Chapter 1

INFORMATION FUSION APPLIED TO SELECTED FINANCIAL PROBLEM DOMAINS.

David Lowe
Aston University, Aston Triangle, Birmingham B4 7ET, U.K.
d.lowe@aston.ac.uk

Abstract

Financial analysis is a domain characterised by multivariate sources which are noisy, nonstationary, and usually nonlinear. The financial application domain has a strong requirement for good methods of data and information fusion, especially with increasing deregularisation, globalisation and automation. Global mass access to financial data via developments in internet technology will have a major driving impact on the way financial markets will be driven in the future. Given the accepted difficulty of problems in financial domains, it is surprising that even simple methods of data fusion analysis have not been widely adopted in this community. In this chapter we briefly review some of the levels in the data fusion hierarchy, discuss their relevance to different types of financial data problems and demonstrate the benefits, both theoretically and experimentally, that can be achieved in just a few selected topics.

1. INTRODUCTION

Technical financial analysis is an ideal domain for the development of data fusion techniques. There exists a wide base of diverse application requirements such as: the analysis of consumer credit, financial economics, options pricing, portfolio analysis, yield curves, risk management, strategic investment and trading models. Across this broad spectrum is a technical drive to incorporate methods of extracting *information* from *data*. Financial analysis invariably has to contend with multiple sources of data (news reports, company accounts, market data, mulitvariate time series), on different time scales (quarterly government indicators vs foreign exchange tick data) and with differing levels of accuracy from the qualitative to the quantitative. Hence it is an ideal medium for the development of pattern recognition and information fusion methods.

There are various frameworks for data fusion hierarchies. This book tends to concentrate upon the JDL model derived from military requirements. The UK Technology Foresight Working Party on Data Fusion [2] provided a 6-level taxonomy of generic data fusion, provided from a non-military-centric starting point. The levels in this taxonomy proceeding from the low level data processing to the high level symbolic processing were defined as Sensing, Signal Processing, Data Feature Extraction, Pattern Processing, Situation Assessment, and finally Decision Making. In the context of financial analysis we can also recognise a mapping of each of these levels to a specific requirement in a financial application. 'Sensing' implies the collation and trawl for relevant raw data sources. 'Signal Processing' includes the obvious basic preprocessing, labelling, enumerating and construction of the data sources into usable data streams. 'Data Feature Extraction' in finance would correspond to tasks such as selection of relevant features and specific data selection such as removal of obvious outliers. 'Pattern Processing' would include the manipulation of data valid streams for the purpose of visualisation, rescaling, renormalisation, SnR enhancement, forecasting modelling, and transformations into more appropriate structures such as wavelet spaces. 'Situation Assessment' would be the level at which various models would be combined with extra knowledge such as levels of risk an institution is prepared to manage, or obtaining measures of uncertainty in markets or other economic arenas, or optimisation of portfolios of investments according to different what-if criteria. The final layer of decision making would employ different types of utility function, combined with the various models' estimates of the financial instrument under study. Hence, it is easy to see how a generic financial application domain can be mapped on to the Data-To-Decision cascade.

Despite this link to an obvious developed activity, the exploitation of fundamental data fusion techniques remains almost non existent in real financial applications. In the ASI meeting itself, I discussed 6 different financial applications spread across the different levels of the fusion hierarchy, and I demonstrated how simple applications of fusion techniques could return significant performance improvements over the more traditional approaches to these problems. In the talk I considered fusion of sources, features and models, regression, optimisation, prediction and classification, and used examples from option pricing, portfolio construction, managed pension fund selection, short and long term prediction of US bonds, FX ($/DM) data and equities (banks and electricals sector) from the UK markets. Within this diversity we observed a generic performance improvement, which makes the lack of the routine application of these methods all the more surprising.

In this chapter I will expand on just some of the approaches and results discussed at the ASI meeting itself. All examples used are based on real data obtained from collaborations with companies within various financial sectors, to whom I am most grateful for the additional comments and provision of background knowledge in addition to access to data. I also discuss a couple of interesting mathematical results appropriate to fusion domains which do not seem to be widely appreciated across the community, but give support for the use of multiple *simple* models, rather than sophisticated models.

Technical financial analysis however has its own specific problems which create difficulties associated with traditional approaches to data fusion. We recognise that financial data is charaterised by the three 'N's': Noise (and specificially non-Gaussian invariably), Nonlinearity (and hence modelling capabilities are complicated - so the Kalman Filter approach is known to be inappropriate for example), and Nonstationarity (so again we should be aware of the tradeoff between stability and plasticity in constructing short-time stationary models for example). These difficulties both constrain and motivate novel developments.

2. FUSION OF SOURCES

We begin by considering a medium-level fusion problem: that of how to deal with multiple sources when those sources are subject to ignorance of specific form. At this level we will have already performed sensing, signal processing and data feature extraction. This is a very common situation in finance. Finance intrinsically deals with multivariate time series, multiple time scale values, and different information sources. Since we are not likely to have good knowledge of the generator which produces the sources of financial time series, we inevitably have to deal with uncertainty. This implies that we should employ a probabilistic approach to looking at the issues of source fusion under uncertainty. Hence , we first consider the mathematics of simple probabilistic approaches to combining inputs. We will see an interesting consequence on decision-making by incorporating an uncertainty of the source models into the analysis.

2.1 SIMPLE COMBINATION STRATEGIES FOR CLASSIFICATION

Ideally in finance we would want to estimate the probability distribution of a given quantity conditioned jointly upon a set of different sources of information, $P(\theta = \omega | x_1, x_2, \ldots, x_S)$. Given this distribution, we would then make decisions based on a Bayesian decision rule or incorporate measures of utility to bias our decisions. Unfortunately we cannot generally estimate such a distribution in practice without severe approximations on the model form. Instead it is common to make assumptions on the statistical relationships between the information sources, and use the simpler models derived from the individual information sources. One usual method is to assume conditional independence between inputs (which could be multiple data sources or multiple models). We follow through this common assumption and also discuss an even simpler model. In practice, we often find that simpler data fusion models which are not as well 'principled' as the more rigorous approximtions usually are more reliable in real world problems. The reason for this will become apparent in the analysis. We follow an elegant approach which has recently been discussed in [5] in the context of multisensor person identification, and we expand on that approach in what follows.

Let us assume that we want to obtain an estimate of a posterior probability of an event ω_k (buy/sell signal, or stock k is likely to gain most) occurring given a pattern vector $\boldsymbol{x}$. The Bayesian approach would be to choose the event which maximises the posterior distribution. Under the independence assumption we have:–

$$P(\omega_k|x_1,\ldots,x_S) = \frac{P(\omega_k)\prod_{i=1}^{S} p(x_i|\omega_k)}{\sum_{k=1}^{S} P(\omega_k)\prod_{i=1}^{S} p(x_i|\omega_k)}$$

So the decision rule would be to choose the event which maximises

$$max_{k=1}^{K} P(\omega_k)\prod_{i=1}^{S} p(x_i|\omega_k)$$

In terms of posterior probabilities produced by separate source models, using $p(\omega_k|x_i) = p(x_i|\omega_k)P(\omega_k)/P(x_i)$, gives the decision as

$$max_{k=1}^{K} P(\omega_k)\prod_{i=1}^{S}\left[\frac{P(x_i)}{p(\omega_k|x_i)}P(\omega_k)\right]$$

$$= max_{k=1}^{S} P^{1-S}(\omega_k)\prod_{i=1}^{S} P(x_i)p(\omega_k|x_i)$$

This is the **Product Rule**.

We will assume that $P(x_i) = const = 1/S$. i.e we have no prior belief to select one of the sources as more important than any other.

Note an important point about this 'product rule' for combining information sources: it is very severe since an event may be completely inhibited if just one of the independent source models returns an almost zero value. It also assumes we know, accurately, what the individual probability distributions of each source model is. This is most unlikely. Hence we have to assume the existence of uncertainty on the values of these source models. We can perform an error sensitivity on this product rule by assuming that the *estimated* aposterior probability $\hat{p}(\omega_k|x_i)$ is close to the desired value $p(\omega_k|x_i)$ by writing $\hat{p}(\omega_k|x_i) = p(\omega_k|x_i) + \epsilon_{ki}$. Substituting and linearising around P assuming that the error is small, we find

$$max_{k=1}^{S} P^{1-S}(\omega_k) \prod_{i=1}^{S} P(x_i)\hat{p}(\omega_k|x_i)$$

$$\approx max_{k=1}^{S} P^{1-S}(\omega_k) \prod_{i=1}^{S} P(x_i)p(\omega_k|x_i) \left[1 + \frac{\epsilon_{ki}}{p(\omega_k|x_i)}\right]$$

For small ϵ the product components may be separated as

$$\approx max_{k=1}^{S} P^{1-S}(\omega_k) \left[\prod_{i=1}^{S} P(x_i)p(\omega_k|x_i)\right] \left[1 + \sum_i \frac{\epsilon_{ki}}{p(\omega_k|x_i)}\right]$$

We see from this that the effect of allowing for uncertainty in the source distributions is to allow for a modulation of the decision rule by the term

$$1 + \sum_{i=1}^{S} \frac{\epsilon_{ki}}{P(\omega_k|x_i)}$$

It is now obvious that whenever a source model returns a very small value for the likelihood of an event, then that particular error in estimation is magnified.

Interestingly we can introduce a robustness into the decision making by incorporating a more severe simplification than the independence assumption. If we now assume that the posterior does not really deviate much from the prior of the event, so that $P(\omega_k|x_i) \approx P(\omega_k)[1 + \delta_{ki}]$ where δ_{ki} is assumed a small perturbation, then we find that the decision rule can be expressed as :–

$$max_{k=1}^{S} P^{1-S}(\omega_k) \prod_{i=1}^{S} P(x_i)p(\omega_k)[1 + \delta_{ki}]$$

$$\approx max_{k=1}^{S} P(\omega_k)[1 + \sum_i \delta_{ki}]$$

(assuming that $P(x_i) = const$). Substituting in for δ gives:–

$$\approx max_{k=1}^{S} P(\omega_k)[1 + \sum_i \left\{\frac{p(\omega_k|x_i)}{P(\omega_k)} - 1\right\}]$$

or,

$$max_{k=1}^{S} \left\{(1 - S)P(\omega_k) + \sum_{i=1}^{S} p(\omega_k|x_i)\right\}$$

This is now a *sum rule* for decision making. Again it assumes that all source probabilities are known accurately. As before, we can examine the effect of this assumption on the decision making process by performing an error analysis. Assuming a small error of ϵ_{ki} in $p(\omega_k|x_i)$, the sum rule becomes:–

$$max_{k=1}^{S}\left\{(1-S)P(\omega_k)+\sum_{i=1}^{S}p(\omega_k|x_i)[1+\frac{\sum_i \epsilon_{ki}}{\sum_i p(\omega_k|x_i)}]\right\}$$

This expression shows that the decision rule is modulated by the error term of the form:

$$1+\frac{\sum_{i=1}^{S}\epsilon_{ki}}{\sum_{i=1}^{S}p(\omega_k|x_i)}$$

Now note that *all* source probability models have to return an almost zero value to distort the error. Generally the sum over all sources of the probabilities will be greater than unity and so this will have the effect of damping down the errors. Hence, even though this is a much simpler model for source fusion, it is more robust and reliable in practice since we can never approximate the true distribution function with any degree of accuracy, and hence the decision making is usually dominated by the sensitivity to errors. We will observe this effect when we give an example of combining models later.

2.2 PRICING OF FINANCIAL OPTIONS CONTRACTS

In the previous section we discussed the consequences of the statistical distribution of separate source models. An alternative strategy occurs when there exists a deterministic (though inevitably nonlinear) relationship linking together the various information sources. The financial application domain we use is a very simple example which may be posed as a straightforward nonlinear multivariate regression problem. It is complicated by the fact that one of the required input variables cannot be measured, but must be inferred from other sources. The example considered is a problem based on the pricing of financial instruments known as *Options*. An option on a commodity is a contract which gives the owner of the contract the right (but not the obligation) to either buy or sell the underlying commodity at a predetermined price at some time in the future. The price in the contract is known as the *exercise price* or the *strike price*. The date in the contract is known as the *expiration date*, *exercise date* or *maturity*. A contract giving the owner the right to buy by a certain date is known as a ‘Call’, and the right to sell is known as a ‘Put’ options contract. If the contract can only be exchanged at the expiry date of the contract, then it is known as a *European* option. If the contract can be exercised at any time up to the expiry date it is known as an *American* option. Options on stocks were first traded on an organised exchange in 1973, since when there has been a dramatic growth in the options markets. The underlying assets of the contracts now include stocks, stock indices, currencies, debt, commodities and futures contracts. Although often used for balancing risk by constructing portfolios, options can be used for speculation due to their inbuilt gearing of capital. When used to speculate it is possible to gain significantly but also lose all the initial investment. Hence mispricing or incorrectly estimating the value of an options contract can have significant consequences, as the 1995 saga of the Baring Bank has emphatically illustrated.

This particular area of financial mathematics is of specific interest, as there exists a widely accepted model which estimates the price of European call or put options by an analytic formula which involves five variables. The model is known as the Black–Scholes pricing formula (these proponents were recently awarded the Nobel Prize for Economics for this work) and was derived in 1973 [1]. The model makes various assumptions, such as the existence of equilibrium markets, stock prices assumed to follow a random walk in continuous time and are lognormally distributed, the existence of constant short term risk free interest rates, and the assumption of no dividends or transaction costs. The model also assumes a rational investor.

Under these assumptions, Black and Scholes produced an analytic formula for the value of a European call or put option in terms of five quantities: the underlying stock price, S, the strike price X, the risk free interest rate r the time to expiry T and the 'volatility' σ (usually taken as an estimate of the standard deviation of the logarithm of price fluctuations). The value of a European call option under the model based approach is predicted to be

$$C = S \times N(d_1) - X \times \exp^{-r(T-t)} N(d_2)$$

where $d_1 = [ln(S/X) + (R + \sigma^2/2)(T - t)]/\sigma\sqrt{T-t}$ and $d_2 = d_1 - \sigma\sqrt{T-t}$, and $N(x)$ is the cumulative probability distribution function for a standardised normal variable. However this formula, though widely in use today, is known to consistently underestimate or overestimate the market depending upon whether the option contract is a call or a put.

Figure 1.1 shows the actual and predicted values for 755 Put options on $/DM exchange rates which demonstrates the consistent bias. Despite many efforts to improve on the basic model (for example, see articles in [3]), there is still no reliable model to estimate the market value of options contracts.

As an alternative to the model based approach, we can view the problem of estimated the price of a Put contract P (say) as one of nonlinear regression by the fusion of the five driving variables, $\{S, X, r, T, \sigma\}$. Since we require a flexible semiparametric model capable of approximating the underlying (but unknown) data generator, we choose to employ a radial basis function spline network. Figure 1.2 depicts the simple fusion process.

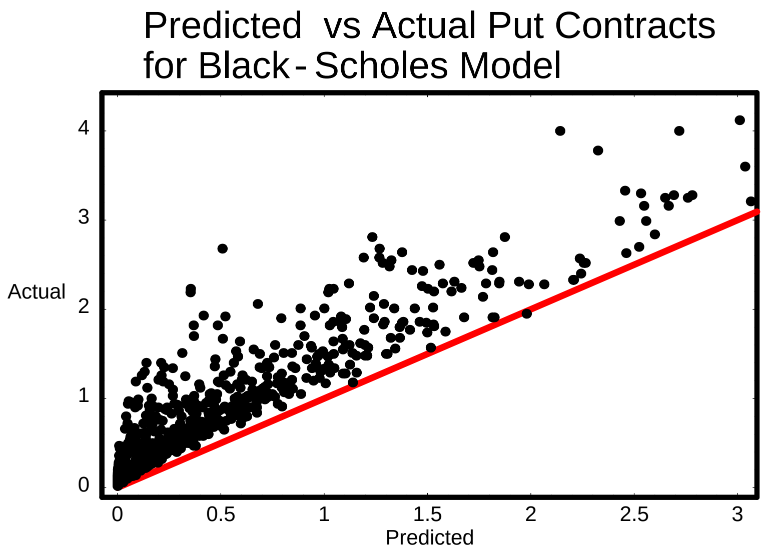

Figure 1.1 Scatter plot of actual and predicted Put options contracts obtained by the standard Black-Scholes model.

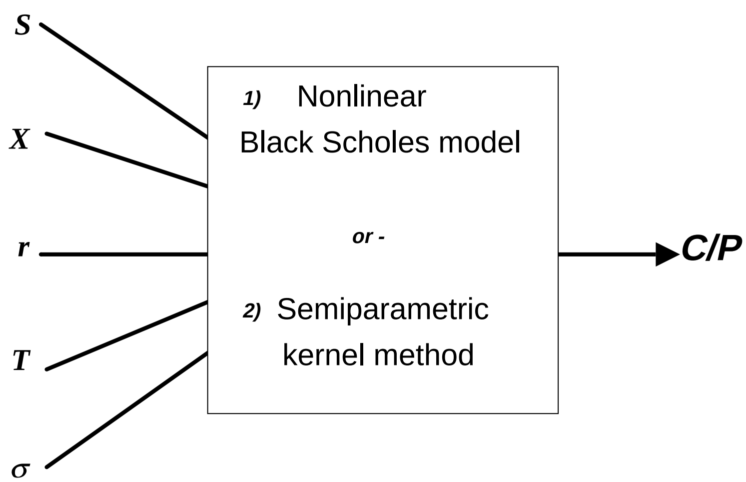

Figure 1.2 Option pricing by model-based or semiparametric-based regression

As a demonstration we choose to estimate the market value of European put options on $/DM exchange rates. Training data was taken from 42 consecutive days between February 8th 1989 and April 18th 1989. This amounted to 1592 separate contracts. The risk free interest rates were taken from the contemporary values of 3 month Treasury Bills as a good proxy for true risk free interest rates, and volatility was calculated using an optimally chosen sliding window on historic data.

Note that volatility is the only parameter not directly accessible and significant activity occurs around the financial world to estimate accurate volatilities. Volatility is generally (though not by this author) considered to be a key parameter for option pricing, to the extent that options are also often quoted in volatility values. We examined a variety of methods to estimate volatility which we could also use in out-of-sample experiments (and so *implied* volatility calculations are not available since for implied volatility we also need the cost of the contract in the first place. We eventually settled on a chosen sliding window calculated on historic values on the basis of overall minimum mean square error between estimated and actual option values on an out-of-sample set. All other information was easily available.

The network performance was tested against the consecutive trading days, between May 24th 1989 and June 30th 1989, amounting to 752 different contracts. The input data was prewhitened to be zero mean, unit variance on a channel by channel basis (with parameters evaluated in the training data only).

The 'fusion' in this example is simply achieved by the single RBF network acting as a universal approximator to produce an estimate of the (unknown) nonlinear regressor, or data generator, which connects the 5 input variables with the single output (the price of the option contract). Combining multiple models to improve output estimates will be demonstrated later in this chapter. Note that in this example, since we are only using a single model to achieve the fusion of the inputs, it is quite important to match the complexity of the model (as determined essentially by the number of basis functions) to the complexity of the problem. There are several ways we could do this. For the purposes of this experiment we used cross validation on a separate contiguous data set of values and then fixed the model complexity when applied to the separate test set which we used to present the results in the figures.

Figure 1.3 shows out of sample performance of the semiparametric approach. The scatterplots show the actual market value of the contracts plotted against the Radial Basis Function predictions. Clearly the model has produced very good estimates of the underlying values of the contracts with a Spearman rank order correlation of 0.9753, hence indicating very good correlations between the network models and the data.

What is particularly noteworthy is the robustness of estimates even for high priced contracts where the data density is clearly more sparse. These results

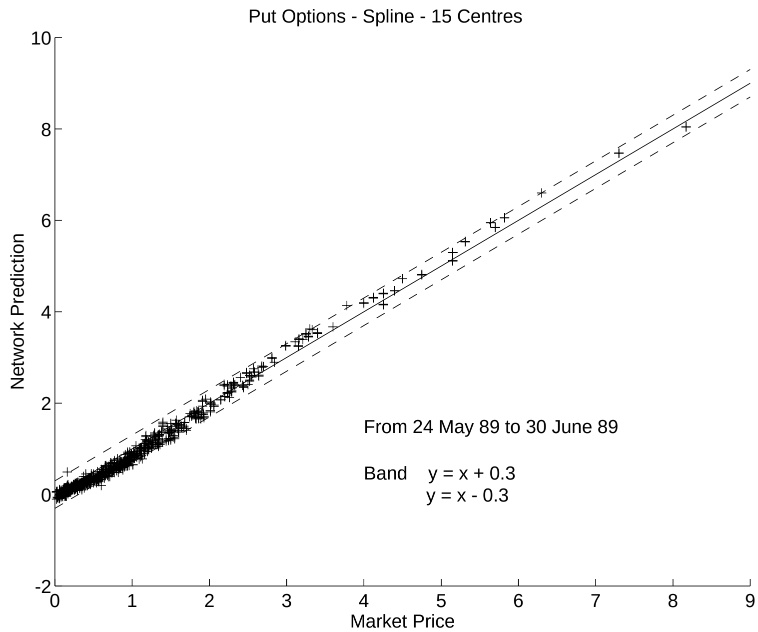

Figure 1.3 Scatter plot of actual and predicted Put options contracts obtained by the simple fusion model.

provide a major improvement over the standard Black-Scholes model which has a consistent bias away from the true values. Results on Call contracts are similar though not as good as these obtained on Put contracts. Further details can be obtained in [7].

An interesting aside developed as part of these nonparametric models is a demonstration of the benefits of unconstrained semiparametric data fusion. Specifically, the main problem with option pricing models lies in the determination of the *volatility*. The remaining parameters may be directly obtained or inferred from reliable sources. Much effort is involved in attempting to produce better interpretations and estimates of volatility, as it is widely perceived as an indication of risk. It also dominates the uncertainty in the Black-Scholes formula. However, given that volatility is an inherent property of the time series, and we employ the time series in the semiparametric model, then is it feasible that volatility is already implicitly contained within the underlying stock price?

Since the semiparametric approach is not constrained by the model based assumptions of the Black-Scholes model we may investigate this question. The results of Figure 1.4 indicate that a 4 parameter model excluding volatility still performs significantly better than the Black-Scholes model. These ramifications have yet to be properly investigated.

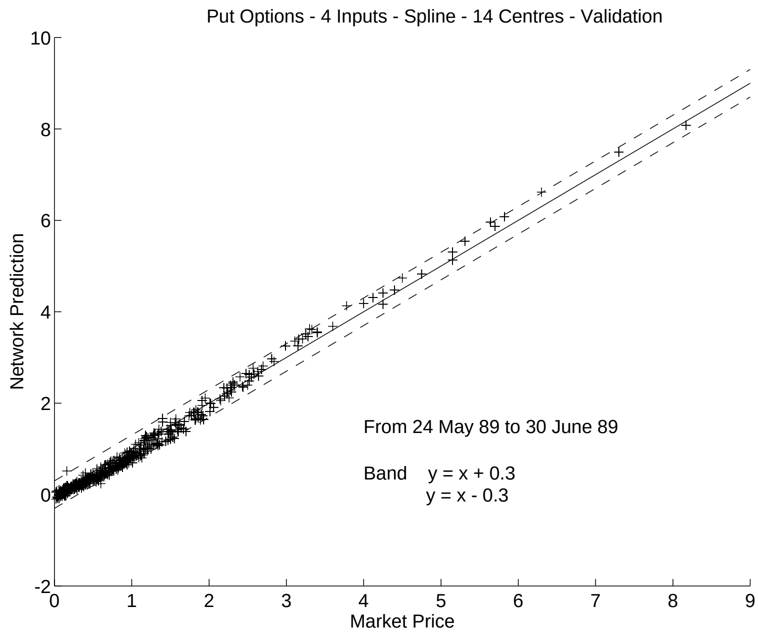

Figure 1.4 Scatter plot of actual and predicted Put options contracts obtained by the simple fusion model, but only using a 4 parameter regression.

3. MODEL COMBINATION

In finance, it is common practice to employ a variety of models when trying to produce forecasts or inferences, and then select those models which the user has most confidence in. Interestingly, model selection is not a particularly good method of arriving at a decision, even when some of the models give quite poor results. Instead, a simple 'model fusion' result of simple averaging of the predictions of a variety of models – especially when those models have a diverse basis – usually returns superior performance. We now examine this higher level data fusion issue of model averaging, and demonstrate the main result on a single-equity iterated forecasting problem.

3.1 ERROR AVERAGING

Let us first provide some background mathematics justifying the combination of models - even when it is known that some of the models are inferior (e.g. linear models). We first consider the case of a simple, uniform averaging of multiple models. Consider the case when we have M distinct models (we are not concerned about the specific forms here), each one of which produces an output value $y_i(\boldsymbol{x}) = r(\boldsymbol{x}) + \epsilon_i(\boldsymbol{x})$. $r(\boldsymbol{x})$ is the true (unknown) function of the input data $\boldsymbol{x}$.

The average (squared) error produced by each model i is $E_i = \langle [y_i(\boldsymbol{x}) - r(\boldsymbol{x})]^2 \rangle = \langle \epsilon_i^2 \rangle$ where the expectation $\langle \ldots \rangle$ is over the distribution of the data $\boldsymbol{x}$. The average (squaared) error over these models would be $\hat{E} = \sum_{i=1}^{M} E_i / M$.

Now consider the expected error produced by a simple average fusion over all model predictions. By averaging over all models, a modified prediction value would be just $\hat{y} = \sum_{i=1}^{M} y_i(\boldsymbol{x})/M$. This gives an error due to this 'committee' of equally voting models as $\hat{E}_M = \langle (\sum_{i=1}^{M} \epsilon_i / M)^2 \rangle$. If we make the assumption that each model's errors are independent at second order, then this expected committee error is simply related to the error of each committee member as:–

$$\hat{E}_M = \frac{1}{M} E_i$$

This result seems to indicate that a significant reduction (a factor of M for M models) in expected mean error can be achieved by a simple averaging over multiple models. Of course the real error reduction is much smaller than this as the assumption that the committee model errors are independent is in practice false. Nevertheless, in practice it is found that performance improvements can be made by this simple averaging approach. Since the major gains are to be made when the errors are uncorrelated then this leads to the desire of using multiple diverse models, since this generally increases the chance of the error distributions being different. It also justifies the use of even keeping weak models (e.g. linear models) in the committee scheme. This reduction in error is derived from a reduced variance. So this tells us that it is preferable to choose models with small bias and reduced the overall error by variance reduction. In terms of neurl network models, this implies using a collection of models which are too complex for the problem at hand.

A word of warning, however. Committee averaging is good to improve the performance of a collection of weak models. It is an expectation *on average* across models. Committee averaging will reduce the effectiveness of using a set of very good models – or equivalently, if one model is known to be the expert, it is best to use that one model rather than heed the committee!

3.2 WEIGHTED ERROR AVERAGING

Despite the previous result, we might imagine that a better fusion model exists if we allow for differentially weighting those committee members who we believe give better predictions more heavily in the averaging process. So more generally we still consider a simple linear averaging, but in the form $\hat{y}_W(\boldsymbol{x}) = \sum_{i=1}^{M} \alpha_i y_i(\boldsymbol{x})$. Under the constraint that $\sum_i \alpha_i = 1$, minimising the weighted average committee error with respect to these parameters we find a solution for the weighting coefficients in the form:

$$\alpha_i = \frac{\sum_{j=1}^{M} (\boldsymbol{C}^{-1})_{ij}}{\sum_{i=1}^{M} \sum_{j=1}^{M} (\boldsymbol{C}^{-1})_{ij}}$$

which gives a value for the expected minimum error as $\hat{E}_W = \left(\sum_{i=1}^{M} \sum_{j=1}^{M} (\boldsymbol{C}^{-1})_{ij} \right)^{-1}$. In these expressions $\boldsymbol{C}^{-1}$ is the inverse covariance matrix between the different model errors: $C_{ij} = \langle \epsilon_i(\boldsymbol{x}) \epsilon_j(\boldsymbol{x}) \rangle$. Hence, either knowing something about the correlation between the errors of the different models (or estimating the covariance matrix from sampling of the trained models), allows us to construct a better weighted average committee.

From an analysis of the error of the weighted committee network [6] we can show that the expected error by weighting exploiting the covariance knowledge will in general be better than the simple averaging of committee members. The weighted committee error may be decomposed into two terms:

$$\begin{aligned} \hat{E}_W = & \sum_i \alpha_i \langle (y_i(\boldsymbol{x}) - r(\boldsymbol{x}))^2 \rangle \\ & - \sum_i \alpha_i \langle (y_i(\boldsymbol{x}) - \hat{y}_W(\boldsymbol{x}))^2 \rangle \end{aligned}$$

This is reminiscent of the bias-variance decomposition, though the negative sign assists in reducing the overall error by allowing for a more diverse variance, provided the bias is not increased. The first term is the error due to the individual models around the true regression. The second term is the effect of the spread of predictions of the committee members around the weighted average committee prediction – the 'variance'. Hence if we can increase the spread of predictions around the committee prediction without increasing the individual member prediction errors, then we should be able to reduce the overall committee error.

3.3 EXAMPLE OF MODEL WEIGHTING

As a simple example of a prediction-fusion model, we considered the forward iterative prediction of the quoted daily closing price of HSBC plc from the FTSE-100 Price Index based only on historic information. In this example we used the previous two years of price information. Individual predictive models were obtained based on 10 multilayer perceptron networks separately optimised by standard methods. A model with 10 hidden units was employed. Note that model order selection techniques were not emphasised as the model averaging compensates for small bias models. Each model was then initialised using data up to the end of the training period, and then the outputs of each model were recursively fed back into the inputs of the models, thus allowing each model to produce an iterated daily forecast of HSBC prices for each of 21 days in

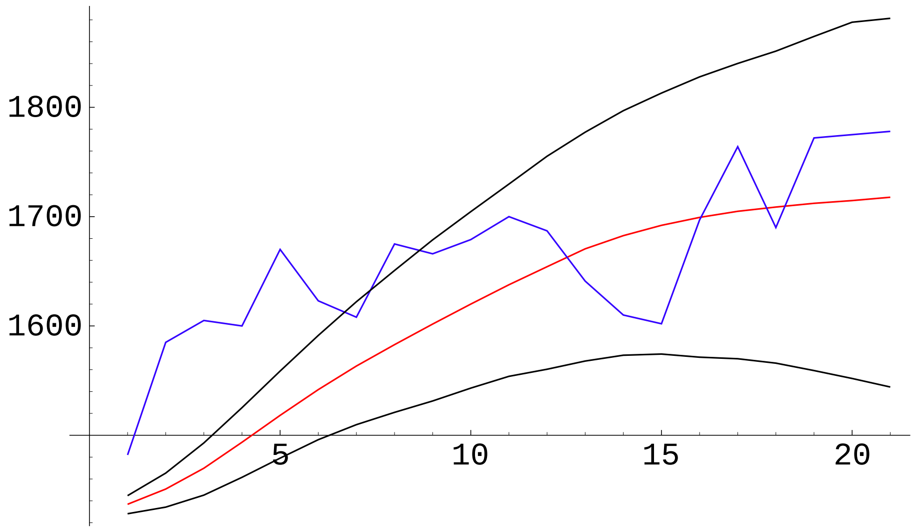

Figure 1.5 The mean forecast using the fusion of model outputs (the middle smooth line) iterated 21 days ahead and the true price trajectory (the top jagged line). Also shown are the confidence intervals.

the future. These predictions were then averaged using the simple weighting model and the results displayed in Figure 1.5.

This figure also shows an estimate of confidence intervals. There are many techniques for obtaining estimates for the confidence of predictions using non-linear models [8] – some including fusion techniques as indicated in the previous variance analysis. Each has respective advantages and disadvantages depending upon the assumptions which are used in their derivations. Figure 1.5 employed a bootstrap technique [9] to produce the confidence bands. Results obtained using this method of committee fusion on out-of-sample performance have been demonstrated to return consistently better performance than single model estimates and is now part of a live trading system.

As a comparison, Figure 1.6 provides the prediction estimates for three of the individual models.

In this example the model averaging approach produced estimates as good as the *a posteriori* selected best individual model, as evidenced in Figure 1.7 which shows the spread of sum squared errors made by the various models on out of sample performance.

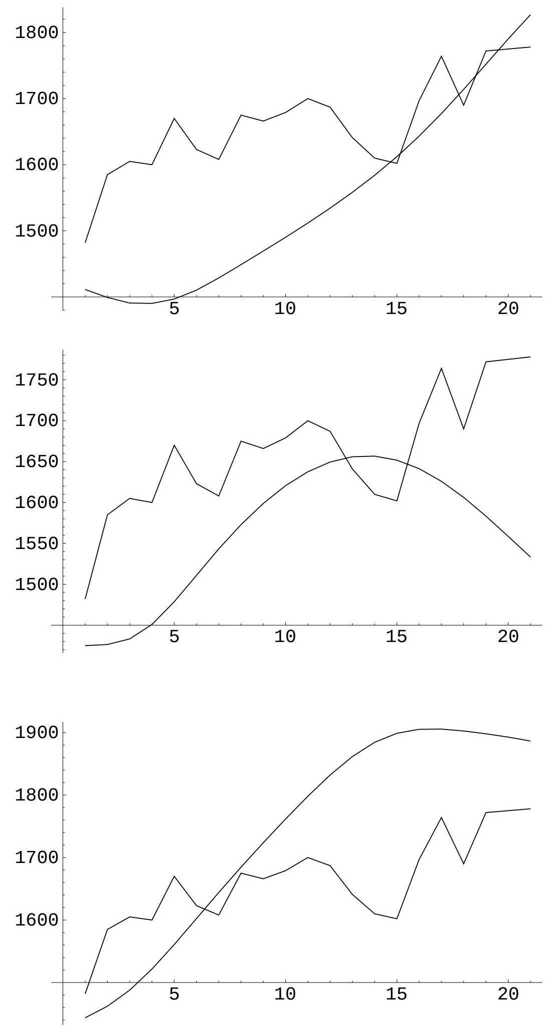

Figure 1.6 Exemplar prediction estimates produced by three of the individual models, demonstrating the diversity of the outputs of the models

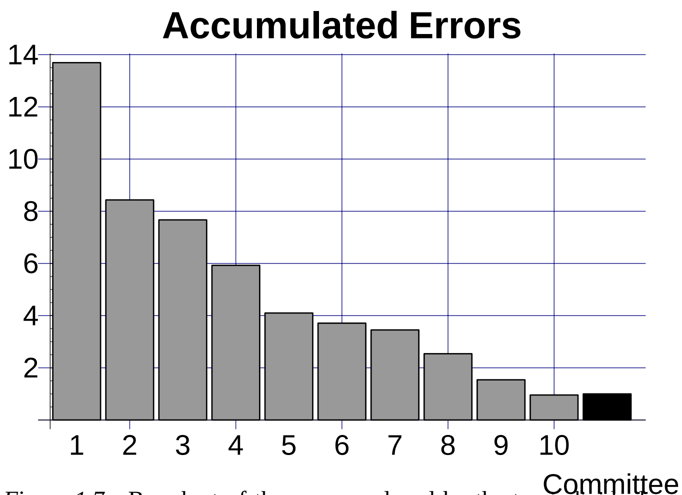

Figure 1.7 Bar chart of the error produced by the ten individual models and the committee model (dark-shaded bar) accumulated over all 21 predictions and normalised to the error of the committee result.

3.4 BAYESIAN COMMITTEES

We can pose part of the model fusion problem, especially for semiparametric models such as neural networks, in an approximate Bayesian framework. A basic principle of a Bayesian approach to inference is that integration over models is superior to model selection. It is of no surprise therefore to note that the committee approach to fusion can be shown to emerge from a Bayesian perspective. It is instructive to provide a brief outline sketch. Each of the models in our committee or portfolio is usually characterised by a set of parameters $\boldsymbol{w}$ (e.g. the weights in a neural network) sampled from a corresponding probability density $p(\boldsymbol{w}|data)$. For sufficiently diverse models after training on historic data the chosen weights are likely to be sampled from distinct regions of parameter space. (For example, for neural networks, symmetry and redundancy indicate that there are distinct weight vectors which yield neural network models with simlar or equivalent perfromance in terms of a cost function such as mean square error or mutual information.) We assume that the complete density may be approximated by a set of effectively distinct densities – one for each model m: $p(\boldsymbol{w}|data) \approx \sum_m p(m, \boldsymbol{w}|data) = \sum_m p(\boldsymbol{w}|m, data)P(m|data)$. All observable quantities should be obtainable by averaging over the density function. In general we cannot perform this averaging. However for the simple assumption of separated model regions in parameter space we can decompose the average. In particular, the expected value of the prediction of the committee of models would be just :

$$\begin{aligned}\hat{y}_B(\boldsymbol{x}) &= \int y(\boldsymbol{w},\boldsymbol{x})p(\boldsymbol{w}|data)d\boldsymbol{w} \\ &\approx \sum_m P(m|data)\int_{\Omega_m} y(\boldsymbol{w},\boldsymbol{x})p(\boldsymbol{w}|m,data)d\boldsymbol{w} \\ &= \sum_M P(m|data)\hat{y}_m\end{aligned}$$

where Ω_m is the region of 'activity' of parameter space surrounding model m and $\hat{y}_m$ is the expected prediction due to model m.

Note that this simple outline provides the best solution as a fusion of individual committee member predictions where the weighting depends on the likelihood of each particular committee member occurring.

It is interesting that committee averaging can be obtained from a Bayesian perspective, where a different weighting scheme is suggested for the source models. However we note that the Bayesian perspective has assumed that the individual source models are accurately known. We saw earlier in this chapter that decision-level fusion was significantly distorted if we allowed for errors in the source models. This result may account for the observation that pragmatically we have found that weightings determined by the previous methods (and usually just the simple uniform weighting) typically produce better predictions than this Bayesian separable approximation. However we recognise this as an interesting open issue which needs further study - theoretically and numerically.

4. CONCLUSIONS

In this short chapter we have merely selected a few examples of where data fusion methodologies are likely to be useful in the technical analysis of financial problems. We have completely excluded fascinating discussions of other topics. Amongst these we mention a few obvious candidates for the application of data fusion techniques. These include: high level decision making by exploiting alternative utilities under uncertainty, the exploitation of subjective information combined with data analysis for visualisation, the tradeoffs in portfolio construction, selective source selection by confidence ranking, model output moderation by the incorporation of uncertainty modelling, and indeed the whole area of uncertainty modelling in financial fusion. The potential for exploitation of fundamental data fusion techniques in this arena is enormous. Given the obvious benefit and potential gain to be derived, it is extremely surprising that there is hardly any real activity in this domain. It is common folklore that advances in financial engineering are not made public if they are a success, because they are used for profit 'behind closed doors'. In which case, perhaps data fusion techniques really are being exploited within the industry successfully. The alternative hypothesis is that there are very few useful principled techniques being applied and it is the knowledge of this paucity that is being concealed. I favour the latter.

We should also stress that we have only demonstrated very simple information fusion strategies, and concentrated upon those approaches which can be readily applied to the real world problems. Yet despite the simplicity (which in itself has some theoretical support as indicated in this chapter), the experimental evidence has demonstrated the effectiveness of applying such techniques. It is our hope that this brief overview might stimulate some small interest in the development of information fusion towards technical financial analysis.

Acknowledgments

This work was supported in part by the EPSRC contract GR/K51815, *"Nonstationary Feature Extraction and Tracking for the Classification of Turning Points in Multivariate Time Series"*. I would also like to extend my thanks to my student Krzysztof Zapart for helping me with some of this work, and to several collaborators from various financial domains who have provided me with some small amount of their expert domain knowledge to allow me to progress slightly with some of their problems. Of specific note has been the guidance provided by Dr C Satchwell of Technical Forecasts Ltd.

References

[1] Black F, and Scholes M, "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, **81**, 637–659, (1973).

[2] "Defence and Aerospace Foresight Technology Working Party: Data Fusion and Data Processing", Royal Aeronautical Society, ISBN 1857680685, (1997).

[3] Hand DJ, Jacka SD, (eds), *Statistics in Finance*, Arnold, (1998).

[4] Haussler D, and Opper M, "Mutual Information, Metric Entropy and Cumulative Relative Entropy Risk", *The Annals of Statistics*, **25**(6), 2451–2494, (1997).

[5] Kittler J, "Combining classifiers: A theoretical framework", *Pattern Analysis and Applications*, **1**, 18–27, (1998).

[6] Krogh A and Vedelsby J, "Neural network ensembles, cross validation and active learning", In D S Touretzky, G Tesauro, T K Leen (eds) *Advances in Neural Information Processing Systems*, **7**, Cambridge MA: MIT Press, (1995).

[7] Lowe D, "On the use of nonlocal and nonpositive definite Basis functions in RBF networks",206–211, *Proceedings 4th IEE International Conference on Artificial Neural Networks*, Conference Publication number **409**, The Institute of Electrical Engineers, (1995).

[8] Lowe D and Zapart K, "Validation of Neural Networks in Automotive Engine Calibration", *5th IEE International Conference on Artificial Neural Networks*, **440**, 221–226, The Institute of Electrical Engineers, (1997).

[9] Zoubir A M and Boashash B, "The Bootstrap and its Application in Signal Processing", 56–76, *IEEE Signal Processing Magazine*, **15**(1), (1998).