FORECASTING, THEORY AND PRACTICE

Director, Technical Forecasts Ltd.
Commercial House
19 Station Road
Bognor Regis PO21 1QD

ChrisSatch@aol.com

What is knowledge? To understand that you know something when you do, and admit that you do not know something when you do not. That is knowledge.

(Sun Tzu, 500BC)

INTRODUCTION

This paper aims to make some new forecasting techniques comprehensible to the widest possible audience and so contains no formulae.

Technically, we aim to provide useful forecasts. The utility of a forecast depends on a user's objectives. Our present work is targeted at users whose decisions can be made on the basis of a general view of whether prices will go up, down or stay about the same. We also have other methods, currently in a research phase, which are arguably more accurate, but which produce less-smooth forecasts from which a general view of price movement is harder to infer. Users requiring limits to price movements (e.g. for setting trading stops) may find those forecasts more appropriate. These examples illustrate the general point that it is difficult to define what is meant by a "good" forecast, without some knowledge of its intended use.

Judging a new forecasting technology can involve some unusual problems. A decision to use some new technology should depend on the advantages it offers over existing methods. With forecasting, there is often a tendency to want to act on one's own hunches – a desire serviced by the betting industry and to a lesser extent the equities market. This can sometimes produce a belief that one's own forecasts are always best; leading to exaggerated claims and a tendency to underestimate the value of other forecasts or forecasting technology

As developers of forecasting methods, we also are subject to the temptation to self-deceive. Consequently, we shall try to spell out both the shortcomings and capabilities of forecasting methods presented in this paper. To assess technology objectively, a potential user should first attempt to define a benchmark for the best forecasting technology they can realistically implement, and then judge the advantages of any new forecasting technology against that benchmark. A decision to implement a new financial forecasting technology will probably always be a choice between imperfect methods – implying that it may be better to do something rather than suffer the indefinite wait for the perfect forecasting method to be developed.

STATISTICAL INSIGHTS NEEDED TO DEVELOP GOOD FORECASTING MODELS

The process of forecasting involves the use of prior experience to predict future values of a time series. This means that any prediction can only be as good as the prior knowledge the forecasting method has seen. Where a forecasting model is developed, some of the data will be used to estimate parameters of the model, which can then be applied to previously unseen (i.e. out of sample) data to test its efficacy. **All forecasts in this paper are generated from out of sample data.**

In accepting that a forecast is possible at all, it is also accepted that the time series from which it is derived must be capable of mathematical representation. We call a mathematical function that generates time series data, a **regression function**. Clues to the nature of that regression function are assumed to come from the time series, whose data may be contaminated by **noise**. When noise is subtracted from time series data, we are left with a **signal**. Therefore, we seek an unknown regression function, from a noise-contaminated sample of data it produced, which may or may not be representative of the full range of data it is capable of producing. Lack of knowledge of whether time series data is representative of all data its regression function can produce is a severe impediment. It means that all forecasts derived without that knowledge need to be qualified with words to the effect “this forecast is derived from past experience, which may not be representative of future price behaviour”.

When we try to model an unknown regression function, without knowledge of its mathematical form, our universal function approximators produce a **network function**, which may be incapable of representing its behaviour. Any difference between network and regression functions is called **the bias.**

Even if there appears to be sufficient representative data, noise contamination can still mean that the sample of data that is available for training may not be representative of average data generated by the regression function. Consequently, there may be a difference between network function outputs for a particular data set, and network function outputs for the average of all data sets produced by the regression function. The square of this difference is called **the variance.**

In the drawings below, the dashed line represents the regression function. If the network function is a straight line (left drawing), it will not represent the regression function well, resulting in high bias. Network functions, calculated from different (noise contaminated) data samples generated by the regression function, are unlikely to change much, so the variance will be low.

If the network function is made up of a series of straight lines joining data points of a specific sample (right drawing), it should represent the regression function better, resulting in lower bias. When a new sample of data is used and a new network function calculated in the same way, they will be likely to differ considerably, resulting in high variance.

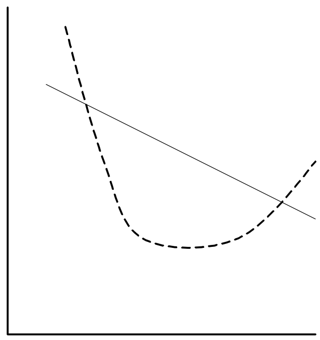


Fig 1 High Bias/ Low Variance

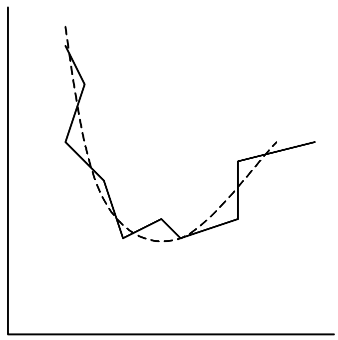


Fig 2 Low Bias/ High Variance

A fuller account of bias and variance can be found in (1). Bias and variance provide insights into the problems of modelling data, but no precise advice on how to do it. General advice on this point comes from a result known as Occam’s Razor, which can be paraphrased by saying that the simplest model capable of representing data, is usually the best. That general advice pre-supposes knowledge of signal and noise components of the time series and the appropriate level of complexity for a network function. Clearly, a highly parameterised network function will allow enormous freedom to represent the time series and can do so very accurately, representing both its signal and noise components. The problem with that is that its predictive abilities are impaired because so much parametric capital has been used to represent <u>both</u> a repeatable signal <u>and</u> the unrepeatable noise contaminating it. Occam’s Razor tells us to find a simple network function, sufficient only to represent the signal component of the time series. That ideal may also be achieved by averaging multiple models, each one of which is over-fitted (exemplified by Fig 2), but where the errors associated with fitted noise self-cancel. This idea is known as a committee of networks, and is one that we employ in our non-linear models.

One consequence of Occam’s Razor is a dilemma between explanation and prediction. Most people like explanations and dislike “black boxes”. To produce a forecasting model, with explanations, means formulating a model that is capable of representing those explanations, initially without knowledge of their relevance. The greater the number of potential explanations modelled, the greater the number of independent parameters that are needed, and the greater the noise that can also be modelled. Occam’s razor tells us to use the simplest possible model if we want good predictions. Without some incredible luck, the simplest possible model has to be found as part of a systematic data modelling process, rather than something that models pre-conceived explanations. In practice, there are very few reliable non-linear data modeling techniques that are both usable and well understood. Those that exist have tended to be developed to model arbitrary data rather than explain economic phenomena. The result of this is that a good explanatory model will usually predict poorly, whereas a good predictive model will usually have very limited powers of explanation. For this we coin the phrase of "explanation/ prediction dilemma". It implies that if we want good predictions we have to learn to live with black boxes.

DETERMINISTIC OR STOCHASTIC FORECASTS?

Our methods of predicting time series are deterministic, i.e. we assume the existence of a signal that is capable of being forecast. In reality, financial time series have noise, which could be described by forecasting a sequence of probability density functions, but which we are ignoring when making our forecasts. In addition, there will usually be uncertainty both about the best way of modelling the data, and about which of several different candidate models is most appropriate in different circumstances. This question can be partially addressed by looking at forecasts from several different, but similarly plausible models, and observing the divergence in their results. This is a grey area in the deterministic/stochastic question but one which demands a deterministic modelling capability. We have followed a deterministic approach for a number of other reasons. Firstly, the market we are aiming at is more used to deterministic technical indicators than stochastic, and should find it easier to incorporate deterministic forecasts into their (technical) decision making. Secondly, financial decision-makers are likely to find it easier to fuse "soft" information (e.g. impressions gained from reading news reports) into a decision-making process based on deterministic indicators. Thirdly, deterministic results from different models and starting points can be used to obtain an impression of the reliability or otherwise of the forecasting process, as well as providing valuable visual insights into the nature of the markets. Some of these insights will be described later.

LEADING INDICATORS

Where two time series are related, and movement in one of those series precedes movement in the other, the technique of leading indicators can be used. This is particularly useful for short time series, which are known to be related to longer, leading time series. The availability of a long leading time series means that less data is required to analyze the lagging series, when compared with other methods.

The first step is to examine either correlation or mutual information between a lagging series and candidate leading indicators. This determines whether a leading indicator is present, if so the best one, and what the lag is between it and the series of interest. Various windows of values from the two series are then chosen and used as inputs for a non-linear mapping with future values of the lagging time series. There is a great deal of trial and error in the process of finding window sizes to produce the best mapping. Our experience with this technique has been mostly in the property field where it has produced some very useful forecasts, (2) from relatively short time series. Figs 6, 7 and 8 show a number of leading indicator forecasts, which are appropriately captioned to highlight our experience with the method.

DYNAMICAL SYSTEMS CONCEPTS

Figs 3 -- 4 show how time series phase portraits are produced. The essential point to be inferred from them is that the behaviour of a complex time series may be capable of being represented by consistent trajectories forming a multi-dimensional manifold. Where those trajectories are both consistent and capable of being modelled, future values of a time series can be predicted from a combination of current and previous values. This provides a theoretical basis for developing predictive models from a window (delay vector) of previous time series values. There is a mathematical term associated with this concept, known as an embedding, and we sometimes refer to this work as embedding technology.

Given a delay vector (i.e. a window of time series values) our aim is to find a network function to model future values. That can be done using the delay vectors themselves, or by transforming them into some other space, doing predictions in that space and transforming those predictions back into the original space. The advantage of the latter method is that some transformations, such as a Fourier Transform, facilitate the removal of noise from delay vectors, prior to their use in the development of a network function. For reasons previously outlined, we only attempt to model the signal component of the delay vectors. Our modeling process is therefore:

1. Choose a space in which to do the predictions.
2. Transform delay vectors into that space, removing any noise. (This means that the original time series is converted into a multiple time series, corresponding to the transformations of successive delay vectors.)

3. Develop a predictive model in the transformed space.

Given a novel delay vector, the procedure for getting a prediction is:

1. Convert the delay vector into the transformed space.
2. Use the predictive model to obtain forecast in the transformed space.
3. Convert the forecast back into the original space.

The ability to choose the space in which to do the predictions enables the best possible match to be made between a modelling technique, and inputs and targets for which a mapping is needed. It was stated earlier that our options for effective non-linear modelling techniques are limited, so choosing an optimal space in which to develop the model can lead to improvements in results. An overview of the transformation, modelling and predictive methods is shown as a flowchart in Fig 5. Background to non-linear modelling techniques can be found in (1). Mostly, we use Radial Basis Function networks developed from work described in (3).

EXPERIENCE WITH MODELLING DYNAMICAL SYSTEMS

Dynamical systems theory suggests the existence of a regression function, related to a window size, which produces the price time series. If such a regression function exists, it should be capable of modelling prices and should predict significantly better than all other window sizes. Examination of RMS prediction errors, for various window sizes and forecast horizons, fails to reveal the existence of any window size that is universally best. This may be due to deficiencies in the modelling process or may simply mean that the dynamical systems concept is inappropriate for financial time series. There are a number of significantly different window sizes whose corresponding models result in significantly lower prediction errors than those of their immediate neighbours. This leads us to believe there are a number of similarly plausible models, which at times all predict fairly well, but which at other times produce very different forecasts. Consistency in the forecasts produced by different models is a very useful property for assessing their reliability.

Forecasts are sensitive to the point from which they are made as well as the window size used in the model from which they are produced. It is possible, sometimes even easy, to produce a forecast that can subsequently be shown to be excellent. It is much more difficult to know that a forecast is excellent at the time it is made. Some idea of the accuracy of a forecast can be obtained by comparing those produced from models of different window sizes, possibly starting from different dates. Where a large-window forecast corroborates a small-window forecast, possibly starting from an earlier date, the result is likely (but not certain) to provide a good forecast. Figs 10 and 11 show examples of this for UK Gilts forecasts.

The result is that on some financial instruments, some of the time, good forecasts can be produced and, through corroboration with other forecasts, an idea obtained of their quality at the time they are produced. There is a conflict between the number of forecasts that are usually needed and the degree of corroboration that is simultaneously needed. For most starting points in a time series, good corroboration between forecasts will usually be absent, but those starting points may well already be covered by corroborated forecasts from a previous starting point. Inevitably, there will be occasions when forecasts are required for periods for which corroborated forecasts just do not exist. If forecasts are used in these periods, inaccuracies must be expected and appropriate contingency plans available to cope with their consequences.

Our experience with this forecasting method has led to various procedures and observations:

1. We examine forecasts from a number of different models, possibly using different starting points, and look for corroborated results.

2. In a sideways market, results of different forecasts may appear to meander aimlessly. This can be taken as an indication that a market is either in, or about to enter, a sideways period. (Fig 11)

3. During a correction in a trend, either the original trend or a sideways market may be forecast.

4. After a shock or anomaly in the market, forecasts tend to be unreliable.

5. The validity of any single forecasting model can change rapidly – even after a lengthy period in which it has consistently out-performed all other forecasting models.

6. After forecasting a turning point, the remainder of a forecast becomes less reliable.

7. The latest starting point does not always produce the best forecast, and frequently does not.

8. The extent to which these observations apply varies for different financial instruments. This means that no single set of trading rules, based on this forecasting technology, is likely to be applicable to all financial instruments.

USING FORECASTS FOR FINANCIAL DECISION MAKING

Sun Tzu's definition of knowledge: *"To understand that you know something when you do, and admit that you do not know something when you do not"* is an ideal to which we would like to aspire, but which the vagaries of the marketplace inhibit. Our knowledge of the behaviour of our forecasts, as summarised previously, is fuzzy and decisions will need to be made against that background of uncertainty. A realistic expectation of our forecasting is to reduce, but not remove, the uncertainty associated with financial decisions.

In the case of property price forecasts (Figs 6, 7 and 8) we see why we need to emphasize that all forecasts are developed from previous market experience. Most of the time, leading-indicator forecasts provide some indication of an impending move. Sometimes a turning point is missed, but the move is confirmed with the next forecast. Markets occasionally behave out of character, turning points are not confirmed and forecasts are "wrong." In these cases, non-confirmations of price moves suggest the market is anomalous. Usually, anomalous moves are short lived and, where they are not confirmed by our forecasts, tend to peter out or correct rapidly, often presenting major trading opportunities. Here we have the dilemma of an inaccurate forecast providing the clue to an impending move, showing that the usefulness of a forecast lies in its correct interpretation as well as its accuracy.

It is easy to be seduced by the technical elegance of dynamical systems theory, but where leading indicators exist, their simpler technology tends to give a more reliable result and is our preferred forecasting method. Where a leading indicator can be forecast by our embedding technology, and that forecast is corroborated, it can be used to generate a considerably longer forecast for the lagging time series. In practice this leaves us with a consistent ability to forecast a lagging series to the extent of the lag, and an occasional capability to forecast a lagging series to the extent of the lag plus the period for which a corroborated forecast of the leading indicator exists. Mostly, these forecasts are used for property investment decisions, either by advisors or investors.

For the majority of financial time series, few, if any useful leading indicators are available and we have to analyze them with our single time series embedding technology. At this point it is important to review what these results are: a number of similarly plausible forecasts, based on previous market experience, for different window sizes. The next question is how to make the best use of them? For a hedger or futures trader, the primary question is will prices move? If so, how much? Earlier, we mentioned that an impression of price movement was easier to infer from a smooth rather than jagged line, but that inference had to be traded for accuracy. Our present system produces smooth lines, whose accuracy may be improved later. For an options trader wanting to sell calls above the market and puts below, jagged-line forecasts are important for strike price selection. Similarly, a futures trader may need the same information to set MIT (market if touched) or stop orders. We envisage requirements for both regularized and unregularized forecasts. For the moment, we present regularized results aimed at the futures' traders' perspective; i.e. wondering whether to enter or exit a market.

The procedure for inferring a price move is to see if all the forecasts agree on the move. When they do, and there is other net evidence to support them, a move is considered likely and a trader might take a position. Figs 9 and 10 show examples of such cases.

During a sideways market, trend-following systems over-trade and lose money. The onset of a sideways market is indicated by inconsistent forecasts, mirroring its own confusion. Forecast inconsistencies tend to remain throughout sideways periods, with consistency returning at their end to indicate both that a position should be taken and its direction. Figs 11 and 12 show the system's progress through a sideways market.

When, in a trend, prices reverse, there is a question of whether their movement indicates a trend reversal or correction. Simple trend-following trading systems tend to get sucked into loss-making trades during trend corrections. With multiple forecasts, the key to making such a decision comes from noting what the predictions say and when they say it. Figs 13 - 16 show the system's progress through a trend correction. Note that forecasts never confirm the correction as a trend reversal and forecast the continuation of the trend at the right time.

We mentioned earlier that we do a signal/noise decomposition to obtain a filtered version of closing prices. The combination of filtered prices and noise estimates can be used to help place stops or MIT orders. A labelled diagram indicating these features is shown in Fig 17, where forecasts are suggesting an upturn in the market is about to occur. That upturn happens, and a stop can be progressively raised underneath price, but the move eventually peters out, as indicated by the forecasts in Fig 18. When the justification for staying in a position disappears, the filtered line and noise level can be used to assess a level for an MIT order to exit the position. That tends to be a better strategy than allowing a stop to be triggered, as the exit should occur at a higher price and at a time when the market is moving with the trade, lessening slippage.

Our program is currently developed to be an aid to discretionary financial decisions, but may be further developed in the future. In order to gain experience in using our forecasts, trading situations need to be simulated. That means that forecasts and only that data a trader would have in real life can be shown, simulated decisions made from that information, and financial consequences computed. Figs 17 and 18 show the program's display in this simulated trading mode. During simulated trading, the number of bad decisions usually reduces, as the trader becomes more familiar with using the forecasts and combining them successfully with other information.

CONCLUSIONS

In this paper, we have addressed the statistical background to forecasting involving bias, variance, Occam's Razor, and have explained how these considerations have led us to identify the explanation/prediction dilemma.

A forecast can only be as representative as the data from which it is derived, which leads us to caution against relying on forecasts derived from short time series. For short time series, we advise trying the leading indicator approach, rather than embeddology or any other approach based on a single time series.

We have made the point that "bad" forecasts occur during periods of anomalous market behaviour, and they can signal the likelihood of a significant correction and trading opportunity.

With sufficient data, a dynamical systems approach seems to be capable of modelling the market some of the time. When forecasts from two or more such models, of differing but appropriate window sizes, possibly starting from different dates, agree in their results, our experience suggests that a predicted move will occur. This has led us to the idea of corroborated forecasts, derived from models using different window sizes and starting from different dates.

When forecasts disagree, and appear to meander sideways, a sideways market is likely. The futures' trader may wish to stay out of this market.

To trade with all of these forecasts at present, subjective interpretation and the possible incorporation of other data is needed. As an example, for property forecasts, we advise being extremely sensitive to any hint of a turning point, and prepared for it to come sooner and sharper than predicted.

Filtered prices and noise estimates have an important part to play in developing an optimal trading strategy and help overcome many of the inherent limitations of deterministic forecasts.

Markets cannot be forecast reliably from all starting dates, but good forecasts appear possible from some starting dates. Use of the latest information may not lead to the best forecast, and frequently does not. The practical consequences of this for trading are that entries can be made from times when forecasts appear reliable, but exits may have to be made at times when forecasts are unreliable. A stop loss policy is therefore advocated for all trading based on these forecasts, and a higher level of subjective appraisal of forecasts may be needed for exits. Put simply, good exits are harder to make than good entries.

The training of users of financial forecasts is therefore a serious issue, which usually needs to co-exist with a requirement for commercial secrecy. To allow decision-makers to become familiar with our results, we have developed a program where forecasts, filtered prices and noise estimates, can be displayed as they would be in live trading, but on historical and out of sample data. Financial decisions can then be taken and their consequences evaluated. This simulated trading is one way of allowing users to become familiar with using forecasts in combination with other information relevant to financial decisions.

Finally, we come back to a point made in the introduction. A decision to implement a new financial forecasting technology will probably always be a choice between imperfect methods; hence the need for a benchmark. The issue is whether a new forecasting technology offers sufficient advantages over the benchmark to justify its implementation.

ACKNOWLEDGEMENT

Approaches to forecasting described in this paper have been developed with the active help of my co-directors, Prof. D. Lowe and Mr. H. James. Their assistance is gratefully acknowledged.

REFERENCES

1. Bishop C. M. (1995) “Neural Networks for Pattern Recognition” Oxford University Press pp164-171, 332-342.

2. Connellan, O.P. and James, H. (1998) “Estimated Realisation Price (ERP) by Neural Networks”, Jol. of Property Valuation and Investment, MCB University Press, Vol 16, No. 1, pp71 – 86.

3. Broomhead, D. and Lowe, D. (1988) “Multivariable Functional Interpolation and Adaptive Networks”, Jol of Complex Systems, Vol. 2 Issue 3, pp321-355.

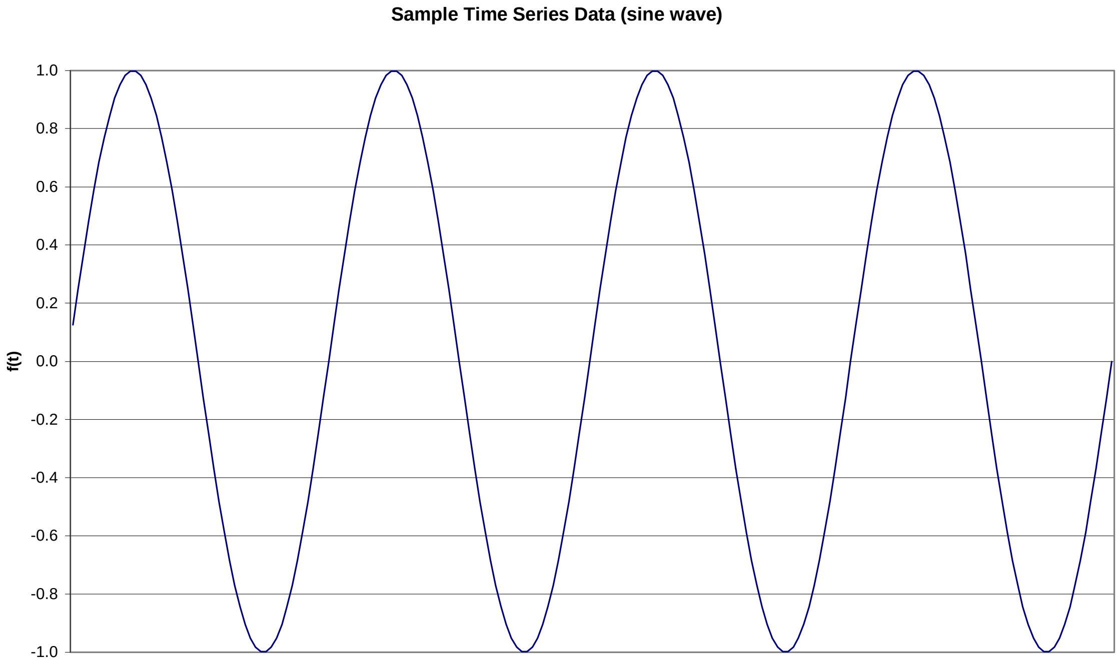


Fig 3a: To illustrate a principle, we use the example of very simple sine wave data.

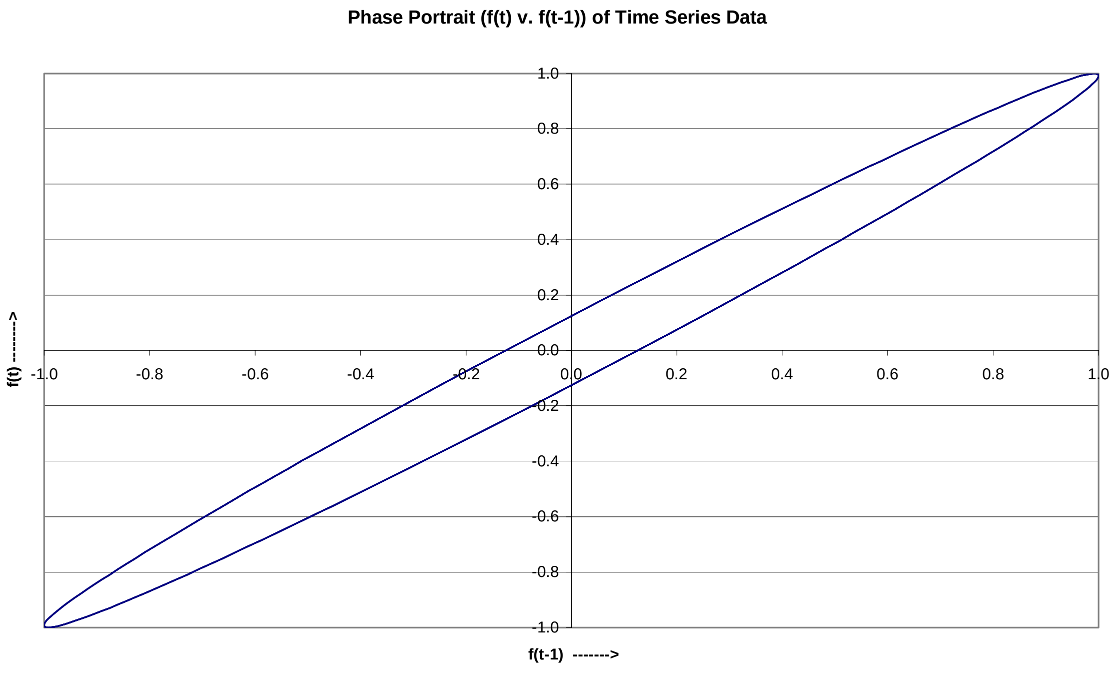


Fig 3b: A phase portrait is simply a plot of a current value, f(t), against the preceding value, f(t-1). For this time series, the result is a regular ellipse. Successive elements of the time series describe trajectories around this ellipse. Consequently a knowledge of this ellipse's properties is sufficient to predict f(t+1) from f(t).

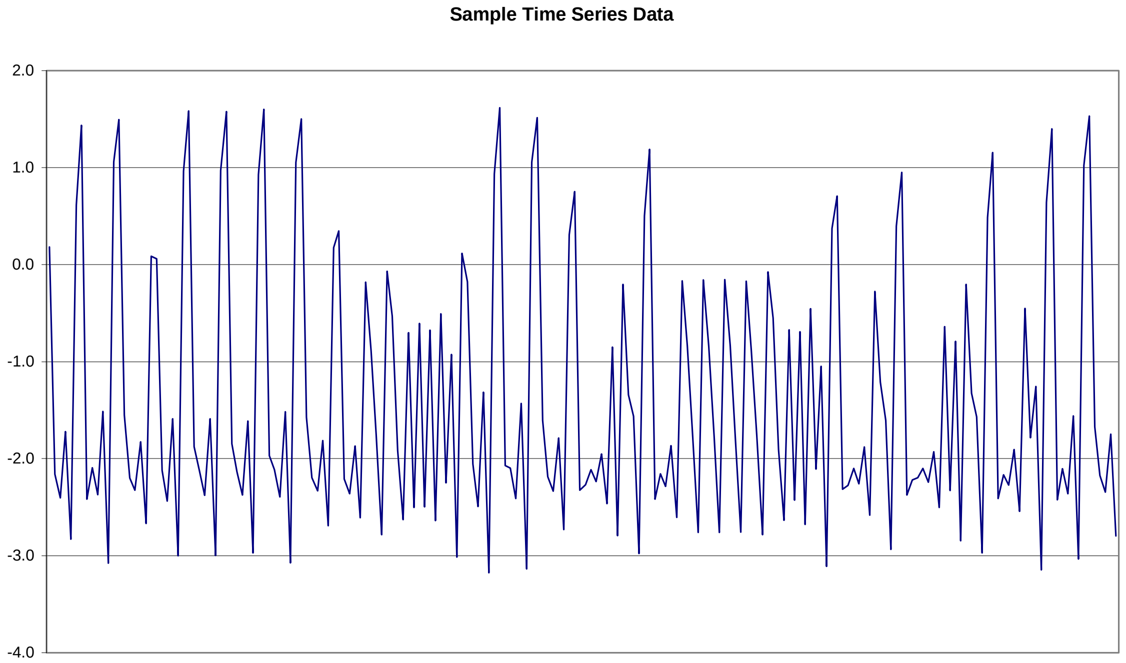


Fig 4a. This time series (f(t) plotted against time) is more complex than that shown in Fig 6.

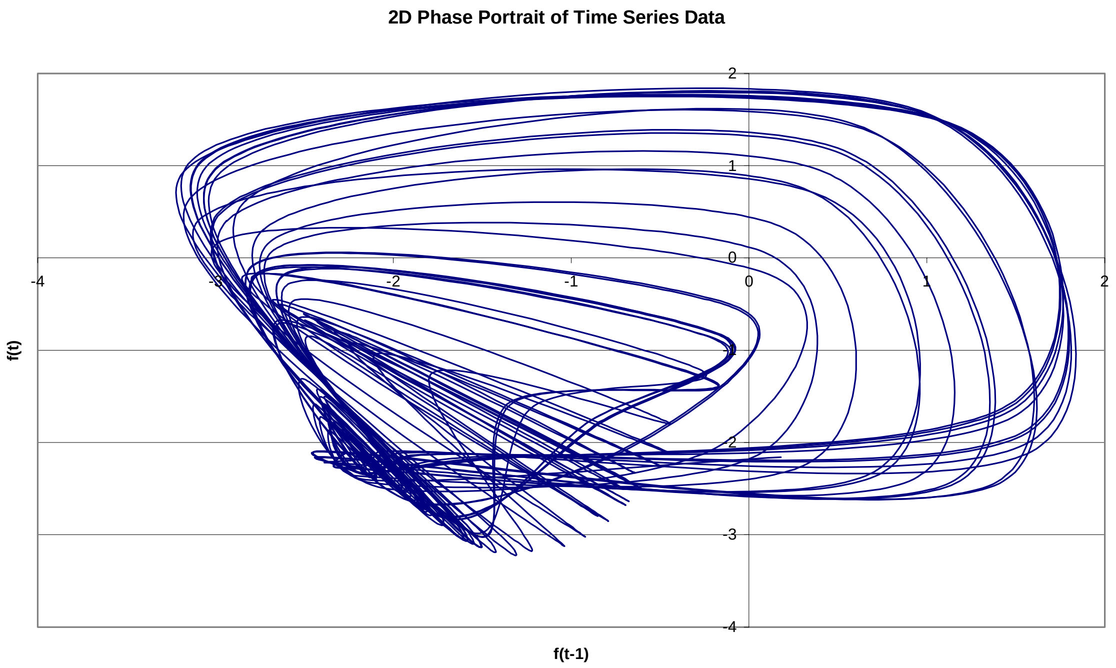
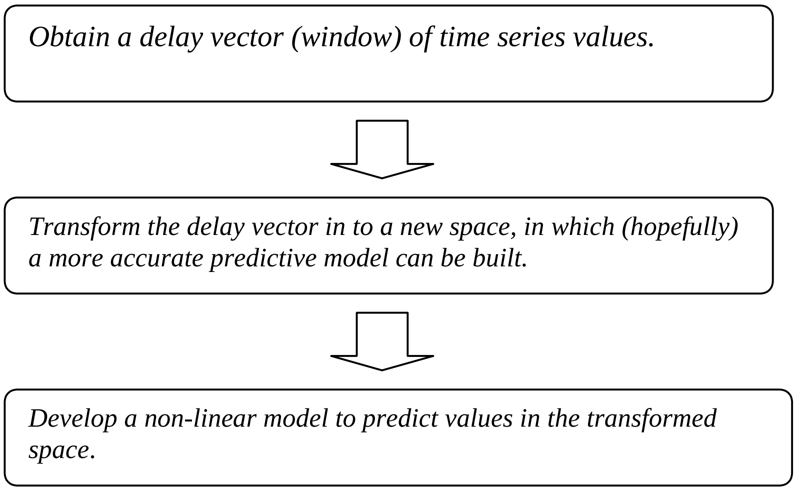


Fig 4b. The phase portrait of this time series (Fig 8) involves intersecting trajectories. Each intersection represents an ambiguity in a two dimensional predictive model. If successive time series values are plotted in a higher dimensional space, the intersections disappear. To visualize this, imagine these trajectories represent a plan view of a bent and distorted coiled spring, where the coils are not touching. The implication is that this time series can be predicted if an appropriate size can be found for delay vectors and a representative network function calculated.

1. Modelling



2. Prediction

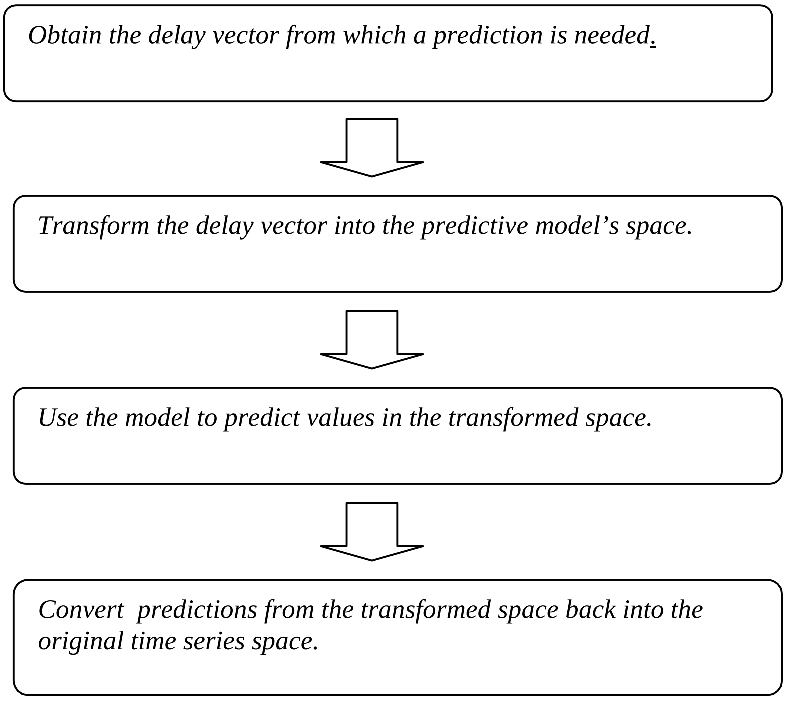


Fig 5 -- Flow chart, showing the use of transformations for modelling and prediction.