BUILDING FINANCIAL TIME SERIES PREDICTIONS WITH EVOLUTIONARY ARTIFICIAL NEURAL NETWORK

Serge Hayward

Department of Finance
Ecole Supérieure de Commerce de Dijon, France

Abstract: Price forecasting and trading strategies modeling are examined with major international stock indexes under different time horizons. Results demonstrate that an accurate prediction is equally important as a stable saving rate for long-term survivability. The best economic performances are achieved for a one-year investment horizon with longer training not necessarily leading to improved accuracy. Thin markets' dominance by a particular traders' type (e.g. short memory agents) results in a higher likelihood to learn with computational intelligence tools profitable strategies, used by dominant traders. An improvement in profitability is achieved for models optimized with genetic algorithm and fine-tuning of training/validation/testing distribution.

Keywords: Neural Networks; Genetic Algorithms; Finance; Forecasts; Models

1. INTRODUCTION

Recent advances in computational intelligence (CI) stimulate new development in financial modeling and forecasting. This research is motivated by an assertion (Blume and Easley, 1992) that agents' long-term fitness is a function of appropriate preferences rather than accurate predictions; and a computational experiment conformation of this claim by (Chen and Huang, 2003), where a key factor of survivability is given by the stability of traders' saving rate. Further inspiration for this paper comes from the supposition (Chen and Huang, 2003) that, as longer validation periods get agents' beliefs closer to the true process, strategy's accuracy increases with the memory length.

In finance this paper is interested in examining relationships between stock traders' investment periods, memory lengths and long/short term fitness. In CI it is driven by a search for statistic and economic foundation of an evolutionary / artificial neural network (E/ANN).

2. TIME HORIZONS

Stylized facts often highlight that financial assets prices exhibit non-stationary behavior with profound structural brakes. Thus, in a market, with a frequently changing data generating mechanism it could be advantageous to look at the limited past. From an experiment design perspective, a model with short term training is more likely to over-fit the data, where, training a model over long term could result in missing out potentially useful (although, only developing) contemporary relationships.

With regards to investment horizons, we examine the behavior of short term speculating traders, defined by a one year forward period and long term investing traders, defined by a three years forward horizon. Long term traders are represented by three types: those who make investment decisions once every three years; those who make only portfolio decisions at the end of each year, reinvesting all the capital generated from an yearly trading and those who make portfolio and saving decisions at the end of each year, with reinvestment equal to $w_t(1-v_t)$, where w_t is wealth, accumulated at the end of trading period t and v_t is the saving rate. In the experiment we condition on the 'optimal' (for long-term survivability) saving rate, suggested in (Chen and Huang 2003). Inheriting agents with such an 'optimal' saving rate, allows us to examine profitability of trading decisions over short and longer terms.

A trading strategy choice (with regards to the time horizons) is a function of market conditions, which themselves are a function of strategies used by agents, populated this market. In these settings market conditions (or strategies used by the dominant type of traders) determine the optimal memory length. This approach is an effort towards considering a market environment endogenously in financial data mining.

3. METHODOLOGY

For our experiment we build ANN forecasts and generate a posterior optimal rule. The rule, using future information to determine the best current trading action, returns a buy/sell signal (B/S) today if prices tomorrow have increased/decreased. A posterior optimal rule signal (PORS) is then modeled with ANN forecasts, generating a trading B/S signal. Combining a trading signal with a strategy warrants a position to be taken. We consider a number of market timing strategies, appropriate for different strengths of the B/S signal. If we have a buy (sell) signal on the basis of prices expected to increase (decrease) than we enter a Long (Short) position. Note that our approach is different from standard B/S signal generation by a technical trading rule. In the latter it is only a signal from a technical trading rule that establishes that prices are expected to increase/decrease. In our model we collaborate signal's expectations of price change (given by PORS) with a time-series forecast.

To apply our methodology we develop the dual network structure, presented in Fig. 1. The forecasting network feeds into the action network, from which the information set includes the output of the first network and PORS, as well as the inputs used for forecasting, in order to relate the forecast to the data upon which it was based.

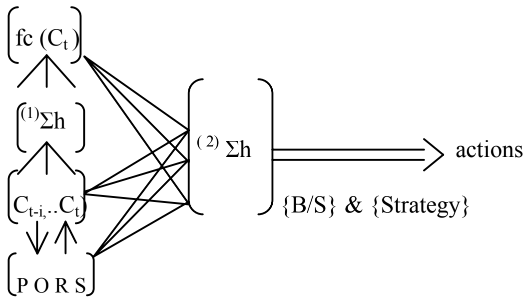

Fig. 1. Dual ANN: (1) forecasting network; (2) acting network

This structure is an effort to relate actions' profitability to forecasting quality, examining this relationship in computational settings. The model is evolutionary in the sense it considers a population of networks (individual agents facing identical problems/instances) that generate different solutions, which are assessed and selected on the basis of their fitness. Backpropagation is used in the forecasting net to learn to approximate the unknown conditional expectation function (without the need to make assumptions about a data generating mechanism and beliefs formation). It is also employed in the action net to learn the relationship between forecasts' statistical and actions' economic characteristics. Lastly, agents discover their optimal models with genetic algorithm (GA); applying it for ANN model discovery makes technical decisions less arbitrary. The structure seems to be intuitive and simple to generate results independent from a chosen architecture. The results produced are sufficiently general, being stable for multiple independent runs with different random seeds for a dual forecasting/action net and a single forecasting net.

3.1 Generating Posterior Optimal Rule Signal

PORS is a function of a trading strategy adopted and based on the amount of minimum profit and the number of samples into the future. Stepping forward one sample at a time, the potential profit is examined. If the profit expected is enough to clear the minimum profit after transaction costs (TC), a PORS is generated. The direction of PORS is governed by the direction of the price movement. Normally, the strength of the signal reflects the size of underlying price changes, although, we also examine signals without this correlation to identify when profit generating conditions begin. Lastly, we consider PORS generated only at the points of highest profit to establish the maximum profit available.

4. DESCRIPTION OF THE ENVIRONMENT

Let Y be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. Ω is a space of outcomes, $\mathcal{F}$ is a σ-field and P is a probability measure. For a space $(\Omega, \mathcal{F}, P)$ a conditional probability $P[A|\mathcal{F}]$ for a set A, defined with respect to a σ-field $\mathcal{F}$, is the conditional probability of the set A, being evaluated in light of the information available in the σ-field $\mathcal{F}$. Suppose economic agents' utility functions are given by a general form:

$$U(W_{t+s}) = g(Y_{t+s}, \delta(fc_{t+s})) \quad (1)$$

According to (1), agents' utility depends on: a target variable Y_{t+s}; a decision/strategy variable, $\delta(fc_{t+s})$, which is a function of the forecast, fc_{t+s}, where $s \geq 1$ is a forecasting horizon. Setting the horizon equal to 1, we examine the next period forecast (when this simplification does not undermine the results for $s \geq 1$). A reward variable W_{t+s} is sufficiently general to consider different types of economic agents and includes wealth, reputation, etc. $w_{t+1}(y_{t+1}, fc_{t+1})$ is the response function, stating that at time $t+1$ an agent's reward w_{t+1} depends on the realization of the target variable y_{t+1} and on the accuracy of the target's forecast, fc_{t+1}. Forecasting is regarded as a major factor of a decision rule, being close to the reality in

financial markets. Also, it has a developed statistical foundation in econometrics allowing its application in computational settings.
Let $fct_{+1}=\theta' X_t$ to be a forecast of Y_{t+1} conditional on the information set $\mathcal{F}_t$, where unknown m-vector of parameters, $\theta \in \Theta$, with Θ to be compact in $\mathbb{R}^k$ and observable at time t n-vector of variables, X_t. X_t is $\mathcal{F}_t$-measurable and might include some exogenous variables, indicators, lags of Y_t, etc. An optimal forecast does not exclude model misspecification, which can be due to the form of fc_{t+1} or failure to include all relevant information in X_t. With imperfect foresight, the response function and, therefore, the utility function are negatively correlated with forecast error, $e_{t+1} \equiv y_{t+1} - fc_{t+1}$; $|e_{t+1}| > 0$. A mapping of the forecast into a strategy rule, $\delta(fc_{t+1})$ (combined with elements of X_t) determines a predictive density g_y, which establishes agents' actions.
In this setting, maximizing expected utility requires us to find an optimal forecast, fc_{t+1} and to establish an optimal decision rule, $\delta(fc_{t+1})$. Note that optimality is with respect to a particular utility function, implemented through a loss function, in the sense that there is no loss for a correct decision and a positive loss for incorrect one. Given a utility function, expected utility maximization requires minimization of the expected value of a loss function, representing the relationship between the size of the forecast error and the economic loss incurred because of that error. A strategy development (mapping of the forecast into a decision rule) is another way to minimize the expected value of a loss function.
A loss function, L: $\mathbb{R} \to \mathbb{R}^+$, related to some economic criteria or a statistical measure of accuracy, takes a general form:

$$L(p,\ a,\ e) \equiv [a + (1\text{-}2a)1(e < 0)]e^p, \qquad (2)$$

where p is a coefficient of risk aversion; e is the forecast error; $\alpha \in [0,1]$ is the degree of asymmetry in the forecaster's loss function. $L(p,\ \alpha,\ e)$ is $\mathcal{F}_t$-measurable. It could also be presented as:

$$L(p,\ a,\ \theta) \equiv [a + (1\text{-}2a)1(Y_{t+1} - fc_{t+1}(\theta) < 0)]\,|Y_{t+1} - fc_{t+1}(\theta)|^p, \qquad (3)$$

where α and p are shape parameters and a vector of unknown parameters, $\theta \in \Theta$. For given values of p and α an agent's optimal one-period forecast is

$$\min_{\theta\in\Theta} E[L(\rho,\ \alpha,\ \theta)] = E[L(Y_{t+1} - fc_{t+1})] = E[L(e_{t+1})]. \quad (4)$$

Training EANN with different training, validation and testing durations allows us to examine how agents' statistical and economic performances relate to their past and forward time horizons.

5. EXPERIMENTAL DESIGN

We use ANN with GA optimization for the building/evolution of price forecast and trading strategy development/evolution upon relevant forecast. The mechanism appears to be an intuitive way to deal with agents' cognitive limits in forecasting and optimization, modeling the traders' learning process to approximate the unknown conditional expectation function. It also provides a natural procedure to consider decisions' heterogeneity by agents viewing similar information. A single hidden layer ANN seems to be sufficient for our problem, particularly considering the universal approximation property of feedforward nets. GA facilitates an optimal choice of network settings and adds additional explanatory power to the analysis.

6. PERFORMANCE SURFACE

The performance of ANN learning is monitored by observing how the cost changes over training iterations. The learning curve presents the internal error over each epoch of training, comparing the output of the ANN to the desired output. In price forecasting, the target is the next day closing price, where in signal modeling, the target is the current strategy. Achieving an accurate representation of the mapping between the input and the target might not necessarily lead to a forecast to be exploitable or a strategy using that forecast to be profitable.
We consider that evaluation criteria should measure not so much absolute effectiveness of the model with respect to the environment[1] but rather its relative effectiveness with respect to other models. Although we train ANN with the goal to minimize internal error function, we test and optimize its generalization ability by comparing its performance with the results of a benchmark, an efficient prediction (EP). In forecasting prices, EP is the last known value. For predicting strategies, it is the buy/hold (B/H) strategy. The degree of improvement over efficient prediction (IEP) is calculated as an error from a de-normalized value of the ANN and a desired output, then normalizing the result with the difference between the target and EP value.

7. PROFITABILITY AS PERFORMANCE MEASURE

To make the final goal meaningful in economic terms we use profitability as a measure of overall success. We examine the following forms of cumulative and individual trades return measures: non-realized simple aggregate return (r); profit/loss factor; average, maximum gain/loss. In addition we estimate exit efficiency, measuring whether trades may have been held too long, relative to the maximum amount of profit to be made, as well as the

[1] (Chen and Huang, 2003) found correlation between the Kolmogorov-Smirnov statistics and the length of validation period. Assuming that traders' beliefs with longer validation periods get closer to the true process in simulations and agents' accuracy increases, they consider the time horizon that agents use for validation as a representation of the accuracy of prediction.

frequency and the length of trades, including out of market position. To assess risk exposure we adopt common 'primitive' statistics, the Sharpe ratio (SR)[2] and the maximum drawdown (ξ). The latter, calculating the percentage loss relative to the initial investment for the date range, measures the size of losses occurred while achieving given gains. It demonstrates how prone a strategy is to losses. To overcome the Fisher effect we consider trading positions with a one-day delay.

TC is assumed to be paid both when entering and exiting the market, as a percentage of the trade value. TC accounts for broker's fees, taxes, liquidity cost (bid-ask spread), as well as costs of collecting/analysis of information and opportunity costs. According to (Sweeney, 1988) large institutional investors achieve one-way TC about 0.1-0.2%. Often TC in this range is used in computational models. Since TC (defined above) would differ for heterogeneous agents, we report the break-even TC that offsets trading revenue with costs leading to zero profits.

Thus, in this paper profitability is a function of return, risk and transaction costs. The classification of the ANN output as different types of B/S signals determines the capability of the model to detect the key turning points of price movement. Evaluating the mapping of a forecast into a strategy, $\delta(fc_{t+1})$, assesses the success in establishing a predictive density, g_y that determines agents' actions.

8. TRADING STRATEGIES STYLES

Both long and short trades are allowed in the simulation. Investing total funds for the first trade, subsequent trades (during a year) are made by re-investing all of the money returned from previous trades. If the account no longer has enough capital to cover TC trading stops. For agents making annual portfolio and saving decisions the saving rate, v_t = 0.20773 and the risk free interest rate, r_s = 0.1.

Long-term traders with annual portfolio (saving) decisions use a sliding window reinvestment scheme, presented in Fig. 2. Training/validation/testing (Tr/V/Ts) periods indicate that following a yearly investment (24.01.01-23.01.02), agents reinvest their wealth for another year (24.01.02-23.01.03) and then for one more year (24.01.03-23.01.04).

26.07.98**Tr**25.07.00**V**24.01.01**Ts**23.01.02
26.07.99 **Tr** 25.07.01**V**24.01.02**Ts**23.01.03
26.07.00 **Tr** 25.07.02**V**24.01.03**Ts**23.01.04

Fig. 2. Sliding window reinvestment scheme

[2] Given by the average return divided by the standard deviation of that return.

9. GENETIC TRAINING OPTIMIZATION

In this research EC is used for ANN model discovery, considering GA optimization for: network's topology; performance surface; learning rules; number of neurons and memory taps; weight update; step size and momentum rate. GA tests various settings from different initial conditions (in the absence of a priori knowledge and to avoid symmetry that can trap the search algorithm). Since the overall objective of financial forecasting is to make a trading decision, based on that forecast profitable, economic criteria rather than statistical qualities need to be employed for the final goal. We use GA optimization with the aim to minimize IEP value and profitability as a measure of overall success.

10. EMPIRICAL APPLICATION

10.1 Data

We consider daily closing prices for CAC 40, DAX, AEX, FTSE ALL, NASDAQ 100, S&P 100 and RUSSELL 2000 share indexes obtained from Yahoo Finance. The choice of series is given by similarities in their statistical characteristics with associated diverse behavior of market participants. The time period under investigation is 23.01.93-23.01.04. There were 2770 observations in each row data sets. Examining the data graphically reveals that the stock prices exhibit an upward, but non-linear trend until around the year 2000 and a downward trend until the spring 2003. Persistent fluctuations around the trends, which increase in variability with downward moves, can also be seen. Descriptive statistics for each index confirm that the unit-root hypothesis cannot be rejected at least at 5% critical level. Visual examination also identifies the three common structural brakes around: 07.98, 03-09.00 and 03-04.03.

Changes in prices exhibit clustering volatility. The daily returns display excess kurtosis and the null of no skewness is rejected for all data. The tests statistics lead to rejection of the Gaussian hypothesis for the distributions of the series. It confirms that high-frequency stock returns follow a distribution incompatible with normality assumed often in the analytical literature.

11. EXPERIMENTAL RESULTS

ANN with GA optimization was programmed with various topologies[3]. We have generated and considered 63 forecasting and 107 trading strategies'

[3] Programs in Visual C++, v. 6.0 are available upon request. We have run tests on TradingSolutons, v. 2.1 and NeuroSolutions v. 4.22.

settings for each series. Effectiveness of search algorithm was examined with multiple trials for each setting. On average 87% of 10 individual runs produce identical results, confirming the replicability of our models. Efficiency of the search was assessed by the time it takes to find good results. The search with ANN unoptimized genetically took a few minutes, where the search with GA optimization lasted on average 180 minutes on a Pentium 4 processor.

By simulating traders' price forecasts and trading strategies development, agents' economic performance was found to be the best for a one-year forward time horizon, and it deteriorates significantly in testing exceeding two years, supporting the idea of frequent structural brakes. Over a one year testing period on average 9 trading strategies of each index considered were able to outperform in economic terms the B/H benchmark, with an investment of $10,000 and a TC of 2% of trade value. The primary strategies superiority over the benchmark was on average 53%. The range of break-even TC for primary strategies was [2.6; 5]. Thus, the break-even TC for a number of indexes appears to be high enough to exceed actual TC. GA model discovery reveals that Multilayer Perceptron and Time-Lag Recurrent Network (with Focus Laguarre memory) and Conjugate Gradients learning rule generate the best performance in statistical and economic terms for forecasting and acting nets. Profitability produced by our simple architecture supports computational model development based on economic and statistical foundations.

Table 1 presents annualized economic performance of long-term investing traders. The first figure describes traders attempting to learn a function for three years investment horizon. The second figure characterizes agents who make only portfolio decisions at the end of each year, reinvesting all of the capital generated from the yearly trading. The last figure corresponds to traders who make portfolio and saving decisions at the end of each year, whose total returns include the risky return from stock indexes and risk-free return on the amount saved from investing in stock markets.

Over three years investment horizon, the inferior economic performance characterizes traders without regular annual investment decisions. Profitability of agents, saving an 'optimal' percentage of wealth at a risk-free interest rate, is not conclusively superior to the performance of those reinvesting the entire capital. These results contradict the claims that agents' long-term fitness is not a function of an accurate prediction, but only depends on an appropriate risk aversion, through a stable saving rate. We explain our results through the relationships between riskless, risky returns and the maximum drawdown measure.

Risk free interest rate of 10% seems to be high enough to make saving decisions attractive. Nevertheless, when a risky return is well above a riskless return and a strategy is sufficiently prone to losses while achieving given gains, situations leading to significant decrease in wealth will not necessarily appear. (Note that this explanation does not challenge the fact that investments including savings at a risk-free rate are less risky than total funds reinvestments in stock indexes, as illustrated with Sharpe ratio.) Simulations with different investment periods produce similar results up to 5 years of forward horizon. We conclude, therefore, that a profitable (in terms of risk adjusted returns) strategy development is not less important than an 'optimal' saving decision for reasonably long investment horizons.

Our experiment has identified some relationships between market size and optimal memory horizons. Strategies' simulations demonstrate that a memory length is negatively correlated with a daily trading volume. For instance, DAX trading volume exceeds NASDAQ 100 and even more S&P 100 volumes. At the same time, among strategies superior to B/H, 36.6% were with the shortest training duration for S&P 100, 25% for NASDAQ 100 and 12% for DAX. Among the five worst in economic terms strategies DAX simulation has 60% with the shortest training; where S&P 100 simulation includes no short-training strategies, but 40% of their under-performing strategies with the longest training.

The above results allow us to deduce that in thinner markets there is higher likelihood that short memory horizons traders take the dominant position, influencing price movements. On the other hand, thick markets have more diverse influences from various agents' types. So, strategies of short memory traders have less impact on price movements. Our assertion is also confirmed indirectly by price variances. The volatility of thick markets, considered, exceeds the volatility of thinner markets, indicating diverse pressures of heterogeneous agents, without clear dominance by a particular type.

Considering maximum return available, given by PORS generated at the points of highest profit, we notice that thin markets' modeling potentials exceed those of thick markets. This result corresponds to agent-based computational modeling findings

Table 1 Economic performances of long term traders (annualized return-r; Sharpe ratio-SR; maximum drawdown-ξ)

	CAC 40	DAX	AEX	FTSE ALL	NASDAQ 100	S&P 100	RUSSELL 2000
r	0.76-14.77-13.65	1.34-16.48-21.72	4.02-22.56-29.11	3.02-15.82-12.07	10.2-18.62-18.98	4.82-14.58-17.66	9.26-19.37-16.25
SR	0.01-0.1-0.12	0.06-0.11-0.14	0.08-0.11-0.13	0.1-0.12-0.12	0.7-0.11-0.12	0.14-0.11-0.13	0.1-0.12-0.13
ξ	55.41-21.1-18.48	44.93-32.26-16.34	39.87-19.79-22.3	27.53-18.46-25.49	20.06-23.67-22.6	22.4-33.71-26.42	35.21-24.9-28.12

(Chen and Liao, 2002) that price deviations from the equilibrium are negatively correlated to the market size. We conclude, therefore, that thin markets' dominance by a particular agents' type facilitates a better environment for learning with CI tools of profitable strategies, used by dominant agents.

GA model discovery reveals that recurrent networks' memory depth is a function of market conditions. The environments with frequent structural brakes have the optimal memory depth on average 38% shorter than the markets with less frequent shocks. To test how the frequency of novel concepts' arrival affects modeling of the environment with structural brakes we have run simulations with different GA probabilities of mutation (PM). Runs with a high PM (0.05) have produced the highest returns for four out of five indexes that have experienced frequent shocks. Thus, to model the turmoil in an economic system that experience frequent shocks, shorter memory depth for recurrent networks and high probability of GA mutation might be considered optimal.

GA model discovery didn't identify higher memory depth as an optimal for long training periods in comparison to short ones. At the same time, the optimal number of hidden layers' neurons was found proportional to the length of training: [5, 7] for short training periods and [6, 15] for longer training durations. Thus, longer training results in the increased complexity of the relationships, where older data is not necessarily useful for the current/future state modeling/forecasting.

For the four memory horizons considered, the most accurate forecasts are with two shorter training durations. Regarding strategies' accuracy all memory lengths are well presented among the best and worst results. In profitability terms, for the whole set of strategies investigated, there is no dominance by strategies with a particular training horizon. Therefore, our results reject a claim that longer training generates more statistically accurate or profitable strategies.

To maximize ANN generalization, with dividing the data into Tr/V/Ts, we have considered a number of distributions. Splitting the data the way presented in Table 2 in comparison to 60%/20%/20% distribution resulted in improved economic performance.

Table 2 Tr/V/Ts distributions

Period	23.01.93-23.01.04	23.07.95-23.01.04	23.01.98-23.01.04	23.07.00-23.01.04
Years	7½/2½/1	5/2½/1	2½/2½/1	2½/½/1
Distribution (%)	64/27/29	59/29/12	42/41/17	57/14/29

Considering rules outperforming B/H benchmark, improvement in annualized return was registered: 6 out of 7 strategies for CAC 40; 6 out of 8 strategies for DAX; 7 out of 10 strategies for AEX; 7 out of 10 for FTSE ALL; 6 out of 8 for NASDAQ 100; 5 out of 11 for S&P 100 and 7 out of 8 for RUSSELL 2000. Thus, financial modeling and forecasting with CI tools benefit in profitability terms with some fine-tuning of Tr/V/Ts distribution.

12. CONCLUSION

The results of our experiment contradict the assertions that agents' long-term fitness is not a function of an accurate prediction, but only depends on an appropriate risk aversion, through a stable saving rate. 'Optimal' locked-up saving contracts would not, therefore, substitute a search for profitable strategies.

To model the turmoil in an economic system with frequent shocks, short memory horizons are considered optimal, as older data is not necessarily informative for the current/future state modeling/forecasting.

In thinner markets there is higher likelihood that short memory horizons agents take the dominant position, influencing price movements. Thin markets' dominance by a particular traders' type facilitates a better environment for learning with CI tools of profitable strategies, used by dominant traders.

REFERENCES

Blume, L. and D. Easley (1992). "Evolution and Market Behavior." Journal Economic Theory 58: 9-40.

Chen, S.-H. and Y.-C. Huang (2003). Simulating the Evolution of Portfolio Behavior in a Multiple-Asset Agent-Based Artificial Stock Market. The 9th International Conference on Computing in Economics and Finance, University of Washington, Seattle, USA.

Chen, S.-H. and C.-C. Liao (2002). Price Discovery in Agent-Based Computational Modeling of the Artificial Stock Market. Genetic Algorithms and Genetic Programming in Computational Finance. S.-H. Chen, Kluwer Academic Publishers: 335-356.

Sweeney, R. J. (1988). "Some Filter Rule Tests: Methods and Results." Journal of Financial and Quantitative Analysis 23: 285-301.