Financial trading systems: Is recurrent reinforcement learning the way?

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

WORKING PAPER SERIES

Francesco Bertoluzzo and Marco Corazza

Financial trading system:
is recurrent reinforcement learning
the *via*?

Financial trading systems:

is recurrent reinforcement learning the *via*?

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

[1]Consorzio Venezia Ricerche
Via della Libertà, 12 − 30175 Marghera (VE), Italy
Phone: (+39) 041 5093019 – Fax: (+39) 041 5093074
fbertoluzzo@yahoo.it

[2]Department of Applied Mathematics − University Ca' Foscari of Venice
Dorsoduro, 3825/E − 30123 Venice, Italy
Phone: (+39) 041 2346911 – Fax: (+39) 041 5221756
corazza@unive.it

[3]School for Advanced Studies in Venice Foundation
Dorsoduro, 3488/U − 30123 Venice, Italy
Phone: (+39) 041 2201242 – Fax: (+39) 041 2201249

Abstract – In this paper we propose a financial trading system whose trading strategy is developed by means of an artificial neural network approach based on a learning algorithm of recurrent reinforcement type. In general terms, this kind of approach consists: first, in directly specifying a trading policy based on some predetermined investor's measure of profitability; second, in directly 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 the measure of profitability we consider, and obtain all the related learning relationships; third, we propose a simple procedure for the management of drawdown-like phenomena; finally, we apply our financial trading approach to some of the most prominent assets of the Italian stock market.

Keywords – Financial trading system, recurrent reinforcement learning, no-hidden-layer perceptron model, returns weighted directional symmetry measure, gradient ascent technique, Italian stock market.

---

[*] Corresponding author.

INTRODUCTION

When an economic agent invests capital in financial markets, she/he has to make decisions under uncertainty. In such a context, her/his task consists in suitably dealing 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, transaction volumes, …) related to the time-behaviour of one or more suitable quantities of interest (trading signals, …);

– in utilizing the current and the past values of these variables to forecast (or, more in general, to extract information concerning with) the future values of the suitable 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 obviously 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, Wilder, 1978, Appel, 1979, Nison, 1991, and Murphy, 1999).

Nevertheless, although only in recent years, also the academic world has began to (partially) recognize the soundness of some of the features related to the FTSs (see, among the first ones, Lee *et al.*, 2000, and Lo *et al.*, 2000). 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, Plihon *et al.*, 2001, and Corazza *et al.*, 2002).

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 learnings of recurrent reinforcement type. In general terms, these approaches consist:

– in directly 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 predictions/information extraction of/concerning with such quantities);

- in directly 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 Moody *et al.*, 1998, Moody *et al.*, 2001, and Gold, 2003. In general, they show that such strategies perform better than the ones based on supervised learning methodologies when market frictions are considered. In Moody *et al.*, 1998, the Authors develop and utilize a recurrent reinforcement learning (RRL) algorithm in order to set an FTS that, taking into account transaction costs, maximizes an appropriate investor utility function based both on the well-known Sharpe ratio and on its differential version. Then, they show by controlled experiments that the proposed FTS performs better than standard FTSs. Finally, the Authors use their FTS to make profitable trades with respect to assets of the U.S.A. financial markets. In Moody *et al.*, 2001, the Authors mainly compare FTSs developed by using RRL methodologies with FTSs developed by using stochastic dynamic programming methodologies. In general, they show by extensive experiments that the former approach is better than the latter one. In Gold, 2003, the Author considers a FTS similar to the one develop in Moody *et al.*, 1998, and apply it to financial high frequency data, obtaining profitable performances.

In this paper, with respect to the cited contributions:

- instead of considering as measure of profitability the Sharpe ratio (which is the only one used in the quoted literature), we take into account the reciprocal of the returns weighted direction symmetry index reported in Abecasis *et al.*, 1999 (see the third section);
- we obtain the differential version of the measure of profitability we consider and obtain all the new related learning relationships (see the third section again);
- 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 most prominent assets of the Italian stock market (see the fourth section again).

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

RECURRENT REINFORCEMENT LEARNING: A SHORT RECALL

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

Generally speaking, 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 positive or negative reward (in our case the investor's gain or loss). 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, Sutton *et al.*, 1997).

It is to notice that, given the well-known stochastic nature of financial markets, detecting such an optimal trading policy is equivalent to solve a suitable stochastic dynamic programming problem. Under this point of view, recurrent reinforcement learning provides approximate solutions to stochastic dynamic programming problems (see, for more details, Bertsekas, 1995, and Bertsekas *et al.*, 1996).

THE FINANCIAL TRADING SYSTEM

In this section, at first we describe our discrete-time trading strategy (see the first sub-section), then we obtain all the new learning relationships related to the measure of profitability we consider (see the second sub-section).

The trading strategy

Let we start by considering a discrete-time frame $t = 0, \ldots, T$. Our trading strategy at time t, *i.e.* F_t, is simply based on the sign of the output, y_t, of a suitable ANN:[2]

- if $y_t < 0$, then $F_t = -1$ and one short-sells the considered stock or 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 stock or portfolio.[3]

---

[2] When possible we maintain the same notation used in Moody *et al.*, 1998, Moody *et al.*, 2001, and Gold, 2003.

We reasonably assume that this trading strategy depends on the current and past values of one or more suitable variables related to the stock or portfolio to trade, and depends on the previous value of the trading strategy itself. In particular, in this paper we consider only one variable, the logarithmic rate of return of the stock or portfolio to trade.

The ANN we consider is a simple no-hidden-layer perceptron model in which the squashing function is the $\tanh(\cdot)$ one (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,

$x_t, \ldots, x_{t-N}$ are the current and past values of the logarithmic rate of return of the stock or portfolio to trade,

and $w_{M+2,\tau}$ is the threshold of the ANN at time τ.

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

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

where

μ is the (prefixed) amount of capital to invest,

r_τ is the geometric rate of return at time τ of the stock or portfolio to trade,

and δ is the (prefixed) per cent transaction cost related to the stock or portfolio quota to trade.

It is to notice:

–   that R_t is the net reward of the generic time period $]t-1,t]$;

–   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 as follows the (net) cumulative reward from time 1 to time t:

---

[3] It is to notice that F_t plays the role of the action.

$$CR_t = CR_{t-1} + R_t = \sum_{i=1}^{t} R_i \,.$$

It is to notice:

– that CR_t denotes the (net) capital gain at time t;

– that the proposed expression for the net cumulative reward is a particularization of the following more general one:

$$CR_t = CR_{t-1} + R_t = \sum_{i=1}^{t} R_i \left(1 + i_{fr}\right)^{t-i} \,.$$

where

i_{fr} is the free-risk rate of return of period.

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

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

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 [Abecasis *et al.*, 1999], at each time t is given by the ratio between the cumulative "good" (*i.e.* positive) rewards and the cumulative "bad" (*i.e.* not positive) rewards.

It is to notice that a well-working trading strategy should guarantee that:

$$\sum_{i=0}^{t} g_i |R_i| > \sum_{i=0}^{t} b_i |R_i|$$

i.e., after some simple arrangements, that:

$$I_t > 1, \ t = 1, \ldots, T \,.$$

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 a well-founded economic approach consisting in the maximization of a suitable investor's utility function depending, at time t, on R_1, …, R_t. In particular, as 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:

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

(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 the investor's 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):

$$\tilde{I}_t = \frac{A_t}{B_t}$$

(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, $U_t = \tilde{I}_t$.

Then, in order to provide an expression for D_t, we act in a way similar to the one used in Moody *et al.*, 1998, Moody *et al.*, 2001, and Gold, 2003, *i.e.*:

– firstly, we consider the expansion of (3) in Taylor's series about $\eta = 0$;

– secondly, we utilize $d\tilde{I}_t / d\eta \big|_{\eta=0}$ as approximation of D_t, *i.e.*, after some arrangements,

$$D_t \cong \frac{d\tilde{I}}{d\eta}\bigg|_{\eta=0} = \begin{cases} -\dfrac{A_{t-1}(R_t + B_{t-1})}{B_{t-1}^2} \text{ if } R_t \leq 0 \\[2mm] \dfrac{R_t - A_{t-1}}{B_{t-1}} \text{ if } R_t > 0 \end{cases}.$$

At this point, it is possible to prove that (see, for more details, Moody *et al.*, 1998, Moody *et al.*, 2001, and Gold, 2003):

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

where

$$\frac{dU_\tau}{dR_\tau} = \frac{d(U_\tau - U_{\tau-1})}{dR_\tau} = \frac{dD_\tau}{dR_\tau} = \begin{cases} \dfrac{A_{\tau-1}B_{\tau-1}}{B_{\tau-1}^3} \text{ if } R_\tau < 0 \\[2mm] \dfrac{1}{B_{\tau-1}} \text{ if } R_\tau > 0 \end{cases}$$ (it is to notice that $\dfrac{dU_\tau}{dR_\tau} = \dfrac{d(U_\tau - U_{\tau-1})}{dR_\tau}$ as

$U_{\tau-1}$ does not depend on R_τ),

$$\frac{dR_\tau}{dF_\tau} = -\mu\delta\,\mathrm{sign}(F_\tau - F_{\tau-1}),$$

$$\frac{dR_\tau}{dF_{\tau-1}} = P_\tau - P_{\tau-1} - \mu\delta\,\mathrm{sign}(F_\tau - F_{\tau-1}),$$

and $\dfrac{dF_\tau}{dw_{i,\tau}}$, which depends on the chosen squashing function, is easily obtainable (see, for instance, Bishop, 1995).

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:

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

APPLICATIONS

In this section, at first we present the experimental plan of our application, then we propose a simple procedure for the management of drawdown-like phenomena (see the first sub-section), finally we give the results coming from the application of our FTA.

We apply the considered FTA to 13 assets of the most prominent ones of the Italian stock market, *i.e.* to all the assets specifying the MIB30 at September 21, 2006 such that, at that date, were quoted from more than 20 years: Alleanza Assicurazioni (AA), Banca Intesa (BI), Banca Popolare di Milano (BPM), Capitalia (C), Fiat (F), Finmeccanica (Fi), Fondiaria-Sai (FS), Generali Assicurazioni (GA), Mediobanca (M), Ras Fraz (RF), Saipem (S), Telecom Italia (TI), and Unicredito Italiano (UI). In detail, we consider close daily data from January 1, 1973, to September 21, 2006.

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 (see the next sub-section). For carrying out this "optimal" setting, we perform in a manner similar to the one utilized in Moody *et al.*, 1998, and in Gold, 2003, *i.e.* we articulate the whole trading time period (from 0 to T) in the following sequence of overlapping time sub-period:

$$[0, \Delta t_{off} + \Delta t_{on}], \; [\Delta t_{on}, \Delta t_{off} + 2\Delta t_{on}], \; \ldots, \; [i\Delta t_{on}, \Delta t_{off} + (i+1)2\Delta t_{on}], \; \ldots,$$

$$[T - \Delta t_{off} - 2\Delta t_{on}, T - \Delta t_{on}], \; [T - \Delta t_{off} - \Delta t_{on}, T]$$

where

Δt_{off} is the length of the initial part of each time sub-period (from $i\Delta t_{on}$ to $\Delta t_{off} + i\Delta t_{on}$) 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 (from $\Delta t_{off} + i\Delta t_{on}$ to $\Delta t_{off} + (i+1)\Delta t_{on}$) during which the FTA works in an on-line modality for making the financial trading.[4]

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

---

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

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 that first instant time $t \in \{1, \ldots, T\}$ in which the net capital gain, *i.e.* CR_t 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

$$\mu_0 = \left| \min \left\{ \min_{0 < t \leq \Delta t_{off}} \{ R_t : R_t < 0 \wedge CR_t < 0 \wedge F_{t-1} F_t = -1 \}, 0 \right\} \right|, \text{ } i.e. \text{ the absolute value of the largest}$$

loss (associated to the occurrence that the net capital gain 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 $-(\mu + \mu_0)$.

The results

In all the 13 applications, we have used an amount of capital to invest, *i.e.* μ, equal to 1. As far as regards the "optimal" setting of the parameters, we have determined that, in general, $M \in \{3, 4, \ldots, 9, 10\}$, $\delta \in \{0.005, 0.010, \ldots, 0.070, 0.75\}$, $\rho_t = 0.01$, $\eta = 0.01$, $\Delta t_{off} = 500$, and $\Delta t_{on} = 50$; of course, μ_0 varies as the investigated stock asset varies. Moreover, in each application we initialize both the numerator and the denominator of (1) to a positive infinitesimally small quantity.

In Table 1 we present the main results of our applications. In particular, we respect to each investigated stock asset:

−   in the first column we report the identifier;

- in the second column we report the CR_T, with $T = $ September 21, 2006, 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.* $\min_{\Delta t_{on} \leq t \leq T} \{R_t : R_t < 0 \wedge F_{t-1}F_t = -1\}$;
- in the fifth column we report as benchmark the CR_T, with $T = $ September 21, 2006, obtained from the application of our FTS (*i.e.* FTA – drawdown-like phenomenon management). It is to notice that this CR_T has to be considered as a (sometimes unrealistic) upper bound for the CR_T reported in the second column because, being lacking the management of the drawdown-like phenomena, the associated financial trading can unrealistically continue also when $CR_t < 0$, with $t = 1, \ldots, T-1$;
- 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 $R_t < 0 / R_t = 0 / R_t > 0$.

ID	CR_T (FTA)	μ_0	Largest loss	CR_T (FTS)	N. of trades	% $R_t < 0$	% $R_t = 0$	% $R_t > 0$
AA	–0.559	0.143	–0.348	–0.683	201	0.468	0.034	0.498
BI	–1.161	0.161	–0.630	–1.785	187	0.503	0.069	0.428
BPM	0.311	0.179	–0.340	0.367	96	0.510	0.073	0.417
C	0.478	0.185	–0.778	0.567	189	0.444	0.112	0.444
F	1.609	0.083	–0.338	1.742	199	0.492	0.021	0.487
Fi	2.701	0.048	–0.415	2.841	147	0.463	0.088	0.449
FS	1.211	0.075	–0.250	1.303	145	0.448	0.049	0.503
GA	–0.076	0.228	–0.238	–0.093	152	0.520	0.033	0.447
M	1.472	0.146	–0.231	1.687	151	0.444	0.039	0.517
RF	0.988	0.197	–0.270	1.182	42	0.429	0.000	0.571
S	0.126	0.032	–0.488	0.130	344	0.485	0.088	0.427
TI	–0.807	0.497	–1.643	–1.208	236	0.483	0.059	0.458
UI	–1.112	0.207	–0.981	0.013	3	1.000	0.000	0.000

Table 1 – The main results of our applications.

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

- that, although 5 of the 13 investigated stock assets show a negative CR_T, the summation of all these CR_Ts is positive and equal to 5.181. That can be interpreted as the fact that, under a suitable diversification in the investments, our FTA is enough well-working in the long-run term (see, for example, Figure 1);

- that, of course, the magnitude of this summation is not satisfying. It certainly depends on the choice to utilize in the expression for CR_T a free-risk rate of return equal to zero, and also depends 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, the utilization of an amount of capital to invest equal to $\mu + \mu_0$ appears able to reduce all the losses in T (and, as counterpart, also all the gains in T), and, among the three occurred critical cases (namely: Banca Intesa, Telecom Italia, and Unicredito Italiano), it is also able to guarantee the continuation of the financial trading of the one Telecom Italia;

- that in the critical case Banca Intesa, the behaviour of our FTA before the occurring of large losses has been acceptable (see Figure 2 and Figure 3).

CONCLUDING REMARKS

In this section we provide some remarks for possible extensions of our contribution. In particular:

- in order to (attempt 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 $F_t = 1$.

   Of course, in such a case we should need to determine the "optimal" values of two other parameters, *i.e.* ε^- and ε^+;

- in order to (attempt to) improve the learning capabilities of the ANN, we conjecture that could be profitable to use a multi-layer perceptron model (to the best of our knowledge, such a check is carried out in Gold, 2003, 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;

12

- in order to (attempt to) explicitly take into account the risk in our FTA, we conjecture that could be profitable to utilize some suitable risk-adjusted version of the investor's gain index (1);
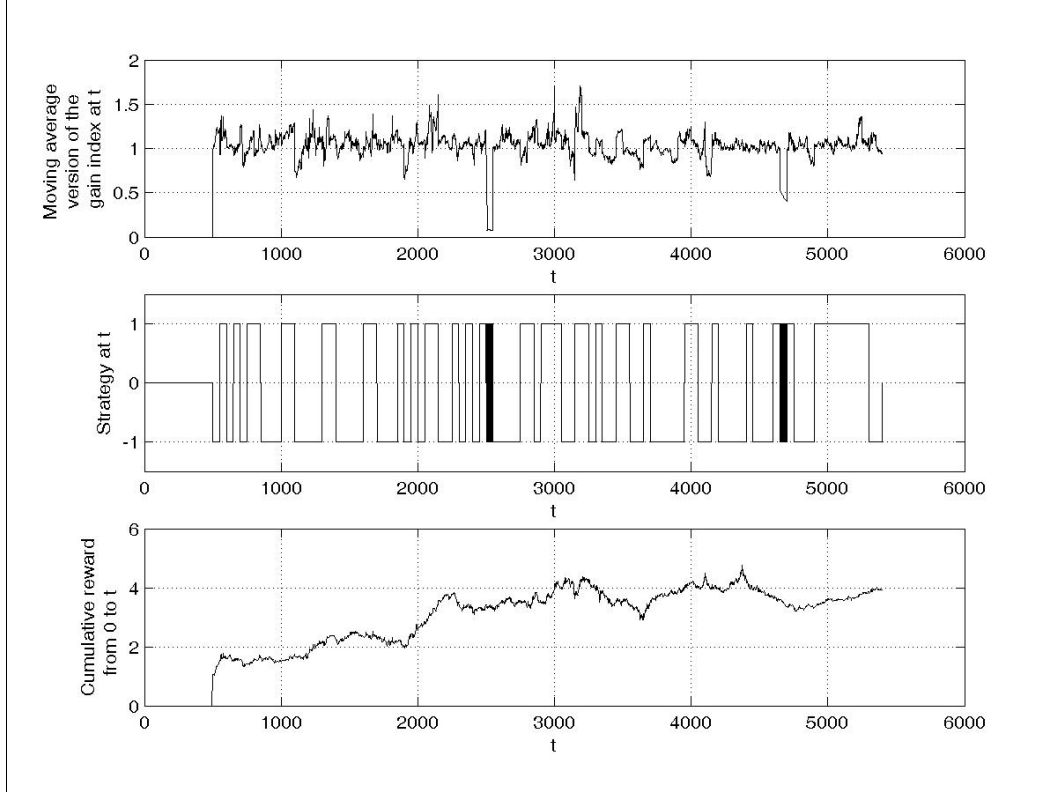

Figure 1 – The case Finmeccanica (it is to notice that the cumulative reward represented in the figure is the gross one, *i.e.* $CR_T + \mu + \mu_0$).

- finally, in order to (attempt to) make "close" our FTA, we conjecture that could be fruitful to (attempt to) make the drawdown-like phenomenon management endogenous to the learning process.

References

Abecasis, S. M., Lapenta, E. S., & Pedreira, C. E. (1999). Performance metrics for financial time series forecasting. *Journal of Computational Intelligence in Finance*, *July/August*, 5-23.

Appel, G. (1979). *The Moving Average Convergence-Divergence Method*. Great Neck, New York (U.S.A.): Signalert.

Bertsekas, D.P. (1995). *Dynamic Programming and Optimal Control*. Belmont, Massachusetts (U.S.A.): Athena Scientific.

Bertsekas, D.P., & Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Belmont, Massachusetts (U.S.A.): Athena Scientific.

Bishop, C. (1995). *Neural Networks for Pattern Recognition.* Oxford, United Kingdom: Oxford University Press.
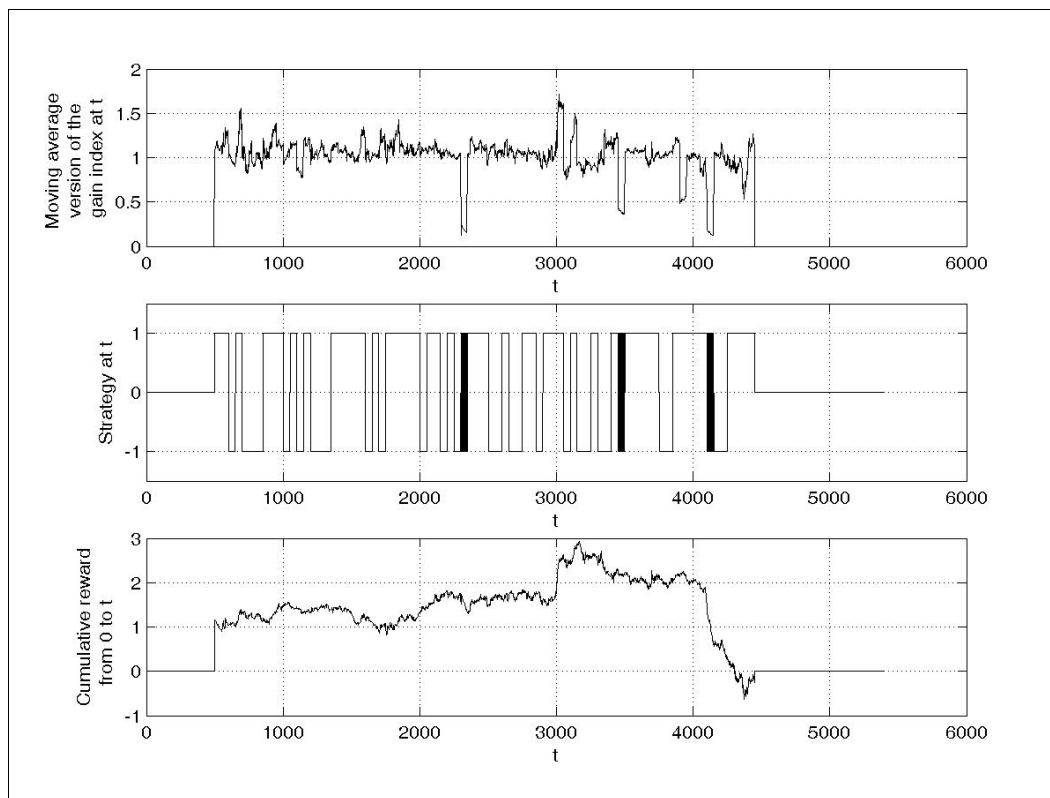

Figure 2 – The critical case Banca Intesa (it is to notice that the cumulative reward represented in the figure is the gross one, *i.e.* $CR_T + \mu + \mu_0$).

Corazza, M., Vanni, P., & Loschi, U. (2002). Hybrid automatic trading system: technical analysis & group method of data handling. In M. Marinaro, & R. Tagliaferro (Eds.), *Neural Nets. 13th Italian Workshop on Neural Nets, WIRN Vietri 2002. Vietri sul Mare, Italy, May/June 2002. Revised Paper* (pp. 47-55). Berlin, Germany: Springer.

Lee, C. M. C., & Swaminathan, B. (2000). Price momentum and trading volume. *Journal of Finance*, *LV*, 2017-2069.

Lo, W. A., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, *LV*, 1705-1769.

Murphy, J.J. (1999). *Study Guide to Technical Analysis of the Financial Markets*. New York, New York (U.S.A.): Prentice Hall Press.

Gold, C. (2003). FX trading via recurrent reinforcement learning. *Proceedings of IEEE International Conference on Computational Intelligence in Financial Engineering, 2003* (pp. 363-370).

Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, *12*, 875–889.

Moody, J., & Wu, L. (1997) Optimization of trading systems and portfolios. In Y. Abu-Mostafa, A. N. Refenes, & A. S. Weigend (Eds.), *Decision Technologies for Financial Engineering* (pp. 23-35). London, United Kingdom: World Scientific.

Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting*, *17*, 441–470.
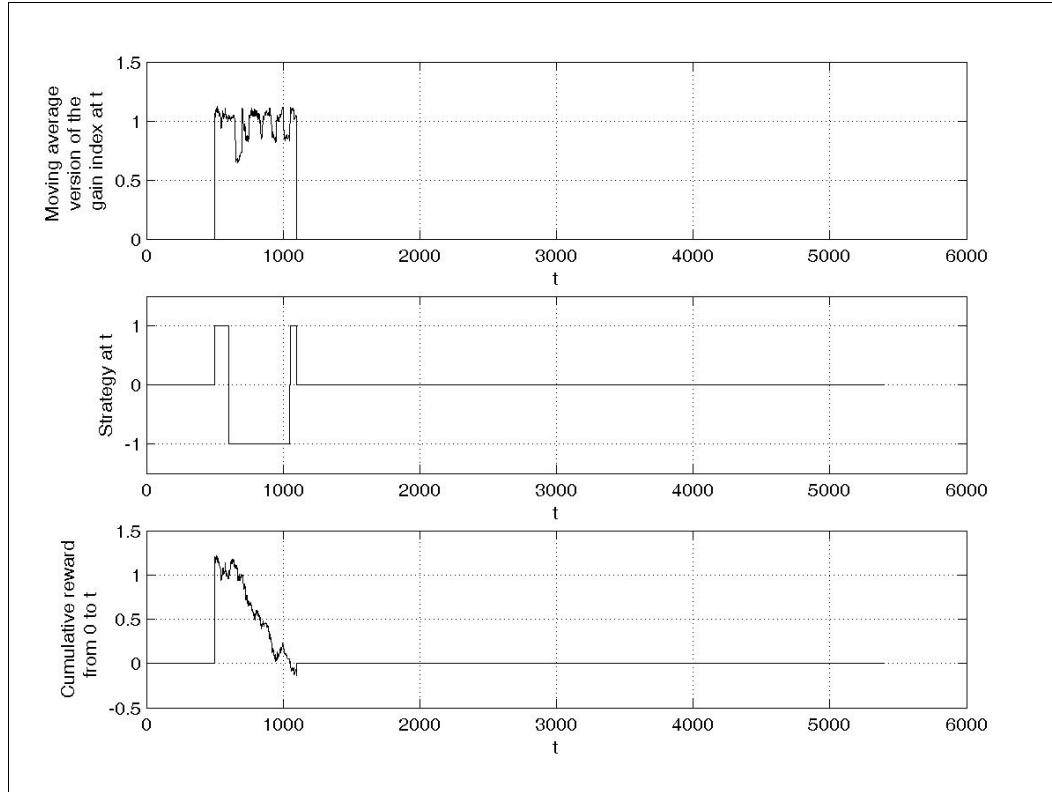
Figure 3 – The critical case Unicredito Italiano (it is to notice that the cumulative reward represented in the figure is the gross one, *i.e.* $CR_T + \mu + \mu_0$).

Nison, S. (1991). *Japanese Candlesticks Charting Technique*. New York, New York (U.S.A.): Prentice Hall Press.

Plihon, V., Wu, F., & Gardarin, G. (2001). A financial data mining trading system. In M. Bouzeghoub, Z. Kedad, & E. Métais (Eds.), *Natural Language Processing and Information Systems: 5th International Conference on Applications of Natural Language to Information Systems, NLDB 2000. Versailles, France, June 28-30, 2000. Revised Paper* (p. 370). Berlin, Germany, Springer.

Sutton, R. S., & Barto, A. G. (1997). *An Introduction to Reinforcement Learning*. Cambridge, Massachusetts (U.S.A.): MIT Press.

Wilder, J. W. (1978). *New Concepts in Technical Trading*. Greensboro, North Carolina (U.S.A.): Trend Research.