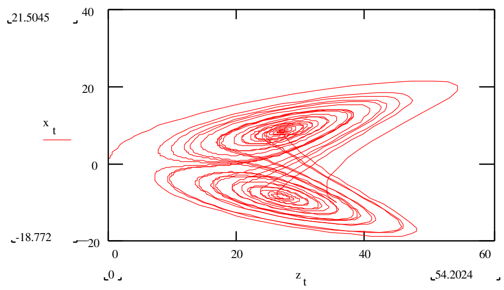


Fig. 2 A time series consisting of 3 sinusoids

Presented at the IEEE World Conference on Computational Intelligence, Orlando, June 94



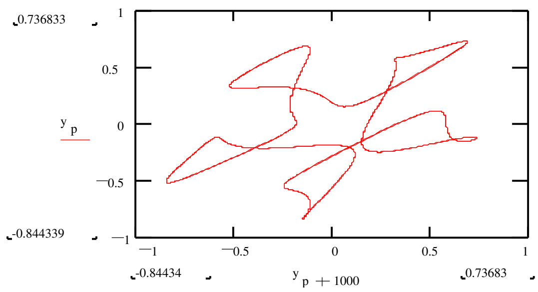


Fig. 3 The Lorenz attractor

Fig. 4 The attractor for the series in Fig. 2

It is interesting to note what happens as a result of embedding a time series. Figure 3 shows a view of the attractor associated with the Lorenz time series. (figure 1). Figure 2 shows a time series composed of several sinusoids. The Series in figure 1. is chaotic, whereas figure 2. contains a cyclic series. The attractor in figure 4 is a simple shape where each point is revisited multiple times. The Chaotic attractor (figure 3) has a "ball of string" like structure where any one point is never revisited, though many threads may lie close together. Whereas good and continual predictions are possible with figure 4, with a chaotic series, since we have only some nearby but not identical trajectories to predict from, the predictions can only be valid for a limited set of values ahead.

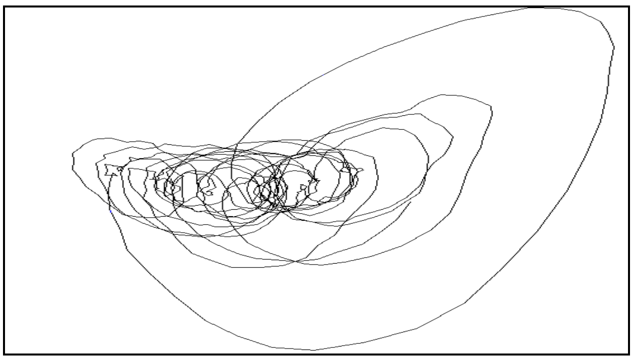


Fig. 5 DM intra-day data.
A picture of an attractor obtained by subtracting a seven point moving average from the raw data, embedding the result in 3 dimensions and smoothing.

Two methods have been published recently for calculating the optimum embedding dimension and separation for further processing.

Auto Mutual information

Mutual information, a concept from information theory, measures the average knowledge in bits (i.e. to the base 2) common to two sequences of numbers. So for example given one series, how much, in bits, do you know about another series. In principle the measure is similar to correlation, except that the result is in bits, and no assumptions about the underlying distributions are made. Two copies of the same time series are compared using this technique, then one series is offset in time from the other by one sample and the measurement is taken again, and again for increasing offsets. The mutual information typically falls and then bounces back for increasing offset. Figure 14 shows a typical sequence.
It is suggested by the authors of the paper first detailing this technique for this purpose, [Kennel2] that the first minimum be used to determine the optimum embedding separation. The logic of this is powerful. We want each dimension to offer as much new information as possible. Arranging the dimension samples to be separated by that separation for the minimum of the AMI curve ensures this, since there the mutual information is at a minimum.

False Nearest Neighbours

This simple and elegant method from [Kennel1] finds the optimum embedding dimension for even noisy or short time series. The separation is first found using Auto-mutual Information (AMI), then the series is embedded with a small dimension, say 2 to begin with. A search is made for each embedded point to find it's nearest neighbour and the Euclidean distance between them noted. The embedding dimension is increased by one for each point, (this requires fetching a further value of the time series) and the new distance calculated. If this new distance is a large increase over the previous distance, the points are marked as False Nearest Neighbours. The number of these found for a given dimension is noted. The process is repeated for larger dimensions. When the number has dropped to close to zero the search is terminated and the current dimension is returned as the result.

The selection of a group of series

Examples of work on multivariate time series prediction using supervised learning are rare. Chakraborty et al. [Chakraborty] and Weigend et al. [Weigend] are examples. Both chose which time series to use by consulting experts in the predicted series. (The series were financial). No attempt was made to evaluate which of the time series were useful in either case. There have been suggestions that series should be chosen to have low correlation, and thus new information in each series. Chakraborty, on the other hand believed that part of the reason for improved predictive power with multi rather than univariate series was that simply more similar samples were available. His series were quite highly correlated.

With reference to the description above of the process of embedding, time series that add structure to the picture of the attractor are most likely to help. Mutual Information seems to be a good indicator of this property, especially since MI is independent of the distribution of the compared series.

Now if we choose time series with good MI with respect to the predictee as helper series we can hope for predictions to improve because of the increase in density of points, as reported by Chakraborty. There is another mechanism that can help predictions, predictive power in the helper series themselves. We saw in figure 14 how rapidly auto mutual information can decline when the two copies of the series are offset in time. A given helper time series may record low MI with respect to the predictee when measured with the two series aligned in time, but give better MI for some time offset. It is proposed that helper time series be "vetted" by searching over some range of offsets, both negative and positive against the predictee. Series that can be shown to depend on the predictee can be discarded, but those who have predictive power should be retained. Furthermore, since we attempt to make predictions only one sample ahead, we should align the helper series in time so that the MI maximum corresponds with the prediction.

Nearest Neighbours information

If we have discovered the optimum data representation for a supervised learning algorithm to make predictions, then we have found the data representation that most succinctly characterises a point on a time series, given a particular sampling period.

We can therefore, using the embedding and Separation information, compare points on a time series in a meaningful way. We now have an answer to the question, "what periods in the past of this time series are most similar to the present?". The criterion we have used for determining similarity is Euclidean distance. Since our time series are normalised to the range [0,1] assuming uniformly distributed points we can express the similarity of two points as:

$\frac{1-i}{2}$ where i is the Euclidean distance between two embedded sample points.

Presented at the IEEE World Conference on Computational Intelligence, Orlando, June 94

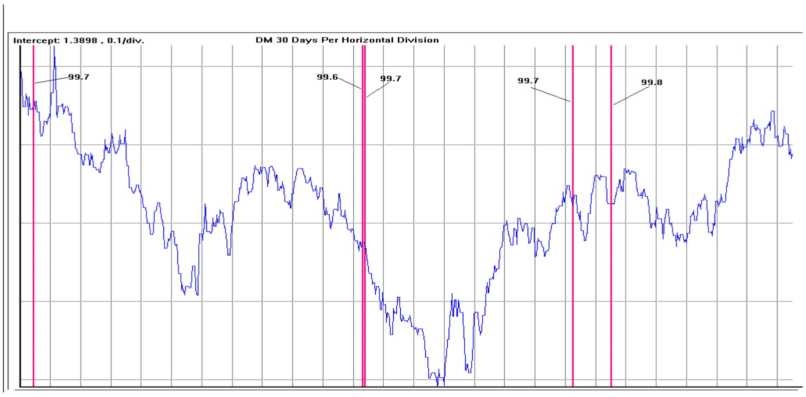


Fig. 6 shows the results of such an analysis where the five most similar embedded vectors on the time series to the present (the far right) are displayed.

Experiments

Five time series were used to test the above, sampled at one minute intervals. They were the spot rates of the DM against the $, the £ against the $, and the 3 month deposit rate for the £,$ and DM.
The samples were taken over a single days trading. The third possibility, the DM against the £, is not included, since we can expect this to keep in line with the other two. These series are widely considered to have strong chaotic components. The graphs of each of these follow:

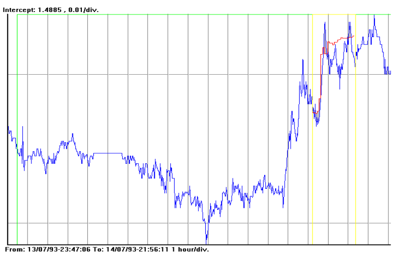


Fig. 7, the DM/$

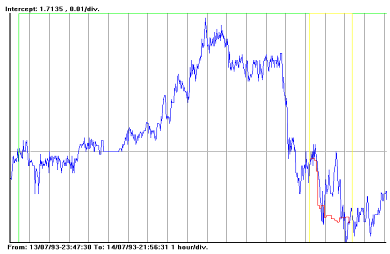


Fig. 8, the £/$

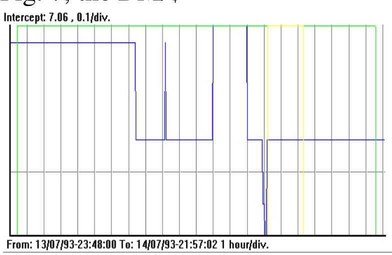


Fig. 9, 3 Month deposit rate, DM

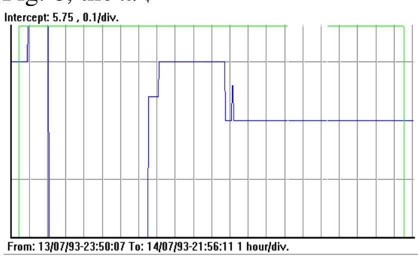


Fig. 10, 3 Month deposit rate, £

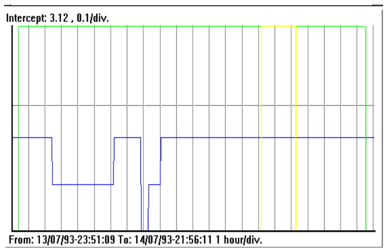


Fig. 11, 3 Month deposit rate, $

The program used to make the evaluations, *The prophet*, was configured to predict both the £ and the DM in 5 minute intervals for a period of two hours. The predictions were iterated, that is the later predictions were based on the previous ones, not on the actual data. The graphs below show the predictions and the actual behaviour on the same graph. The predicted line is the more "steppy" of the two.

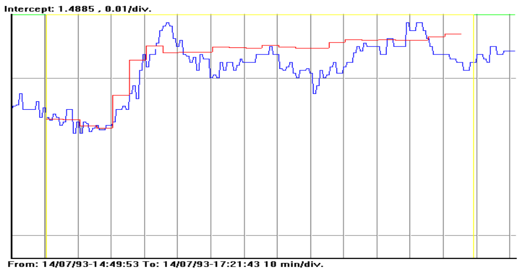


Fig. 12 DM/$ prediction for 2 Hrs in 5 min steps.

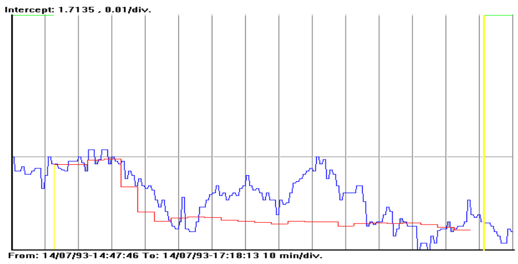


Fig. 13 £/$ prediction for 2 Hrs in 5 min. steps.

The following show the plots of FNN (false nearest neighbour percentage) and AMI (in bits) for different embedding or separation values for each time series:

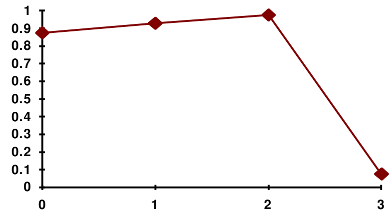


Fig. 14, DM/$ AMI

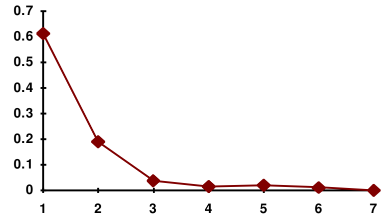


Fig. 15, DM/$ FNN

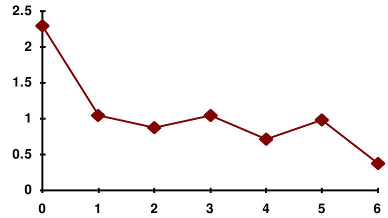


Fig. 16, £/$ AMI

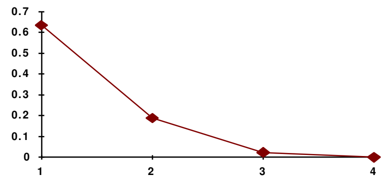


Fig. 17, £/$ FNN

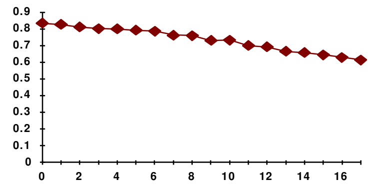


Fig. 18, 3 month DM deposit AMI

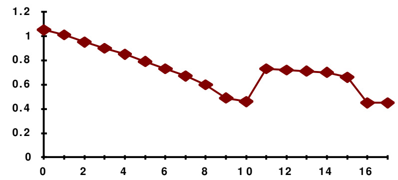


Fig. 19, 3 month £ deposit AMI

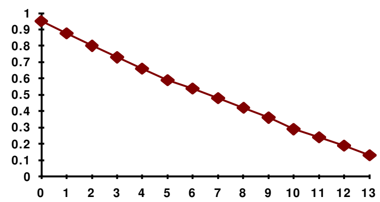


Fig. 20, 3 month $ deposit AMI

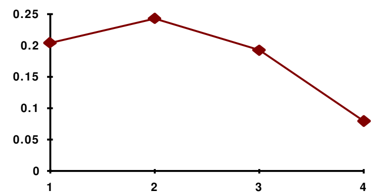


Fig. 21, 3 month $ deposit FNN

The difference between the frequently traded DM and £ and the deposit rates should be clear; the deposit rates have very little variation. The program used to generate AMI and FNN does not go on to perform FNN calculations unless the AMI value is resolved. In two out of the three deposit series the AMI value did not dip sufficiently to be resolved. Looking at the series and the information in them, this is not surprising. Although predictable behaviour may be visible when viewed over days and weeks, at this resolution there is little variation in the data. This is not to say there are no dependencies between the spot rates and the deposit rates, merely that there is no need for embedding to represent these dependencies. The embedding dimension and separation for these series were set to 1.

The software used to make these analyses, *The Prophet*, is limited to a maximum dimension and separation of 16. To exhaustively search for optimum embedding dimensions and separations in combination would require up to 2^{32} trials. The somewhat simpler process of perturbing the optima obtained by our existing methods and observing the effects was used.

The sequence was as follows:

1) The AMI and FNN algorithms were used for each time series.
2) The embedding dimensions and separations suggested were used.
3) Training patterns were generated.

4) The Sugihara and May [Sugihara] algorithm was used to perform predictions.
5) A separate validation period was selected to test the predictive power.
6) Using iterated predictions the performance on predicting the validation set was calculated by comparison to the actual values over the validation period.
7) The R.M.S. error was calculated and subtracted from 1.0, to render a measure of fit.
8) For each helper time series the separation and dimension were perturbed by +/- 1 and +/- 2. Then (2),(3),(4), & (6) where used to evaluate the performance of the whole system.

The Sugihara and May algorithm used above is not a direct copy from their paper. Their algorithm can be split into an embedding part and a table based supervised learning system. Only the latter part, employing nearest neighbour search and interpolation, is used here.The results are shown:

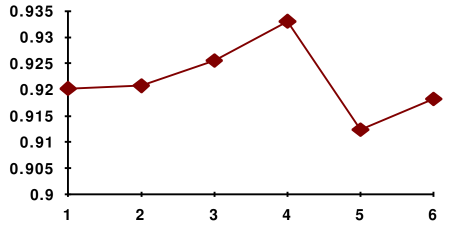


Fig. 22, effects of perturbing DM/$ Dimension

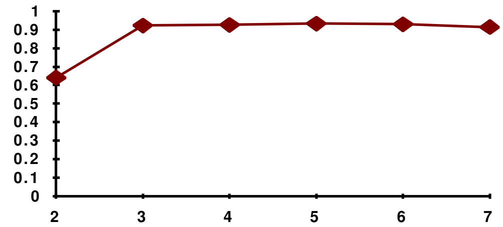


Fig. 23, effects of perturbing DM/$ Separation

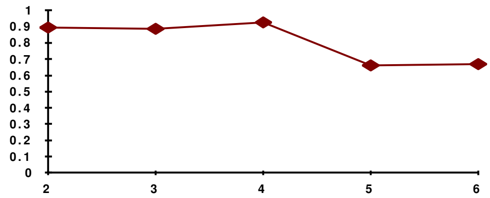


Fig. 24, effects of perturbing £/$ Dimension

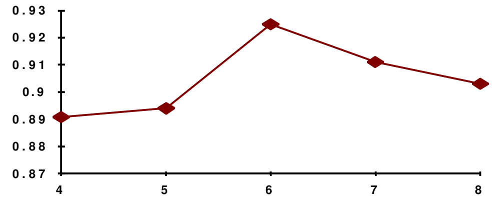


Fig. 25, effects of perturbing £/$ D Separation

In order to further test the efficacy of these methods various other currency pairs were tried. In each case the rate against the American Dollar was used. The Currencies were CAD & CHF, HKD & JPY, NLG & DEM, IEP & JPY, CHF & JPY, CYP & GRD, BEF & DEM, ITL & DEM. The results for perturbation of the dimension gave the peak observed performance 75% of the trials, within 1, 12.5%, and within 2, 12.5%. For separation they gave the peak observed performance 75% of the trials, and were within 1 for the remaining 25%.

Conclusion

There can be no hard and fast conclusions with so little test data. The anecdotal evidence is that AMI and FNN form a strong team for determining embedding dimension and separation, and that those values that are optimal for a series on its own remain optimal when series are used in combination.
The predictive power of *The Prophet* is not an issue here, but the results were good, and interesting in that the performance was exceptional over the first half hour and thereafter increasingly bad. Many mechanisms could be proposed for this, but the chaotic nature of the series has to be the most likely.
The Nearest Neighbour analysis described is an informative and user-friendly technique for displaying, if history is about to repeat itself, what it is that it's about to repeat.

Acknowledgements

Our thanks to David Vanrenen of Equity Systems Limited who first pointed out the diagnostic benefits of the Nearest Neighbours information, and Prof. Reg. King for his support in the preparation of this paper.

References

[Chakraborty] Chakraborty K. Mehrota K., Mohan C., Ranka S. *Forecasting the behaviour of Multivariate time series using Neural Nets*. Neural Networks Vol 5 pp 961-970 1992
[Edmonds1] *The Prophet* time series prediction program. Prophecy Systems Ltd, 34b Market Place, Olney Bucks. MK46 5BU. 1993
[Edmonds2] *Multivariate prediction of financial time series using recent developments in chaos theory.* Proceedings of the 1st international workshop on Neural networks in the capital markets. London Business School, Nov 1993.
[Kennel1] Kennel, M.B., Brown, R., Abarbanel H. *Determining embedding dimension for phase space reconstruction using a geometrical construction*. Physical Review A, Vol 45 No 6 15/03/92 P3403-3411. 1992
[Kennel2] Kennel, M.B. *The Multiple Dimensions Mutual Information Progra*m. Available by anonymous ftp from lyapunov.ucsd.edu. 1992
[Ruelle] Ruelle, D. *Chaotic evolution & strange Attractors*, Cambridge University Press, P28
[Sugihara] Sugihara G., May R. *Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series*. Nature Vol 344 19/4/90 p 734
[Takens] Takens, F. *Dynamical Systems and Turbulence*, Warwick 1980, ed. D.Rand & L.-S. Young, Lecture Notes in Mathematics 898, Springer Verlag.
[Weigend] Weigend A., Rumelhart D., Huberman B. Generalization by weight Elimination with Application to Forecasting. NIPS 3 Morgan Kaufmann p 875-882