# Setting up Performance Surface of an Artificial Neural Network With Genetic Algorithm Optimization: in Search of an Accurate and Profitable Prediction for Stock Trading

Serge Hayward
Department Finance
Ecole Supérieure de Commerce de Dijon, France
shayward@escdijon.com

**This paper considers a design framework of a computational experiment in finance. The examination of relationships between statistics used for economic forecasts evaluation and profitability of investment decisions reveals that only the 'degree of improvement over efficient prediction' shows robust links with profitability. If profits are not observable, this measure is proposed as an evaluation criterion for an economic prediction. Also combined with directional accuracy, it could be used in an estimation technique for economic behavior, as an alternative to conventional least squares. Model discovery and performance surface optimization with genetic algorithm demonstrate profitability improvement with an inconclusive effect on statistical criteria.**

## I. Introduction

Problems with applications of evolutionary computation (EC) in finance are often due to the lack of common methodology and statistical foundations of its numerous techniques. These deficiencies sometimes cast doubt on conjectured results and conclusions. At the same time, relationships between summary statistics used for predictions' evaluation and profitability of investment decisions based on these predictions are not straightforward in nature. The importance of the latter is particularly evident for applications of an evolutionary / artificial neural network (E/ANN) under supervised learning, where the process of network training is based on a chosen statistical criterion, but when economic performance is generally sought. Motivations for this paper come from the ongoing search for the foundation of EC in finance and a claim by [1] that traditional summary statistics are not closely related to a forecast's profit, with the exception of directional accuracy (DA).

Financial assets' prices often exhibit non-stationarity, autocovarience and frequent structural brakes, posing problems for their modeling. This paper also investigates how data mining benefits from genetic algorithm (GA) model discovery, performance surface optimization and pre/pro-processing, improving predictability or/and profitability.

## II. 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 Figure 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.

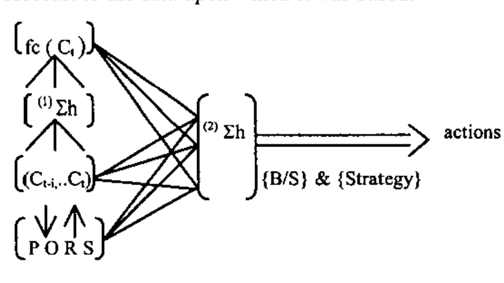


Figure 1. Dual ANN: (1) forecasting network; (2) acting network

This structure is an effort to relate actions' profitability to forecasting quality, examining this relationship in computational settings. The model is evolutionary in the

0-7803-8515-2/04/$20.00 ©2004 IEEE

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 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 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.

### *A. 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.

## III. Description Of The Environment

Let $Y$ be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. $\Omega$ 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:

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

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 evolutionary computation.

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 $\Theta$ to be compact in $R^k$ and observable at time $t$ n-vector of variables, $X_t$. $X_t$ are $\mathcal{F}_t$-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 $X_t$. Under 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 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, $L: R \rightarrow R^+$, related to some economic criteria or a statistical measure of accuracy, takes a general form:

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

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:

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

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

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

Training EANN under different criteria allows us to examine relationships between statistical measures and economic characteristics.

## IV. 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 ads additional explanatory power to the analysis.

*B. Learning Law and Search Algorithm*

Backpropagation is one of the most common algorithms in supervised learning. Although being simple and computationally efficient, the search here can get caught in local minima. Backpropagation is also often criticized for being noisy and slow to converge. To improve the original gradient learning, particularly its slowness of convergence, we consider a number of alternatives.

*C. 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)[2]. In forecasting prices, EP is the last known value[3]. 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.

Making a prediction using a change or a percentage change, the value of IEP is particularly significant. IEP around 1, implying that the ANN predicted a change or a percentage change of zero, indicates that the network does not have adequate information to make a valid prediction. So, it ends up predicting the mean of all changes, zero. Predicting two samples or more in advance, one can have reduction in value of IEP (in comparison to one sample prediction). This does not mean that there is an improvement, since the change in the desired value is typically larger for a longer prediction. We classify our results using the following scale: IEP<0.8 $\Rightarrow$ excellent; IEP<0.85 $\Rightarrow$ very good; IEP<0.9 $\Rightarrow$ good; IEP<0.95 $\Rightarrow$ satisfactory; IEP$\geqslant$0.95 $\Rightarrow$ weak.

*D. Profitability as Performance Measure*

Similar to the performance evaluation criteria of investment managers (total realized returns adjusted for the riskness) the realized total continuously compounded returns or excess returns have been used to review trading rules developed under evolutionary learning. Unlike case-by-case evaluation of actions of portfolio managers, decisions of evolutionary agents are assessed on aggregate, over the entire trading period. Therefore, in computational modeling process/means used by agents need to be explicitly evaluated. Under continuously compounded reinvestment of realized returns, strategies with a higher number of trades and lower returns per trade receive greater fitness. [3] demonstrates that strategies with the lowest mean returns and variances per trade could be evaluated as best.

Simple aggregate realized returns overcome problems with frequent trading. Although the number of trades minimization favors infrequent but prolonged positions. More importantly, realized returns ignore opportunity costs (non-realized losses from missing profitable opportunities), incurred maintaining a certain market position. A proposed solution here is to use non-realized simple aggregate returns.

We examine the following forms of cumulative and individual trades' return measures: non-realized simple aggregate return; 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 frequency and the length of trades, including out of market position. To assess risk exposure we adopt the Sharpe ratio[4] and the maximum drawdown[5], as well as

[1] [2] 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.

[2] Note, market efficiency testing is not an objective of our studies per se. However, learning a profitable forecast/strategy is, in a way, discovering market inefficiency.

[3] If prices exhibit random walk behaviour, equally likely to change up or down, the average forecast has a change of zero from the last value. This makes the last value a good benchmark to determine if the prediction can improve on a random chance.

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

[5] Accesses the size of the individual losses occurred while achieving given gains.

common 'primitive' statistics. 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 [4] 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, $\delta(fc_{t+l})$, assesses the success in establishing a predictive density, $g_y$ that determines agents' actions.

### *E. Time Horizons and Trading Strategies Styles*

Heterogeneous traders in the experiment use different lengths of past and forward time horizons to build their forecasts/strategies. We have run the experiment on stock indexes from a number of markets and found that 'optimal' length of training/validation period is a function of specific market conditions. In this paper we adopt three memory time horizons, [6; 5; 2½] years. We run the experiment with one year testing horizon, as it seems to be reasonable from the actual trading strategies perspective and supported by similar experiments.

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 the previous trades. If the account no longer has enough capital to cover TC, trading stops.

## V. GENETIC ALGORITHM 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 the performance of the following ANN models: Multilayer Perceptron (MLP), Jordan and Elman Networks (J/E), Time-Lag Recurrent Network (TLRN), Recurrent Network (RN), Modular Network (MN) and Support Vector Machine (SVM). We examine the performance surface optimized with GA for DA, discounting the least recent values and minimizing the number of large errors. For learning rule optimization we consider Steepest Descent; Conjugate Gradient; Quickprop; Delta Bar Delta and Momentum.

With GA optimization we test the integer interval [1, 20] for hidden layers' neurons, expecting that a higher number increases the network's learning ability, although at the expense of harder training and a tendency to overspecialization. GA optimization considers the range [1, 20] for the number of taps, affecting the memory of the net. The input layer, having access to the least modified data, has typically the highest number, decreasing in the hidden layers. GA optimization of the weight update for static networks considers whether the weights are updated following all data (batch) or after each piece of data (online) are presented. For dynamic networks GA determines a number of samples to be examined each time ANN updates weights during the training phase.

The step size, controlling the speed of weight adjustment, manages the trade-off between slow learning and a tendency to overreact. Usually the hidden layer has a larger step size than the output layer, and memory components generally have lower step size than other components of the same layer. GA optimizes the step size of the learning rates in the range [0, 1]. The momentum, using the recent weight update, speeds up the learning and helps to avoid local minima. GA searches in the range [0, 1] for the value by which the most recent weight update is multiplied.

In terms of GA parameters, we apply the tournament selection with size 4, {prob=fitness/Σfitness}. Four types of mutation are considered in the experiment: uniform, non-uniform, boundary and Gaussian. Probability of mutation (PM) tested in the range [0, 0.5] and probability of uniform crossover is examined in the range [0.7, 0.95]. We test the effect of the increase in population size in the range [25, 200] on performance and computational time. The training optimization continues until a set of termination criteria is reached, given by maximum generations in the range [100, 500].

When a model lacks information, trading signals' predictions often stay near to the average. If ANN output remains too close to the mean to cross over the thresholds that differentiate entry/exit signals, post-processing is found to be useful (establishing thresholds within the range). Post-processing with GA optimization, examines a predicted signal with simulated trades after each training, searching for the thresholds against the values produced by ANN to generate maximum profit (see Appendix for details).

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 [6].

[6] Another possibility would be to use profit as the performance surface determinant. We leave this option out, since it wouldn't allow us to consider the questions proposed at the beginning. Setting the performance surface determined by the overall objective would not guarantee minimization of the

## VI. Empirical Application

### F. Data

We consider daily closing prices for the MTMS (Moscow Times) share index obtained from Yahoo Finance. The time period under investigation is 01/01/97 to 23/01/04. There were altogether 1575 observations in row data sets. Examining the data graphically reveals that the stock prices exhibit a prominent upward, but non-linear trend, with pronounced and persistent fluctuations about it, which increase in variability as the level of the series increases. Asset prices look persistent and close to unit root or non-stationarity. Descriptive statistics confirm that the unit-root hypothesis cannot be rejected at any confidence level. The data also exhibits large and persistent price volatility with significant autocovarience even at high order lags.

Changes in prices increase in amplitude and exhibit clustering volatility. The daily return displays excess kurtosis and the null of no skewness is rejected at 5% critical level. The tests statistics lead to rejection of the Gaussian hypothesis for the distribution of the series. It confirms that high-frequency stock returns follow a leptokurtic and skewed distribution incompatible with normality assumed often in the analytical literature.

### G. Experimental Results

ANN with GA optimization was programmed with various topologies[7]. Altogether we have generated and considered 93 forecasting and 143 trading strategies' settings. Effectiveness of search algorithm was examined with multiple trials for each setting. 92% 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 120 minutes on a Pentium 4 processor.

Over a one year testing period 19 trading strategies were able to outperform economically the B/H strategy, with an investment of $10,000 and a TC of 2% of trade value. The average return improvement over B/H strategy was 20%, with the first five outperforming the benchmark by 50% and the last three by 2%. The primary strategy superiority over B/H strategy was 72%.

For the five best performing strategies, the break-even TC was estimated to be 2.75%, increasing to 3.5% for the first three and nearly 5% for the primary strategy. Thus, the break-even TC for at least primary strategy appears to be high enough to exceed actual TC. Profitability produced by our simple architecture supports computational model development based on economic and statistical foundations.

The experiment demonstrates that normalization reduces the effect of non-stationarity in the time series. The effect of persistency in prices diminishes with the use of the 'percentage change' in values. Table 1, presenting the average effect of GA post-processing on performance, shows that it has generally improved (positive values) statistical characteristics. Although only accuracy[8] exhibits sizable change, the effects on IEP and correlation[9] were significantly smaller and not always positive.

TABLE 1. GA POST-PROCESSING EFFECT

| ΔStats./Sets | 2000-2004 | 1998-2004 | 1997-2004 |
|---|---|---|---|
| IEP | 0.059 | -0.838 | 0.001 |
| Accuracy (%) | 1.3 | 6.58 | 0.95 |
| Correlation | 0.016 | 0.011 | 0.001 |

The experiment with four types of GA mutation did not identify the dominance by a particular type. We have run simulations with different PM to test how the frequency of novel concepts' arrival affects modeling of the environment with structural brakes. The results, presented in Table 2, show that newcomers generally benefit the system. Although we have expected this outcome, its consistency among all (including short time) horizons was not anticipated. In economic terms, runs with a high probability of mutation {PM=0.5} have produced the highest returns. At the same time, this relationship is of non-linear character (e.g. {PM=0.001} consistently outperforms {PM=0.2}).

Some moderate, although consistent relationship between PM and strategies' risk exposure was found. Higher PM resulted in low riskness, given particularly by Sharpe ratio. We have also noticed some positive correlation between PM and annual trades' quantity, although this relationship appears to be of moderate significance and robustness. Trading frequency in simulations without mutation seems to be set at the beginning and stay until the end either at low or high values. The experiments without mutation have produced strong path-dependent dynamics, though not necessarily with sub-optimal outcome. It seems there exist some 'optimal' PM (in our experiment 0.5 and 0.001) and tinkering with this parameter can improve overall profitability.

We have not found a robust relationship between the

underlying loss function used in forecasting and would not permit us to examine the relationship between statistical qualities and economic profitability.

[7] Programs in Visual C++, v. 6.0 are available upon request. We have run tests on TradingSolutons, v. 2.1, NeuroSolutions v. 4.22 and Matlab v. 6.

[8] Percentage of correct predictions.

[9] Correlation of desired and ANN output.

TABLE 2. ECONOMIC AND STATISTICAL MEASURES UNDER DIFFERENT PROBABILITIES OF MUTATION

| Measures/PM | 0 | 0.001 | 0.2 | 0.5 | 0 | 0.001 | 0.2 | 0.5 | 0 | 0.001 | 0.2 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Return (%) | 76.9 | 85.7 | 76.4 | 99.8 | 65.6 | 75.1 | 62.1 | 86.8 | 68.3 | 74.7 | 60.8 | 82 |
| Sharpe Ratio | 0.13 | 0.15 | 0.15 | 0.16 | 0.13 | 0.13 | 0.14 | 0.16 | 0.13 | 0.13 | 0.13 | 0.14 |
| Trades (N°) | 1 | 3 | 3 | 5 | 9 | 1 | 5 | 10 | 7 | 1 | 4 | 3 |
| IEP | 1.116 | 1.126 | 1.169 | 1.135 | 0.949 | 0.95 | 0.958 | 0.936 | 0.942 | 1.076 | 1.077 | 0.979 |
| Accuracy (%) | 51.5 | 32.9 | 37.66 | 54.98 | 41.2 | 45.92 | 40.77 | 42.06 | 32.38 | 32.9 | 32.9 | 32.4 |
| Data Sets | 2000-2004 | | | | 1998-2004 | | | | 1997-2004 | | | |

memory length and PM>0. Although, the memory length in simulations without mutation was on average 2.5 times shorter than in experiments with mutation. The relationship between PM and common statistical measures was inconclusive at acceptable significance or robustness.

GA model discovery reveals that MLP and TLRN with (Focus) Laguarre memory, with neurons number in the hidden layer in the range [5, 12] and Conjugate Gradient learning rule generate the best performance in statistical and economic terms for forecasting and acting nets. Generally models discovered with GA have lower trading frequencies, but without reduction in riskness. Annualized returns of those models were improved moderately. The effect of GA discovery on models' statistical performance was not conclusive, with a weak tendency towards accuracy amelioration. An increase in population size for GA optimization didn't lead to improvement in results. We explain this by the non multi-modal nature of our problem. Evidently, a higher population size has resulted in longer computational time.

The relationship between statistical measures (accuracy, correlation, IEP) and trading strategies' profitability seems to be of a complicated nature. Among the ten statistically sound price forecasts, there is only one that was used in a trading strategy superior to B/H benchmark. The best five economically strategies are among the worst 50% according to their accuracy. Three of the most accurate strategies are among the worst 25% in terms of their annualized return. Correlation of desired and ANN output characterizes one of the first five strategies with highest return among its best performers, another one among its worst results and the remaining are in the middle. IEP shows some robust relationships with annualized return. All five strategies with highest return have IEP<0.9. Furthermore, one of the first five profitable strategies has one of the three best IEP values. Therefore, if profits are not observable, IEP could be used as an evaluation criterion for an economic prediction.

Regarding the performance surface optimization, two out of the three best strategies included an adjustment to treat directional information as more important than the raw error. We found that training ANN with the performance surface genetically optimized for DA, discounting least recent values or minimizing number of large errors generally improves profitability. Among 25% of the economically weak strategies' annualized returns, there is none with learning criteria optimized. Our experiment has shown that among three optimizations of the performance surface considered, strategies trained on learning the sign of the desired output were generally superior to those trained to reduce the number of large errors or focusing learning on recent values. At the same time, the impact of optimization for DA on common statistical measures was insignificant, conforming that DA only weekly relates to conventional statistical criteria.

Our simulation generally supports a claim that DA relates to forecast profits more than mean squared or mean absolute errors criteria. At the same time, the experiment rejects an assertion that all other summary statistics are not related to forecast profit, as was demonstrated by the IEP relationship with profitability. As the results show that DA (alone or always) does not guarantee profitability of trading strategies trained with this criterion, it might be ineffective to base empirical estimates of economic relationships only on that measure. If conventional least squares are to be considered inadequate, an alternative estimation technique for economic behavior might use a combination of measures, demonstrated to have certain relationships with profitability; IEP and DA have been identified so far.

## VII. Conclusion

Profitability results produced by our simple architecture seem to be sufficiently general to support computational model development, based on economic and statistical foundations. The break-even TC, for at least primary strategy, appears to be high enough to exceed actual TC.

GA post-processing has generally improved statistical characteristics. Novel concepts' arrival, determined by PM, benefits the system in economic terms, but is inconclusive statistically. It seems there exist some 'optimal' PM and tinkering with this parameter has a positive effect on profitability.

Models discovered with GA have moderately higher profitability, but the impact on their statistical characteristics was inconclusive. GA optimization of performance surface (particularly for DA) has a positive effect on strategies' profitability, though with little impact on their statistical characteristics. Since DA does not guarantee profitability of trading strategies trained with this criterion, it might be ineffective to base empirical estimates of economic relationships only on that measure.

When profits are not observable, IEP is proposed as an evaluation criterion for an economic prediction, due to its robust relationships with annualized returns. If conventional least squares are to be considered inadequate, an alternative estimation technique for economic behavior might use a combination of measures, demonstrated to have certain relationships with profitability; IEP and DA have been identified so far.

The performance surface set-up is viewed to be a crucial factor in search of a profitable prediction with an evolutionary model. Measures of trading strategies' predictive power might significantly differ from criteria leading to its profit maximization. The choice of evaluation criteria combining statistical qualities and economic profitability is viewed as essential for an adequate analysis of economic structures.

Presence of at least two objectives (statistical and economic) to be satisfied at the same time could be considered as a multiobjective optimization problem for further research. It seems, evolutionary algorithms, capable generating the Pareto optimal set in a single run, might be particularly appropriate for this task.

## APPENDIX

Use of thresholds within the ranges against values produced by ANN ($\in [-1, 1]$) allows us to set different levels for predicted signals. For enter Long outputs the range is $\{\geq 0.5\}$, with scaling based on the distance between the enter Long and enter Short thresholds (enter Long and zero if thresholds are equal). Exit Short range is $[0.2, 0.5]$, with scaling based on the distance between the exit Short and enter Long thresholds. For exit Long the range is $[-0.2, -0.5]$, with scaling based on the distance between the exit Long and enter Short thresholds. Enter Short range is $\{\leq -0.5\}$, with scaling based on the distance between the enter Long and enter Short thresholds (enter Short and zero if thresholds are equal). For Hold outputs the range is $[-0.2, 0.2]$, with scaling based on the distance between the exit Short and exit Long thresholds.

## REFERENCES


[1] G. Leitch and E. Tanner, "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, vol. 81, pp. 580-590, 1991.

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

[3] S. Bhattacharyya and K. Mehta, "Evolutionary Induction of Trading Models," in *Evolutionary Computation in Economics and Finance, Studies in Fuzziness and Soft Computing*, S.-H. Chen, Ed.: Physica-Verlag, 2002, pp. 311-331.

[4] R. J. Sweeney, "Some Filter Rule Tests: Methods and Results," *Journal of Financial and Quantitative Analysis*, vol. 23, pp. 285-301, 1988.