Performance Measures in an Evolutionary Stock Trading Expert System

Piotr Lipinski[1, 2] and Jerzy J. Korczak[1]

[1] Université Louis Pasteur, LSIIT, CNRS, Strasbourg, France
[2] University of Wroclaw, Institute of Computer Science, Wroclaw, Poland
{lipinski, jjk}@dpt-info.u-strasbg.fr

Abstract. This paper addresses the problem of investment assessment and selection. A number of various performance measures are evaluated and studied. The goal of these investigations is to compare these performance measures on real-life data and to discover an optimal performance measure for selecting investment strategies in an evolutionary stock trading decision support system. Evaluations have been performed on financial time series from the Paris Stock Exchange.

1 Introduction

A trader on the stock market deals with the problem of selecting one of a number of possible investments. The trader evaluates the performance of investments, and decides to invest in the one whose performance is the highest. The performance measure depends on the trader's profit preferences, risk aversion and utility function.

Traders are interested in the profitability of their investments. Since the future return rates are unknown, traders must estimate them on the basis of a number of past observations. Certainly, traders also estimate the risk related to achieving the desired returns [4]. The expected return rate and the risk constitute the main factors of performance measures.

There are a large number of performance measures [1, 2, 12, 13] coming from several well-known stock market models. Each model makes some assumptions on, among other things, investment return distributions. In practice, the stock market does not perfectly suit any of these models, so the prediction may fail.

Traders use models and their theoretical background as the means to achieve profits. They are particularly interested in obtained return rates. This does not mean that traders are interested in high **expected** return rates – it means that traders are interested in high **achieved** return rates. The **expected** return rate is the rate estimated *a priori*, before a trader begins playing on the stock market over a specific time period. Certainly, it depends on the stock market model assumed. The **achieved** return rate is the rate evaluated *a posteriori*, after a trader completes playing on the stock market over a specific time period. It does not depend on the stock market model assumed because it is evaluated *a posteriori* on the basis of known return rates.

A trader selects an investment according to a chosen performance measure. The pragmatic question is how the performance relates to return rates achieved [8, 9]. Is the performance measure appropriate?

In this paper, a number of various performance measures are investigated and the relation between the evaluated performance and return rates achieved is studied. The goal of these investigations is to compare these performance measures on real-life data and to discover the optimal performance measure for selecting investment strategies in an evolutionary stock trading decision support system [6, 7]. This is crucial to the efficiency and the quality of elaborated trading expertise. Experiments were performed on financial time series from the Paris Stock Exchange.

This paper is structured in the following manner: Section 2 defines the problem. Section 3 presents first investigations on selecting one of two possible investments. Section 4 discusses applying various performance measure to four financial time series, which include price quotations of four stocks from the Paris Stock Exchange. Section 5 presents an evolutionary stock trading expert system. Section 6 discusses applying various performance measures in the system. Section 7 concludes the paper.

2 Problem Definition

At time t, a trader evaluates performance of all possible financial assets, and decides to invest in an asset A whose performance $P_t(A)$ is the highest. He may also estimate *a priori* a return rate $r^{(e)}_{t'}(A)$ of the asset A over the future time period (t, t'). Later, at time t', the trader evaluates *a posteriori* a return rate $r_{t'}(A)$ of the asset A achieved over the recently-passed time period (t, t') and compares it to the expected return rate estimated *a priori*. Let $\delta_{t, t'}(A) = r_{t'}(A) - r^{(e)}_{t'}(A)$.

Even though the asset A had been selected because its performance $P_t(A)$ was the highest, sometimes either $r_{t'}(A)$ was average in comparison with return rates achieved by other investments, or $\delta_{t, t'}(A)$ was too large with respect to the risk estimated. If this occurs too often, the chosen performance measure may be inappropriate.

In this paper, classic performance measures are investigated, such as the Sharpe ratio (the ratio of the return rate to the standard deviation of the return rate), the Treynor ratio (the ratio of the return rate to the beta coefficient), the Jensen alpha, performance measures with asymmetric preferences, such as the Sortino ratio (the ratio of the return rate to the SSD), the ROAS (the ratio of the return rate to the AS), the ROPS (the ratio of the return rate to the PS), as well as practitioner performance measures, such as the Calmar ratio (the ratio of the return rate to the MMD) and the Sterling ratio (the ratio of the return rate to the average MMD), where

$$SSD = \sqrt{\text{avg}[(t-r)^2 : t > r)]}, \ AS = \text{avg}[(t-r) : t > r)], \ PS = 13/20,$$

MMD is the maximum drawdown (i.e. maximum sustained drop over relevant period) and alpha and beta come from the $CAPM$ model.

The first goal is to compare these performance measures using the financial time series from the Paris Stock Exchange. The second goal is to discover the optimal performance measure for selecting investment strategies in an evolutionary stock trading decision support system.

3   Selecting One of Two Investments

Let A and B be two available investments. At time t, a trader selects one of them according to a performance measure P. Later, at time t', the trader evaluates *a posteriori* the return rate $r_{t'}(A)$ and $r_{t'}(B)$ of both investments achieved over the recently-passed time period $(t, t']$ and checks whether the investment with the higher performance has also the higher return rate.

Such experiments were performed using two stocks, AXA and Peugeot, for various t and t' in a period from January 4, 1999 to November 7, 2003, using the Sharpe ratio as the performance measure P.

When AXA had a higher P_t (546 observations), the return rate of AXA was higher than the return rate of Peugeot in 299 observations. When Peugeot had a higher P_t (653 observations), the return rate of Peugeot was higher than the return rate of AXA in 369 observations. That gives the accuracy of the Sharpe ratio $\eta = (299 + 369) / (546 + 653) = 0.5571$. That means that the Sharpe ratio led to the correct investment selection in 55.71% of observations.

In further experiments, the trader selects an investment only if the difference between its performance and the performance of the other investment is significant, i.e. it is greater than a given threshold θ. For $\theta = 0.1$, the accuracy of the Sharpe ratio is $\eta = (137 + 233) / (243 + 384) = 0.5901$. For $\theta = 0.2$, the accuracy of the Sharpe ratio is $\eta = (53 + 146) / (97 + 232) = 0.6049$. Results of the other experiments are presented in Table 1.

Table 1. Accuracy of performance measures η for selecting one of two investments

P	θ	η	P	θ	η
Sharpe ratio	0.00	0.5571	Sortino ratio	0.00	0.5588
Sharpe ratio	0.10	0.5901	Sortino ratio	0.10	0.5938
Sharpe ratio	0.20	0.6049	Sortino ratio	0.20	0.6197
Sharpe ratio	0.30	0.6929	Sortino ratio	0.30	**0.8571**
Sharpe ratio	0.40	**0.7813**	ROAS	0.00	0.4395
Treynor ratio	0.00	0.5388	ROAS	0.20	0.3786
Treynor ratio	0.25	0.5518	Sterling ratio	0.00	0.5613
Treynor ratio	0.50	0.5780	Sterling ratio	0.15	0.6505
			Sterling ratio	0.30	**0.7273**

Table 1 shows that relation between certain performance measures and return rates is weak (e.g. the ROAS ratio). However, some performance measures, such as the Sharpe ratio, the Sortino ratio and the Sterling ratio, may be used to select one of two investments if a proper threshold θ is defined.

4   Performance Measures on Financial Time Series

Let P be a specified performance measure. Let A be a financial asset. At time t, a trader evaluates a performance $P_t(A)$ and decides to invest in A. He also estimates *a priori* a return rate $r^{(e)}_{t'}(A)$ over the future time period $(t, t']$. Later, at time t', the

trader evaluates *a posteriori* the return rate $r_{t'}(A)$ achieved over the recently-passed time period $(t, t']$ and compares it to the expected return rate $r^{(e)}_{t'}(A)$ estimated *a priori*. Let $\delta_{t, t'}(A) = r_{t'}(A) - r^{(e)}_{t'}(A)$.

Such experiments were performed using four stocks from the Paris Stock Exchange, AXA, Credit Lyonnais, Peugeot and STMicroelectronics, for various t and t' in a period from January 4, 1999 to November 7, 2003 using various performance measures. A set of observations, consisting of $P_t(A)$, $r^{(e)}_{t'}(A)$, $r_{t'}(A)$ and $\delta_{t, t'}(A)$, was obtained for each stock and each performance measure. Since the results were quite similar, only experiments concerning AXA are presented in detail.

Figure 1 shows how $\delta_{t, t'}$ is related to the performance measure values P_t. Although no direct relation is evident, some dependencies may be observed. For instance (Figure 1a), one can notice that investments with a Sharpe ratio of $P_t < 0.02$ generally have a larger $\delta_{t, t'}$ than investments with a Sharpe ratio of $P_t > 0.02$. Thus, traders should rather focus on investments with a Sharpe ratio of $P_t > 0.02$. The remaining question is whether greater values of the Sharpe ratio P_t entail smaller values of $\delta_{t, t'}$. In general, the answer is no because, for instance, the spread of $\delta_{t, t'}$ seems to be quite similar for $P_t = 0.02$ and $P_t = 0.03$.
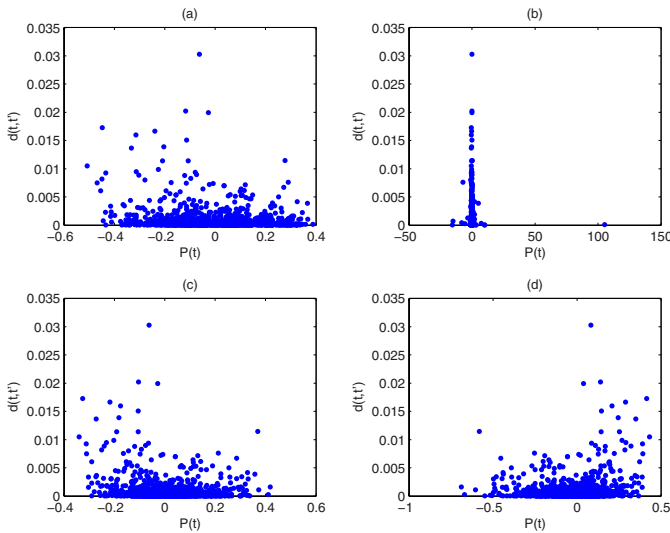

Fig. 1. Dependency of the difference d(t, t') between the achieved return rate and the expected return rate on the performance P(t) (a – Sharpe ratio, b – Treynor ratio, c – Sortino ratio, d - ROAS)

Moreover, experiments similar to those from the previous section were carried out. For each two observations, their performances and their achieved return rates were compared in order to verify whether the investment with the higher performance has also the higher return rate. Results of these experiments are presented in Table 2.

Table 2. Accuracy of performance measures η for selecting one of two investments

P	θ	η	P	θ	η
Sharpe ratio	0.00	0.5517	Sortino ratio	0.15	0.5930
Sharpe ratio	0.20	0.5974	Sortino ratio	0.30	0.6567
Sharpe ratio	0.40	0.6761	Sortino ratio	0.45	**0.7080**
Sharpe ratio	0.60	**0.7676**	ROAS	0.00	0.4496
Treynor ratio	0.00	0.5447	ROAS	0.10	0.4305
Treynor ratio	0.25	0.5713	Sterling ratio	0.00	0.5512
Treynor ratio	0.50	0.5872	Sterling ratio	0.15	0.6215
Sortino ratio	0.00	0.5509	Sterling ratio	0.30	**0.7454**

Table 2 conforms to results presented in the previous section. The Sharpe ratio, the Sortino ratio and the Sterling ratio are more accurate than the other measures.

Although the direct dependence $r_{t'}$ on P_t were not evident, it is interesting to estimate and compare the conditional probabilities $\mathbf{P}[r_{t'} > \theta_r \mid P_t \le \theta_P]$ and $\mathbf{P}[r_{t'} > \theta_r \mid P_t > \theta_P]$ for certain thresholds θ_P and θ_r. Unfortunately, it requires some assumptions on return rate distributions, which might be hard to verify. However, frequencies of occurrences of observations with high $r_{t'}$ in a set of observations with high P_t were investigated. Let $\mathbf{U}$ denote the set of all observations. Let $\mathbf{U'}$ denote a set of observations with high P_t (i.e. $P_t > \theta_P$) and $\mathbf{V}$ denote a set of observations with high $r_{t'}$ (i.e. $r_{t'} > \theta_r$). Let $\alpha = |\mathbf{V} \cap \mathbf{U'}| / |\mathbf{U'}|$ and $\alpha' = |\mathbf{V} \cap (\mathbf{U}-\mathbf{U'})| / |\mathbf{U}-\mathbf{U'}|$. Table 3 presents results for various performance measures.

Table 3. Frequency of occurrence of observations with high $r_{t'}$

P	θ_P	θ_r	α	α'
Sharpe ratio	0.30	0.01	0.4063	0.3213
Sharpe ratio	0.30	0.02	0.3750	0.1894
Sortino ratio	0.25	0.01	0.5405	0.3167
Sortino ratio	0.25	0.02	0.4595	0.1859
Sterling ratio	0.20	0.01	0.4483	0.3205
Sterling ratio	0.20	0.02	0.3793	0.1897

Experiments prove that observations with high $r_{t'}$ occur much more frequently in the set of observations with high P_t than in the set of observations with average P_t. Certainly, the frequency is different for different performance measure.

To conclude, none of the performance measures considered in this paper constitute perfect criteria for investment selection. However, they significantly do increase the chance of selecting profitable investments. Applications of these performance measures in an evolutionary stock trading decision support system is presented in next sections.

5   An Evolutionary Stock Trading Expert System

Traders on the stock market observe quotations of stock prices with the aim to sell an item if it tends to lose value, to buy an item if it tends to gain value, and to take no

action in the remaining cases. Traders often assume that future values can be, more or less accurately, predicted on the basis of past observations. Many methods that analyze past data behavior were introduced [3, 11, 14]. These methods, herein referred to as trading rules, attempt to detect trends or discover contexts leading to occurrences of particular events, which for instance might cause a rise or fall in stock prices.

Let K_t denote the knowledge available at time t. This knowledge may represent historical data, previous stock quotations, or other information on the considered stock. The concept of a stock market trading rule may be formalized by a function f, which computes a decision $f(K_t) \in \{-1.0 \equiv sell, 0.0 \equiv do\ nothing, 1.0 \equiv buy\}$ on the basis of the knowledge K_t available at time t. Naturally, the function f may be defined in a variety of ways.

Although the number of trading rules commonly used by financial experts and market traders is large, some rules may be effective during one time period and less effective during another time period, some rules may work well only in combination with others. At time t, each trader bases his decision on a certain set of trading rules. The trader chooses the set of rules according to criteria defined by his preferences, concerning, for instance, expected profit rate and risk aversion. Depending on the choice of rules, the trader may receive different advice. The question is which rules the trader should choose.

Let e be a subset of the entire set of trading rules $\{ f_1, f_2, \ldots, f_N \}$. Such a subset will be referred to as a stock market trading expert. In a natural way, the expert e can be presented as a binary vector of length N. The i-th coordinate of the vector corresponds to the i-th rule, where 0 stands for absence of the rule and 1 stands for presence of the rule.

A result $r_e(t)$ of the expert e at time t is defined as the average of results of trading rules included in the expert. The result $r_e(t)$ close to -1 corresponds to advice to sell, close to 1 to advice to buy, and otherwise indicates no recommendation, i.e. is interpreted as advice to do nothing. Let $d_e(t)$ denote the decision of the expert e at time t.

Since the number of trading rules is large ($N = 350$ in these investigations), the number of possible experts is enormous (equal to $2^N = 2^{350}$), making the process of manual selection impractical, so a computer decision support system is proposed to make this selection.

The optimal expert is chosen by an evolutionary algorithm [5, 10] according to a given performance measure. Certainly, different performance measures lead to different optimal experts. The question is which performance measure should be applied in order to obtain the most profitable experts.

6 Performance Measures in the Evolutionary Expert System

This section discusses applying performance measures presented in Section 2 to assessment and selection of experts in the evolutionary expert system.

Experiments were performed on four stocks from the Paris Stock Exchange, AXA, Credit Lyonnais, Peugeot and STMicroelectronics, for specified t and t' in a period

from January 4, 1999 to November 7, 2003 using various performance measures. For each stock and each performance measure, a set of observations, which consist of P_t and $r_{t'}$ evaluated on 8000 trading experts, was obtained. 4000 of experts were randomly generated, so their values P_t were average, and 4000 were optimized by the evolutionary algorithm with a relevant objective function, so their values P_t were high.

Although the direct dependence $r_{t'}$ on P_t were not evident, frequencies of highly profitable expert occurrences in a set of experts with high P_t were investigated. Let $\mathbf{U}$ denote the entire set of all 8000 experts. Let $\mathbf{U'}$ denote a set of experts with high P_t (i.e. $P_t > \theta_P$, for a certain threshold θ_P) and $\mathbf{V}$ denote a set of experts with high $r_{t'}$ (i.e. $r_{t'} > \theta_r$, for a certain threshold θ_r). Let $\alpha = |\mathbf{V} \cap \mathbf{U'}| / |\mathbf{U'}|$ and $\alpha' = |\mathbf{V} \cap (\mathbf{U-U'})| / |\mathbf{U-U'}|$. Table 4 presents results for various performance measures.

Table 4. Frequency of highly profitable expert occurrences

P	θ_P	θ_r	α	α'
Sharpe ratio	0.10	0.045000	0.9514	0.7481
Sharpe ratio	0.10	0.045250	0.9205	0.5912
Sortino ratio	0.06	0.045000	0.9442	0.7280
Sortino ratio	0.06	0.045250	0.9050	0.6131
Sterling ratio	0.06	0.045000	0.9769	0.7479
Sterling ratio	0.06	0.045250	0.9551	0.6006

Experiments prove that highly profitable experts occur much more frequently in the set of experts with high P_t than in the set of experts with average P_t. Although the frequency is different for different performance measures, none of the performance measures may be used itself as perfectly efficient selection criteria. It may be interesting to build a multi-objective performance measure based on performance measures investigated in this paper.

7 Conclusions

In this paper, a problem of investment assessment and selection was considered. A number of various performance measures were evaluated. Relation between the performance evaluated and return rates achieved was investigated. A few relatively efficient performance measures were applied to select investment strategies in an evolutionary stock trading decision support system.

A large number of experiments were carried out in order to compare various performance measures. All experiments were performed on real-life data from the Paris Stock Exchange. Experiments showed that the most efficient performance measures are the Sharpe ratio, the Sortino ratio and the Sterling ratio. However, the efficiency may be slightly different over different time periods. Therefore, the presented hypotheses have to be also tested on other markets with variable time horizon.

Although the direct relation between evaluated performance and future return rates is weak and none of the considered measures may be used itself as a perfectly

efficient selection criterion, several of them significantly increase the chance of selecting a highly profitable investment. Thus, they may be used to build a common multi-objective performance measure. Further research on these issues may lead to valuable results.

References

1.   Aftalion, F., Poncet, P., "Les Techniques de Mesure de Performance", Economica, 2003.
2.   Cantaluppi, L., Hug, R., "Efficiency Ratio: A New Methodology for Performance Measurement", Journal of Investing, 2000, pp.19-25.
3.   Colby, W., Meyers, T., "The Encyclopedia of Technical Market Indicators", Down Jones-Irwin, 1990.
4.   Jensen, M., "Risk, the pricing of capital assets, and the evaluation of investment performances", Journal of Business, 1968, pp.167-247.
5.   Goldberg, D. E., "Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley, 1989.
6.   Korczak, J., Lipinski, P., Roger, P., "Evolution Strategy in Portfolio Optimization", Artificial Evolution, ed. P. Collet, Lecture Notes in Computer Science, vol. 2310, Springer, 2002, pp.156-167.
7.   Korczak, J., Roger, P., "Stock timing using genetic algorithms", Applied Stochastic Models in Business and Industry, 2002, pp.121-134.
8.   Lehman, B., Modest, D., "Mutual Fund Performance Evaluation: a Comparison of Benchmarks and Benchmarks Comparison", Journal of Finance, 1987, pp.233-265.
9.   Lo, A., W., "The Statistics of Sharpe Ratios", Financial Analysts Journal, 2002, pp.36-52.
10. Michalewicz, Z., "Genetic Algorithms + Data Structures = Evolution Programs", Springer Verlag, New York, 1994.
11. Murphy, J., "Technical Analysis of the Financial Markets", NUIF, 1998.
12. Sharpe, W., F., "Mutual Fund Performance", Journal of Businness, 1966, pp.119-138.
13. Sharpe, W., F., "The Sharpe Ratio", The Journal of Portfolio Management, 1994.
14. Weigend, A., S., Gershenfeld, N., A., "Time Series Prediction: Forecasting the Future and Understanding the Past", Addison-Wesley, 1993.