ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/221019882

Making Financial Trading by Recurrent Reinforcement Learning

Conference Paper · September 2007

DOI: 10.1007/978-3-540-74827-4_78 · Source: DBLP

CITATIONS
12

READS
1,294

2 authors:

Francesco Bertoluzzo
Università Ca' Foscari Venezia
7 PUBLICATIONS **27** CITATIONS
SEE PROFILE

Marco Corazza
Università Ca' Foscari Venezia
56 PUBLICATIONS **270** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Benford's law as a hidden behavior detection rule View project

All content following this page was uploaded by Marco Corazza on 20 May 2014.

The user has requested enhancement of the downloaded file.

Making Financial Trading by Recurrent Reinforcement Learning

Francesco Bertoluzzo[1] and Marco Corazza[2, 3]

[1] University of Padua, Department of Statistics,
Via Cesare Battisti 241/243, 35121 Padua, Italy
fbertoluzzo@stat.unipd.it
[2] University Ca' Foscari of Venice, Department of Applied Mathematics,
Dorsoduro 3825/E, 30123 Venice, Italy
corazza@unive.it
[3] School for Advanced Studies in Venice Foundation,
Dorsoduro 3488/U, 30123 Venice, Italy

Abstract. In this paper we propose a financial trading system whose strategy is developed by means of an artificial neural network approach based on a recurrent reinforcement learning algorithm. In general terms, this kind of approach consists in specifying a trading policy based on some predetermined investor's measure of profitability, and in setting the financial trading system while using it. In particular, with respect to the prominent literature, in this contribution: first, we take into account as measure of profitability the reciprocal of the returns weighted direction symmetry index instead of the wide-spread Sharpe ratio; second, we obtain the differential version of this measure of profitability and obtain all the related learning relationships; third, we propose a procedure for the management of drawdown-like phenomena; finally, we apply our financial trading approach to some of the major world financial market indices.

1 Introduction

When an economic agent invests in financial markets, she/he has to make decisions under uncertainty. In such a context, her/his task consists in coping with financial risk in order to maximize some predetermined measure of profitability. A wide-spread class of tools which are used to support such risky decisions is the one of the financial trading systems (FTSs).

A standard approach which is usually followed to specify a FTS consists:

- in identifying one or more variables (asset prices, ...) related to the time-behaviour of one or more quantities of interest (trading signals, ...);
- in utilizing the current and the past values of these variables to extract information concerning with the future values of the quantities of interest;
- in using these predictions/information to implement a trading strategy by which to make effective trades.

The distinctly operative valence of the FTSs has made them popular from long time among professional investors and practitioners. A lot of book has been devoted to the working utilization of these tools (see, among the various ones, [8]). Nevertheless, in recent years also the academic world has began to recognize the soundness of some of the features related to the FTSs (see, for example, [5]). Moreover, the really enormous – and continuously increasing – mass of collected financial data has led to the development of FTSs whose information extraction processes are based on data mining methodologies (see, for example, [3]).

Alternative approaches to the standard building of FTSs have been proposed both in the operative literature and in the academic ones. Among the approaches belonging to the latest category, in this paper we consider the ones which exploit artificial neural network (ANN) methodologies based on recurrent reinforcement learnings. In general terms, these approaches consist:

- in specifying a trading policy based on some predetermined investor's measure of profitability (in such a manner one avoids to have to identify the quantities of interest, and avoids to have to perform the information extraction concerning with such quantities);
- in setting the frame and the parameters of the FTS while using it (in such a way one can avoid to carry out the off-line setting of the trading system).

Among the first contributions in this research field we recall [7], [6] and [4]. In general, they show that such strategies perform better than the ones based on supervised learning methodologies when market frictions (transaction costs, ...) are considered.

In this paper, with respect to the cited contributions:

- we take into account as measure of profitability the reciprocal of the returns weighted direction symmetry index (see the third section) reported in [1] instead of the wide-spread Sharpe ratio, which is the only one used in the quoted literature;
- we obtain the differential version of the considered measure of profitability and obtain all the new related learning relationships (see the third section);
- we propose a simple procedure for the management of drawdown-like phenomena by which to integrate the considered FTS (see the fourth section);
- we apply our financial trading approach (FTA) (i.e. FTS + drawdown-like phenomenon management) to some of the major worls financial market indices (see the fourth section).

Finally, in the last section we provide some concluding remarks.

2 Recurrent Reinforcement Learning: A Short Recall

In this section we give a short qualitative introduction of the recurrent reinforcement learning.

In general terms, this kind of learning concerns an agent (in our case the FTS) dynamically interacting with an environment (in our case a financial market). During this interaction, the agent perceives the state of the environment and undertakes a related action. In its turn, the environment, on the basis of this action, provides a negative or positive reward (in our case the investor's loss or gain). The recurrent reinforcement learning consists in the on-line detection of a policy (in our case a trading strategy) which permits the maximization over the time of a predetermined cumulative reward (see, for technical details, [9]).

3 The Financial Trading System

In this section, we describe our discrete-time trading strategy and we obtain all the new learning relationships related to the considered measure of profitability.

3.1 The Trading Strategy

Let we start by considering a discrete-time frame $t = 0, \ldots, T$. Our trading strategy at time t, F_t, is based on the sign of the output, y_t, of a suitable ANN:

- if $y_t < 0$ then $F_t = -1$, and one short-sells the considered portfolio;
- if $y_t = 0$ then $F_t = F_{t-1}$, and one does nothing;
- if $y_t > 0$ then $F_t = 1$, and one buys the considered portfolio.[4]

We assume that this strategy depends on the current and past values of the logarithmic rate of return of the portfolio to trade, and on the previous value of the trading strategy itself.

The ANN we consider is a simple no-hidden-layer perceptron model (both the architectural structure and the squashing function are the ones commonly used in the relevant literature):

$$y_t = \tanh\left(\sum_{i=0}^{M} w_{i,t} x_{t-i} + w_{M+1,t} F_{t-1} + w_{M+2,t}\right),$$

where $w_{0,\tau}, \ldots, w_{M+1,\tau}$ are the weights of the ANN at time τ; $x_\tau, \ldots, x_{\tau-N}$ are the current and past values of the logarithmic rate of return of the portfolio to trade at time τ; and $w_{M+2,\tau}$ is the threshold of the ANN at time τ.

As net reward at the generic time period $(t-1,\ t]$ we take into account the following quantity:

$$R_t = \mu\left[F_{t-1} r_t - \delta\left|F_t - F_{t-1}\right|\right],$$

where μ is the amount of capital to invest; r_τ is the geometric rate of return at time τ of the portfolio to trade; and δ is the per cent transaction cost related to the portfolio quota to trade.

[4] F_t plays the role of the action.

It is to notice that, with respect to the cited contributions, in this paper we consider a net reward formulated in terms of rate of return instead of price.

Given the net reward of period, it is easy to define the net cumulative reward from time 1 to time t as $CR_t = \sum_{i=1}^{t} R_i$. This expression for the net cumulative reward is a particularization of the expression $CR_t = \sum_{i=1}^{t} R_i \left(1 + i_f\right)^{t-i}$, where i_f is the free-risk rate of return of period.

Finally, we give the new investor's gain index at time t whose utilization permits the determination – *via* recurrent reinforcement learning – of the optimal values of the weights of the considered ANN:

$$I_t = \frac{\sum_{i=1}^{t} g_i \left|R_i\right|}{\sum_{i=1}^{t} b_i \left|R_i\right|}, \text{ with } \sum_{i=1}^{t} b_i \left|R_i\right| \neq 0, \tag{1}$$

where $g_\tau = \begin{cases} 0 \text{ if } R_\tau \leq 0 \\ 1 \text{ if } R_\tau > 0 \end{cases}$; and $b_\tau = \begin{cases} 1 \text{ if } R_\tau \leq 0 \\ 0 \text{ if } R_\tau > 0 \end{cases}$.

This index, which is the reciprocal of the returns weighted directional symmetry measure reported in [1], is given at each time t by the ratio between the cumulative "good" (i.e. positive) rewards and the cumulative "bad" (i.e. not positive) rewards.

3.2 The Recurrent Reinforcement Learning

The considered ANN is characterized by $M+3$ parameters: $w_{0,t}, \ldots, w_{M+2,t}$. For determining their optimal values we perform an economically founded approach consisting in the maximization of an investor's utility function depending, at time t, on $R_1, \ldots, R_t$. In particular, as (non-standard) utility function we take into account (1). At this point, we determine the optimal values of the considered parameters by using an usual weight updating method based on the following gradient ascent technique:

$$w_{i,t} = w_{i,t-1} + \rho_t \frac{dU_t}{dw_{i,t}}, \text{ with } i = 0, \ldots, M+2, \tag{2}$$

where U_τ is the investor's utility function at time τ; and ρ_τ is a suitable learning rate at time τ.

It is to notice that, in each generic time period $(t-1,\ t]$, the investor is (obviously) interested in the marginal variation of her/his utility function, i.e. in $D_t = U_t - U_{t-1} = I_t - I_{t-1}$. Now, in order to provide the expression for $dU_t / dw_{i,t} = dI_t / dw_{i,t}$, it is to notice that computing I_t becomes as harder as t increases. So, we resort to the following exponential moving formulation of (1):

$$\widetilde{I}_t = \frac{A_t}{B_t} \tag{3}$$

where $A_\tau = \begin{cases} A_{\tau-1} \text{ if } R_\tau \leq 0 \\ \eta R_\tau + (1-\eta)\, A_{\tau-1} \text{ if } R_\tau > 0 \end{cases}$ is the exponential moving estimates of the numerator of (1) at time τ; $B_\tau = \begin{cases} -\eta R_\tau + (1-\eta)\, B_{\tau-1} \text{ if } R_\tau \leq 0 \\ B_{\tau-1} \text{ if } R_\tau > 0 \end{cases}$ is the

exponential moving estimates of the denominator of (1) at time τ; and η is an adaptation coefficient.

So, $U_t = \widetilde{I}_t$.

Then, in order to provide an expression for D_t, we act in a way similar to the one used in [7], [6] and [4], i.e.:

- firstly, we consider the expansion of (3) in Taylor's series about $\eta = 0$;
- secondly, we utilize $d\widetilde{I}_t/d\eta\Big|_{\eta=0}$ as approximation of D_t, i.e., after some arrangements, i.e.

$$D_t \cong \left.\frac{d\widetilde{I}_t}{d\eta}\right|_{\eta=0} = \begin{cases} -A_{t-1}\,(R_t + B_{t-1})/\,B_{t-1}^2 \text{ if } R_t \leq 0 \\ (R_t - A_{t-1})/\,B_{t-1} \text{ if } R_t > 0 \end{cases}.$$

At this point, it is possible to prove that (see, for more details, [7], [6] and [4]):

$$\frac{dU_t}{dw_{i,t}} = \sum_{j=1}^{t} \frac{dU_j}{dR_j}\left(\frac{dR_j}{dF_j}\frac{dF_j}{dw_{i,t}} + \frac{dR_j}{dF_{j-1}}\frac{dF_{j-1}}{dw_{i,t-1}}\right)$$

where $dU_\tau/\,dR_\tau = d\,(U_\tau - U_{\tau-1})/\,dR_\tau = \begin{cases} (A_{\tau-1}B_{\tau-1})/\,B_{\tau-1}^3 \text{ if } R_t \leq 0 \\ 1/\,B_{\tau-1} \text{ if } R_t > 0 \end{cases}$ (it is to notice that $dU_\tau/\,dR_\tau = d\,(U_\tau - U_{\tau-1})/\,dR_\tau$ as $U_{\tau-1}$ does not depend on R_τ); $dR_\tau/\,dF_\tau = -\mu\,\delta\,\mathrm{sign}\,(F_\tau - F_{\tau-1})$; $dR_\tau/\,dF_{\tau-1} = P_\tau - P_{\tau-1} - \mu\,\delta\,\mathrm{sign}\,(F_\tau - -F_{\tau-1})$; and $dF_\tau/\,dw_{i,\tau}$, which depends on the chosen squashing function, is easily obtainable (see, for instance, [2]).

Finally, in order to implement (2), we approximate the previous exact relationship, which holds for batch learnings, in the following one, which holds for on-line learnings:

$$\frac{dU_t}{dw_{i,t}} \cong \frac{dU_t}{dR_t}\left(\frac{dR_t}{dF_t}\frac{dF_t}{dw_{i,t}} + \frac{dR_t}{dF_{t-1}}\frac{dF_{t-1}}{dw_{i,t-1}}\right).$$

4 Applications

In this section, at first we present the plan of our experimentation, then we propose a procedure for the management of drawdown-like phenomena, finally we give the results coming from the applications of our FTA.

We apply the considered FTA to 9 of the major world financial market indices, i.e. to some of the ones that at March 31, 2007 were quoted from at least 15 years: Dow Jones Industrials (D), FTSE 100 (F), Jakarta SE Composite (J), Madrid SE General (M), NASDAQ Composite (NA), NIKKEI 225 Stock Average (NI), Shanghai SE Composite (Sh), Swiss Market (Sw), and Tel Aviv SE

General (T). In detail, we consider close daily data from April 1, 1992, to March 31, 2007.

In order to make operative our FTA, we need to determine the optimal values/time-evolution of the parameters M, δ, ρ_t, η and of the one related to the management of drawdown-like phenomena. For carrying out this optimal setting, we perform in a manner similar to the one utilized in [7], [6] and [4], i.e. we articulate the whole trading time period (from 0 to T) in the sequence of overlapping time sub-period $[0,\ \Delta t_{off} + \Delta t_{on}]$, $[\Delta t_{on},\ \Delta t_{off} + 2\Delta t_{on}]$, ..., $[T - \Delta t_{off} - 2\Delta t_{on},\ T - \Delta t_{on}]$, $[T - \Delta t_{off} - \Delta t_{on},\ T]$, where Δt_{off} is the length of the initial part of each time sub-period during which the FTA works in an off-line modality for performing the optimal setting of the considered parameters; and Δt_{on} is the length of the final part of each time sub-period during which the FTA works in an on-line modality for making the financial trading.[5]

By so acting, our FTA performs uninterruptedly from Δt_{off} to T, and in the meantime the parameters are periodically updated.

4.1 The Drawdown-like Phenomenon Management

Any well-working FTS should be able to minimize large losses during its running since these losses could reduce so much the capital at investor's disposal to make impossible the continuation of the trading itself. So, also our FTS should be able to minimize large losses, and in case they occur it should be able to guarantee the continuation of the trading.

In order to minimize large losses, the financial trading is not performed (and is definitively interrupted) at first instant time $t \in \{1, \ldots, T-1\}$ in which the net cumulative reward is negative. In such a case, the loss can be at most $-\mu$.

In order to (attempt to) guarantee the continuation of the financial trading in case a large loss occurs, we utilize as amount of capital to invest $\mu + \mu_0$, where $\mu_0 = \left|\min\left\{\min_{0<t\leq\Delta t_{off}}\{R_t :\ R_t < 0\ \wedge\ CR_t < 0\ \wedge\ F_{t-1}F_t = -1\};\ 0\right\}\right|$, i.e. the absolute value of the largest loss (associated to the occurrence that the net cumulative reward is negative and the trading strategy is changed) happened during the initial part of the first time sub-period. Of course, given such an amount of capital to invest, the loss can be at most $-(\mu + \mu_0)$.

4.2 The Results

In all the 9 applications we have used an amount of capital to invest, μ, equal to 1. As far as regards the optimal setting of the parameters, we have determined that, in general, $M \in \{2, 3, \ldots, 13, 14\}$, $\delta \in \{0.0050, 0.0075, \ldots, 0.0725,\ 0.0750\}$, $\rho_t = 0.01$, $\eta = 0.01$, $\Delta t_{off} = 500$, and $\Delta t_{on} = 65$; of course, μ_0 varies as the investigated financial market index varies. Moreover, in each application we suitably initialize the numerator and the denominator of (1).

In Table 1 we present the results of our applications. In particular, with respect to each investigated financial market index: in the first column we report

[5] Of course, we need also to determine the optimal values of Δt_{off} and Δt_{on}.

the identifier; in the second column we report the CR_T, with $T =$ March 31, 2007, obtained from the application of our FTA; in the third column we report μ_0; in the fourth column we report the largest loss, i.e. $\min_{\Delta t_{off} \leq t \leq T} \{R_t : R_t < 0 \wedge F_{t-1}F_t = -1\}$; in the fifth column we report the CR_T, with $T =$ March 31, 2007, obtained from the application of our FTS (i.e. FTA - drawdown-like phenomenon management);[6] in the sixth column we report the number of performed trades; in the seventh/eighth/ninth column we report the percentages of the performed trades for which $R_t = 0/R_t = 0/R_t > 0$.

Table 1. The results of our applications

ID	FTA-based CR_T	μ_0	Largest loss	FTS-based CR_T	Number of trades	% $R_t < 0$	% $R_t = 0$	% $R_t = 0$
D	0.070	0.000	−0.172	0.043	22	0.455	0.000	0.545
F	0.482	0.003	−0.139	0.545	19	0.579	0.000	0.421
J	0.480	0.000	−0.431	0.480	101	0.436	0.109	0.455
M	0.980	0.000	−0.137	0.782	79	0.430	0.076	0.494
NA	0.229	0.016	−0.402	0.406	86	0.488	0.012	0.500
NI	−1.036	0.036	−0.352	−1.264	86	0.570	0.023	0.407
Sh	0.967	0.048	−0.521	1.295	23	0.478	0.000	0.522
Sw	0.922	0.000	−0.297	0.597	19	0.368	0.000	0.632
T	0.658	0.129	−0.179	0.099	88	0.477	0.091	0.432

As far as concerns these results, it is to notice:

- that only 1 of the 9 investigated financial market indices shows a negative FTA-based CR_T, and that the summation of all the FTA-based CR_Ts is positive and equal to 3.752. That can be interpreted as the fact that, under a suitable diversification in the investments, our FTA is well-working in the long-run term;
- that the magnitude of this summation is not satisfying. It depends both on the choice to utilize in the expression for CR_T a free-risk rate of return, i_f, equal to 0, and on the need to effect the setting of the parameters in some more refined way;
- that our drawdown-like phenomenon management is enough well-performing. In fact, as the summation of the FTA-based CR_Ts is greater than the summation of the FTS-based CR_Ts, our FTA appears able to reduce the losses which occur during its running (although it is not able to guarantee the continuation of the financial trading of the only critical case, the one related to the NIKKEI 225 Stock Average index).

[6] The FTS-based CR_T has to be considered as a comparison term for the FTA-based CR_T. It is to notice that the FTS-based CR_T is sometimes unrealistic because, being lacking the management of the drawdown-like phenomena, the associated financial trading could continue to T even if $CR_t < 0$, with $t \in \{1, \ldots T-1\}$.

View publication stats

5 Concluding Remarks

In this section we provide some concluding remarks. In particular:

- in order to reduce the percentage of performed trades for which $R_t < 0$, we conjecture that could be fruitful to utilize the following so-modified trading strategy:
  - if $y_t < -\varepsilon^-$, with $\varepsilon^- > 0$, then $F_t = -1$;
  - if $\varepsilon^- \leq y_t \leq \varepsilon^+$, with $\varepsilon^+ > 0$, then $F_t = F_{t-1}$;
  - if $y_t > \varepsilon^+$ then $F_t = 1$;[7]
- in order to improve the learning capabilities of the ANN, we guess that could be profitable to use a multi-layer perceptron model (it is to notice that such a check is carried out in [4], but without meaningful results);
- in order to make more informative the set of the variables related to the time-evolution of the quantities of interest, we conjecture that could be fruitful to consider, beyond the logarithmic rate of return, at least the transaction volume;
- finally, in order to explicitly take into account the risk in our FTA, we guess that could be profitable to utilize some risk-adjusted version of (1).

References

1. Abecasis, S.M., Lapenta, E.S., Pedreira, C.E.: Performance metrics for financial time series forecasting. Journal of Computational Intelligence in Finance **July/August** (1999) 5–23
2. Bishop, C.: Neural networks for pattern recognition. Oxford University Press, Oxford (1995)
3. Corazza, M., Vanni, P., Loschi, U.: Hybrid automatic trading system: technical analysis & group method of data handling. In: Marinaro, M., Tagliaferro, R. (eds.): Neural nets. 13th Italian workshop on neural nets, WIRN Vietri 2002. Vietri sul Mare, Italy, May/June 2002. Revised paper. Springer, Berlin (2002) 47–55
4. Gold, C.: FX trading via recurrent reinforcement learning. Proceedings of IEEE international conference on computational intelligence in financial engineering (2003) 363–370
5. Lo, W.A., Mamaysky, H., Wang, J.: Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. Journal of Finance **LV** (2000) 1705–1769
6. Moody, J., Saffell, M.: Learning to trade via direct reinforcement. IEEE Transactions on Neural Networks **12** (2001) 875–889
7. Moody, J., Wu, L., Liao, Y., Saffell, M.: Performance functions and reinforcement learning for trading systems and portfolios. Journal of Forecasting **17** (1998) 441–470
8. Murphy, J.J.: Study guide to technical analysis of the financial markets. Prentice Hall Press, New York (1999)
9. Sutton, R.S., Barto, A.G.: An introduction to reinforcement learning. MIT Press, Cambridge (1997)

[7] Of course, in such a case we should need to determine the optimal values of two other parameters, ε^- and ε^+.