



Università
Ca' Foscari
Venezia

Corso di Laurea Magistrale

in Economia e Finanza

(ordinamento ex D.M. 270/2004)

Tesi di Laurea

**Reinforcement Learning ed Analisi Tecnica
applicati ad un sistema di trading automatico**

Relatore:

Ch. Prof. Marco Corazza

Candidata:

Marta Grossato

856613

Anno accademico:

2019/2020

Sommario

Introduzione	1
1. La Teoria del Mercato Efficiente	3
1.1 La definizione di Mercato Efficiente	5
1.1.1 I gradi di efficienza del mercato	8
1.2 Il mercato è davvero efficiente?	9
1.3 La Teoria del Mercato Adattivo	12
1.3.1 Un nuovo ordine mondiale: le implicazioni della AMH	15
2. Il Reinforcement Learning	17
2.1 Gli elementi del Reinforcement Learning.....	19
2.2 I processi decisionali alla Markov.....	21
2.3 Obiettivo e sistema di ricompensa.....	23
2.4 La funzione di valore	24
2.4.1 L'ottimizzazione della funzione di valore	25
2.4.2 Valutazione e perfezionamento del criterio π	27
2.5 I metodi TD: apprendimento per differenze temporali.....	30
2.6 Algoritmo Q-Learning.....	32
2.6.1 Approssimazione e ottimizzazione della funzione di valore.....	34
2.6.2 Il metodo della discesa del gradiente	35
2.6.3 Approssimazione lineare della funzione di valore.....	37
3. Struttura dell'algoritmo	39
3.1 I descrittori dello stato.....	39
3.2 La struttura di selezione dell'azione.....	42
3.3 La funzione di ricompensa.....	45
3.4 I costi di transazione	46
3.5 La funzione di squashing.....	47
3.6 Presentazione delle serie storiche.....	49
3.7 Fase di calcolo.....	55

4. Applicazione dell’algoritmo e risultati	58
4.1 Presentazione delle statistiche	59
4.2 Davide Campari-Milano	62
4.3 Enel S.p.A.....	69
4.4 Fiat Chrysler Automobiles	77
4.5 Intesa Sanpaolo S.p.A.	82
4.6 Telecom Italia S.p.A.....	88
4.7 Considerazioni finali	95
Conclusioni	100
Bibliografia	102

Introduzione

Chiunque si approcci allo studio dei mercati finanziari inevitabilmente finirà per avere a che fare con la Teoria del Mercato Efficiente, che è sicuramente una delle teorie più accreditate e dibattute sulla scena accademica. Formulata negli anni '60 da P. Samuelson e E. Fama – seppur seguendo studi indipendenti –, tale teoria afferma che in un mercato efficiente non esistano opportunità di profitto sfruttabili dai soggetti che vi partecipano. I prezzi, infatti, incorporerebbero già tutta l'informazione disponibile rendendo di fatto impossibile la presenza di asimmetrie o di qualsivoglia vantaggio economico sfruttabile per ottenere dei profitti. Nonostante negli anni ci siano state evidenze del fatto che questa teoria alle volte non rappresenti correttamente la realtà, essa si presenta ancora come un caposaldo di difficile confutazione nel panorama accademico. Una domanda sorge però spontanea: se questa teoria è vera, quale sarebbe l'interesse a svolgere attività di *active management* o di trading dato che non è possibile ottenere un profitto maggiore di quello che si ottiene con una gestione passiva? Una nuova teoria, però, comincia a farsi strada nel *framework* accademico ed è la Teoria del Mercato Adattivo sviluppata e presentata da A. Lo, la quale mette in discussione il paradigma della razionalità postulato dalla Teoria dei Mercati Efficienti. Gli individui, infatti, in un mercato efficiente agirebbero sempre e soltanto razionalmente, eseguendo un'azione solo dopo aver effettivamente appurato che proprio quell'azione permette di massimizzare l'utilità derivante, ad esempio, da un investimento. La Teoria del Mercato Adattivo, invece, propone una visione diversa: gli individui agiscono secondo processi decisionali alle volte distorti, un individuo difficilmente riesce a determinare quale sarà la scelta più razionale da fare. Uno scenario come quello appena descritto, in cui gli individui sono intelligenti ma soprattutto fallibili, si creano possibilità di profitto ed ecco che le attività di trading ed *active management* trovano il loro spazio.

In questo elaborato, assumendo come *framework* di lavoro la Teoria del Mercato Adattivo, si sviluppano e si applicano sistemi di trading finanziario che. Inizialmente, vengono presentate, seppur in maniera sintetica, le due teorie di cui sopra, evidenziandone

gli aspetti principali. Successivamente, viene presentato il metodo di apprendimento per rinforzo, definito *Reinforcement Learning*, entrando nei dettagli del metodo di apprendimento per differenze temporali chiamato *Q-Learning*. Nell'ambito di queste metodologie verrà implementato un sistema di trading automatico, il quale trova la propria giustificazione proprio nel *framework* teorico descritto dalla Teoria del Mercato Adattivo.

L'obiettivo di questo elaborato è testare due differenti sistemi di trading automatico, entrambi basati su un algoritmo appartenente alla metodologia di apprendimento per differenze temporali, ossia il *Q-Learning algorithm*. Entrambi i sistemi verranno applicati a cinque serie storiche di prezzi di titoli reali quotati sul mercato italiano. Il primo sistema utilizza come input i rendimenti logaritmici calcolati partendo proprio dalle serie storiche dei prezzi; il secondo sistema, invece, utilizza come input degli indicatori di Analisi Tecnica, ossia indicatori che normalmente si usano per prevedere e individuare tendenze e *pattern* nell'andamento dei prezzi. Interessante è, quindi, testare la risposta dell'algoritmo a input diversi, per vedere se effettivamente l'uso di input differenti e leggermente più raffinati possa condurre ad un miglioramento delle performance ottenibili.

La struttura della tesi si articola in cinque capitoli. Il primo capitolo si presenta come una panoramica sul *framework* accademico per quanto riguarda la Teoria del Mercato Efficiente e la più recente Teoria del Mercato Adattivo. Il secondo capitolo presenta il metodo di apprendimento per rinforzo, ossia il *Reinforcement Learning*, spiegandone il funzionamento e le caratteristiche principali. Nel medesimo capitolo vengono presentati anche i metodi di apprendimento per differenze temporali, entrando poi nel merito della struttura dell'algoritmo *Q-Learning* utilizzato nella parte applicativa del presente lavoro nel terzo capitolo. Il quarto capitolo riassume i risultati ottenuti dall'applicazione dell'algoritmo, nelle due varianti implementate, alle serie storiche dei prezzi di cinque titoli quotati sul mercato italiano. L'esposizione dei risultati è impostata come un confronto tra i risultati ottenuti dalla prima applicazione dell'algoritmo e successivamente dalla seconda, in modo da fornire gli elementi necessari a valutare l'effettiva utilità degli strumenti dell'Analisi Tecnica insieme alle metodologie del *Reinforcement Learning*.

Il quinto ed ultimo capitolo è costituito dalle conclusioni e le considerazioni finali riguardo i risultati ottenuti, includendo eventuali suggerimenti per possibili lavori futuri.

La Teoria del Mercato Efficiente

Ad oggi sentir parlare di attività di trading online è diventata una cosa comune, si tratta di un termine entrato a far parte del linguaggio corrente ormai da tempo. Nonostante il trading online sia un'attività estremamente attuale e in continua evoluzione, la sua nascita è da ricercare in tempi non molto recenti. Innanzitutto, il trading – online e non – non è solamente speculazione finanziaria, si tratta di un processo ben più complesso che affonda le proprie radici in elementi di diversa natura. Si tratta di un fenomeno in continua evoluzione che nasce dalla fusione di due principali aree di competenza, rispettivamente la finanza e la tecnologia. Fare trading significa sostanzialmente inviare informazioni da un computer ad un altro, le quali contengono dati relativi a scambi di capitali sui mercati finanziari globali, regolamentati e non. È evidente, quindi, che finanza e informatica sono due elementi imprescindibili dell'attività di trading.

Durante il periodo della Guerra Fredda numerose sono state le tecnologie sviluppate per scopi militari, come ad esempio la grande rete di difesa aerea chiamata S.A.G.E. (*Semi Automatic Ground Environment*) finalizzata alla protezione del suolo americano dai bombardieri sovietici. Grazie al fenomeno del *reverse engineering*, cioè un fenomeno tale per cui le tecnologie e le scoperte finalizzate all'uso in campo militare vengono poi diffuse anche per un uso civile, questi nuovi mezzi tecnologici cominciano a diffondersi anche tra le grandi aziende dell'epoca. Verso la fine degli anni '70 un gran numero di banche già faceva uso dei primi elaboratori di calcolo disponibili per snellire le procedure e i processi interni. Si tratta di un primo, grande passo verso la creazione di quello che consideriamo oggi trading. Al fine di rendere più efficienti gli scambi e le operazioni di negoziazione sui mercati finanziari, le principali borse mondiali cominciarono ad utilizzare a loro volta strumenti tecnologici sempre più all'avanguardia per cercare di creare una rete su cui effettuare transazioni nel modo più semplice e veloce possibile. Fino agli anni '80, rimane comunque il telefono il principale strumento di comunicazione in ambito di compravendita di titoli. Con la graduale diffusione di Internet e di computer sempre più sofisticati in grado di elaborare una quantità

di dati via via crescente si arriva allo scenario attuale in cui praticamente chiunque ha accesso ad Internet e ai servizi finanziari. Oggigiorno, chiunque può decidere di cimentarsi nell'attività di trading, è sufficiente collegarsi ad internet ed iscriversi ad una delle numerose piattaforme che offrono tale servizio. Se agli inizi dello sviluppo di questo fenomeno l'interesse era rendere più veloci gli scambi di informazioni sui mercati finanziari, ai tempi attuali l'attenzione viene per lo più concentrata sull'implementazione di sistemi di trading. Con sistemi di trading si intendono una serie di regole che indicano alla figura del *trader* quali sono le operazioni da eseguire al fine di ottenere un profitto, al giorno d'oggi esistono moltissime piattaforme accessibili da chiunque in grado di fornire sistemi di trading più o meno validi. Una particolare categoria di sistemi di trading è rappresentata dai sistemi di trading automatici, i quali sono il principale interesse del presente lavoro. In un sistema di trading automatico la figura del trader viene sostituita da un agente artificiale in grado di eseguire le operazioni e le transazioni seguendo lo schema e le regole proprie del sistema di trading. Questa particolare categoria dei sistemi di trading risulta essere l'ultima frontiera in ambito di investimenti e speculazione: l'implementazione di sistemi come questi appena descritti richiede l'impiego di tecnologie all'avanguardia e lo studio di discipline e ambiti quali, ad esempio, l'Intelligenza Artificiale. Se si esegue una breve ricerca su Internet, si possono trovare una moltitudine di siti e piattaforme che offrono ai potenziali clienti proprio questo servizio di trading automatico, fatto che dimostra quanto questo settore attiri sempre più l'attenzione di investitori esperti ma anche di chi si avvicina per la prima volta al mondo dei mercati finanziari.

Al termine di questa carrellata sulla nascita e la diffusione del trading, appare evidente come questo fenomeno sia oramai consolidato e diffuso a livello globale; uno sviluppo come quello conosciuto dall'attività di trading e, più recentemente, dai sistemi di trading risulta però essere in contrasto con la più diffusa e accreditata teoria riguardo il funzionamento dei mercati finanziari, vale a dire la Teoria dei Mercati Efficienti.

Come andremo ad evidenziare nel corso del presente capitolo, la Teoria dei Mercati Efficienti, d'ora in poi indicata come EMH¹, viene presentata come la principale teoria in grado di spiegare accuratamente il funzionamento e le dinamiche dei mercati finanziari e ad oggi rimane materia di interesse per studi e ricerche. La EMH, nella sua forma più estrema, afferma che i prezzi di un prodotto sul mercato finanziario incorporano già ogni informazione disponibile al pubblico in quel momento. In altri termini, se i prezzi riflettono già ogni informazione disponibile significa che nessun tipo di speculazione è possibile e

¹ Dall'inglese Efficient Market Hypothesis.

quindi la raccolta di informazioni per cercare di predire i movimenti del mercato finanziario è solamente tempo sprecato. Ma allora come ha potuto il trading svilupparsi e diffondersi così massicciamente se secondo la EMH il trading stesso non può portare ad ottenere profitti?

Negli anni sono sorte numerose critiche alla Teoria dei Mercati Efficienti ed è proprio partendo da queste critiche che nasce una teoria che riprende la EMH e la estende utilizzando concetti propri della finanza comportamentale: si tratta della Teoria dei Mercati Adattivi, d'ora in poi indicata con AMH².

Nel seguito di questo capitolo verranno esposte le principali caratteristiche della Teoria dei Mercati Efficienti; poi, partendo dalle critiche a questa, verrà esposta la Teoria dei Mercati Adattivi, la quale riprende molte delle idee di base della EMH reinterpretandole attraverso una serie di principi evolutivi che giustificano l'attività di trading.

1.1 La definizione di Mercato Efficiente

Cosa significa essere efficienti? In ambito economico il termine “efficiente” viene utilizzato molto spesso. Se andiamo a consultare la voce “efficienza economica” dell'enciclopedia Treccani³ troviamo la seguente definizione:

«Rapporto tra benefici e costi o, più in generale, tra il grado nel quale vengono raggiunti gli obiettivi e i mezzi utilizzati. È realizzata quando, con gli strumenti a disposizione, non si può migliorare il livello di conseguimento degli scopi desiderati.»

In termini economici quindi, l'efficienza si realizza quando l'allocazione delle risorse attuale è la migliore possibile ed è tale per cui non ne esiste una alternativa che migliori la condizione di alcuni individui senza al contempo peggiorare la condizione di altri.

Il concetto di efficienza legata ai mercati finanziari è però un concetto leggermente diverso da quello della classica teoria economica. L'efficienza cui si fa riferimento è di fatto l'efficienza informativa, così come venne definita da Stiglitz nel 1981 [1]. L'efficienza informativa si raggiunge se:

² Dall'inglese Adaptive Markets Hypothesis.

³ http://www.treccani.it/enciclopedia/efficienza-economica_%28Dizionario-di-Economia-e-Finanza%29/

1. sul mercato è possibile ottenere tutta l'informazione disponibile, non esistono quindi asimmetrie informative di nessun genere;
2. i prezzi di mercato riflettono immediatamente e completamente tutte le informazioni che il mercato stesso fornisce;
3. le imprese riescono a veicolare in maniera efficiente le informazioni sulle prospettive di crescita del proprio *business* ai nuovi potenziali investitori.

È proprio questo il tipo di efficienza cui la Teoria dei Mercati Efficienti fa riferimento, ossia un tipo di efficienza legato non tanto alle risorse ma alle informazioni a disposizione in un dato momento sul mercato.

Una prima definizione di cosa significa “mercato efficiente” ci viene fornita da colui che è considerato il massimo fautore della teoria EMH, ossia il premio Nobel Eugene F. Fama. Nel suo articolo del 1965 [2] Fama scrive:

«An “efficient” market is defined as a market where there are a large number of rational, profit-maximizers actively competing, with each trying to predict future market values of individual securities, and where important current information is almost freely available to all participants.»

Nonostante questa venga identificata come la prima definizione di mercato efficiente, esistono riferimenti all'efficienza di mercato – anche se non in maniera diretta – che risalgono a circa un secolo prima dell'articolo di Fama. È infatti da attribuire al *broker* e matematico amatoriale francese Jules Regnault una prima definizione di quello che fu poi definito mercato efficiente. Egli con la sola logica e l'osservazione del mercato borsistico francese era arrivato ad affermare che i prezzi dei titoli in un dato momento riflettevano la cosiddetta *wisdom of the crowd*, ossia la saggezza della folla. Cosa significa che i prezzi riflettono la saggezza della folla? Significa esattamente ciò che Fama quasi un secolo dopo ha definito essere il mercato efficiente, cioè un mercato in cui tutta l'informazione disponibile si riflette immediatamente nei prezzi dei titoli. Per chiarire meglio il concetto pensiamo ad un banale esempio: supponiamo che un modello riesca a prevedere con un elevato grado di confidenza che il prezzo del titolo Alpha, che oggi è pari a 100€, subirà un aumento notevole nei prossimi giorni arrivando a 120€. Questa informazione scatenerà negli investitori una reazione abbastanza ovvia: questi cercheranno di acquistare quanti più titoli della società Alpha oggi per poterli rivendere dopo che il prezzo sarà salito; chi invece già possiede i titoli Alpha non sarà minimamente intenzionato a venderli alla quotazione attuale dato che si prevede che

questa possa alzarsi notevolmente nei prossimi giorni. Cosa accade quindi sul mercato? L'aumento della domanda del titolo Alpha sul mercato causerà un immediato rialzo del prezzo del titolo stesso. L'informazione del possibile rialzo del prezzo viene quindi immediatamente assorbita dal mercato causando un effettivo rialzo del prezzo stesso. Ecco cosa significa "saggezza della folla": se la folla si aspetta che una determinata cosa accada, allora questa accade perché il suo comportamento fa in modo che il mercato reagisca di conseguenza, riflettendo le informazioni che la folla stessa ha a disposizione.

Nonostante i diversi contributi alla teoria nel corso dei decenni, i "padri" della EMH sono considerati effettivamente Eugene F. Fama e Paul A. Samuelson, i quali hanno lavorato in maniera indipendente alla teoria e sono arrivati alle medesime conclusioni nello stesso periodo, pubblicando nel 1965 due articoli che formalizzavano l'idea del mercato efficiente. Anche se Fama e Samuelson sono arrivati alle stesse conclusioni con la loro ricerca, sono partiti da idee molto diverse.

Con un approccio molto tecnico, Samuelson ha esplicitato il *Theorem of Fair Game Futures Pricing*, il quale stabilisce che non è possibile ottenere extra profitti semplicemente analizzando la sequenza dei prezzi passati. Il titolo del suo articolo ben riassume il concetto: "*Proof that properly anticipated prices fluctuate randomly*" [3]. In sostanza Samuelson ha dimostrato che in un mercato il prezzo di domani riflette solo ed esclusivamente le nuove informazioni di domani ed è del tutto indipendente dal prezzo e dalle informazioni di oggi.

Partendo da concetti totalmente diversi, Fama arriva a dimostrare che i prezzi in un mercato efficiente seguono generalmente un modello di tipo *Random walk*. Fama [2] afferma che se la teoria del *Random walk* è un'accurata descrizione della realtà, allora tutte le tecniche e le procedure implementate negli anni per cercare di prevedere i movimenti del mercato sono sostanzialmente inutili e prive di alcun valore. Le tecniche e le procedure cui si fa riferimento sono l'analisi fondamentale e l'analisi tecnica: l'analisi fondamentale di basa sul concetto di "valore intrinseco" di un titolo, il quale va poi confrontato con il valore che questo ha sul mercato per capire se sia sovrastimato o sottostimato; l'analisi tecnica invece si basa sulla convinzione che "la storia si ripete" e il compito dell'analista dovrebbe essere quello di ricercare i cosiddetti *pattern* nelle sequenze dei prezzi in modo da essere in grado di individuarli in anticipo quando questi si ripresentano.

Entrambi gli studiosi sono comunque arrivati alla seguente conclusione: più un mercato è efficiente, tanto più i cambiamenti nella sequenza dei prezzi sono casuali ed imprevedibili.

1.1.1 I gradi di efficienza del mercato

Nella letteratura si distinguono tre diversi gradi di efficienza del mercato, ciascuno basato su un diverso set di informazioni disponibili di diversa capienza. Il primo a parlare di gradi di efficienza fu proprio Fama con un articolo pubblicato nel 1969 [4], presentato come un riassunto completo di tutta l'evidenza teorica ed empirica che fino a quel momento era stata raccolta sulla EMH.

Vediamo nel dettaglio i diversi gradi di efficienza del mercato:

1. efficienza in forma debole: i prezzi riflettono esclusivamente le informazioni desumibili dallo storico dei prezzi stessi e dai volumi. In questo scenario è possibile che alcuni soggetti siano in grado di performare meglio del mercato, in quanto si possono creare delle situazioni di asimmetria informativa attraverso l'uso di informazioni non assimilate dal mercato;
2. efficienza in forma semi-forte: alle informazioni desumibili dallo storico dei prezzi e dei volumi si aggiungono le informazioni pubblicamente disponibili e/o rese note dalle aziende. Si tratta quindi di informazioni derivanti da comunicazioni, fusioni, acquisizioni, report annuali ecc. che un'azienda è tenuta a divulgare. In questo contesto un soggetto potrebbe riuscire a ottenere prestazioni migliori di quelle del mercato solamente sfruttando informazioni private di cui nessun altro è a conoscenza, il che potrebbe anche sfociare in un reato, il cosiddetto *insider trading*;
3. efficienza in forma forte: le informazioni contenute nei prezzi sono tutte quelle disponibili sul mercato, siano esse di pubblico dominio o private. In questo scenario non esiste alcun soggetto in grado di battere il mercato dato che tutte le informazioni disponibili sono già state assorbite dai prezzi.

Nella Teoria dei Mercati Efficienti si fa riferimento a “tutte le informazioni disponibili”, pertanto la rappresentazione empirica del modello viene data considerando lo scenario estremo in cui il grado di efficienza è al massimo.

Dato un set di informazioni Φ_t , è possibile quindi tradurre l'espressione “tutte le informazioni disponibili sono contenute nei prezzi” nel modo seguente:

$$\mathbb{E}(\tilde{p}_{t+1}|\Phi_t) = [1 + \mathbb{E}(\tilde{r}_{t+1}|\Phi_t)]p_t \quad 1$$

dove p_t è il prezzo di un titolo al tempo t , \tilde{p}_{t+1} è la variabile casuale che indica il prezzo del titolo al tempo $t+1$ e \tilde{r}_{t+1} è la variabile casuale che indica il rendimento uniperiodale percentuale atteso⁴.

Se il set informativo Φ_t viene pienamente utilizzato nella formazione del prezzo e nella determinazione del rendimento di equilibrio, allora non esiste alcun sistema di trading basato sul medesimo set di informazioni Φ_t in grado di ottenere performance migliori del rendimento atteso di equilibrio. In altri termini, questo risultato implica che, dato z_{t+1} il rendimento in eccesso in $t + 1$ definito come:

$$z_{t+1} = r_{t+1} - \mathbb{E}(\tilde{r}_{t+1} | \Phi_t), \quad 2$$

si ha il valore atteso condizionato al set informativo Φ_t :

$$\mathbb{E}(z_{t+1} | \Phi_t) = 0. \quad 3$$

La sequenza $\{z_t\}$ risulta quindi essere un *fair game* o gioco equo rispetto alla sequenza informativa $\{\Phi_t\}$. Non esiste quindi alcuna possibilità di usare le informazioni Φ_t per ottenere un rendimento al di sopra del livello del mercato.

1.2 Il mercato è davvero efficiente?

Da quando la EMH è stata formalizzata, sono stati innumerevoli gli studi e le ricerche condotte al riguardo. Per anni si è cercata una prova, un'evidenza che permettesse di invalidare questa teoria e dimostrare che i mercati non seguono un meccanismo come quello descritto da Fama e Samuelson. Questo fatto non dovrebbe sorprendere: pensiamo ad esempio a tutti quei manager e gestori di fondi comuni a gestione attiva; se la Teoria dei Mercati Efficienti avesse ragione, allora il lavoro di queste persone diventerebbe inutile e privo di alcun valore. La EMH non ha quindi suscitato molto entusiasmo nella comunità dei managers e dei gestori di portafoglio professionali. Il dibattito resta aperto: i mercati sono o

⁴ Il rendimento uniperiodale percentuale r_t è calcolato come $(p_t - p_{t-1})/p_{t-1}$.

non sono efficienti? Vediamo a tal proposito qualche critica mossa negli anni da numerosi studiosi e ricercatori.

Lo e MacKinlay nel 1988 [5] implementarono un test per misurare la varianza dei rendimenti di un titolo relativa a diversi *holding periods*. Secondo l'ipotesi *random walk*, il range di fluttuazione della varianza dovrebbe aumentare proporzionalmente al range temporale analizzato e il test implementato dai due ricercatori aveva proprio l'obiettivo di verificare se questo fosse vero. La domanda che si posero è la seguente: la varianza calcolata su rendimenti bi-settimanali è esattamente due volte la varianza calcolata su rendimenti settimanali, come dovrebbe essere in un mercato efficiente? Analizzando i risultati ottenuti si accorsero che la varianza calcolata su rendimenti bi-settimanali era tre volte la varianza dei rendimenti settimanali, e non due volte come avrebbe dovuto essere. Questo risultato dimostra la presenza di correlazione seriale dei rendimenti, il che rigetta l'ipotesi che i rendimenti seguano un modello *random walk*.

Nel 1980 Grossman e Stiglitz [6] arrivarono addirittura ad affermare che il mercato efficiente è impossibile, poiché se i mercati fossero perfettamente efficienti non ci sarebbe alcun profitto per la raccolta di informazioni e l'attività di trading diventerebbe inutile, portando i mercati stessi al collasso. I due ricercatori sostenevano che fosse in realtà il grado di inefficienza del mercato a determinarne l'equilibrio: più è alto il grado di inefficienza, più possibilità di profitto si presentano ai partecipanti; l'equilibrio di mercato, quindi, sussiste in presenza di sufficienti opportunità di profitto sfruttabili. I sostenitori della EMH risposero a questa critica sostenendo che esistono delle inefficienze sul mercato e non si possono negare, tuttavia queste inefficienze vengono prontamente eliminate grazie all'impatto di forze opposte che hanno il compito di sfruttare queste anomalie per eliminarle istantaneamente.

Un ulteriore esempio in contrasto la teoria EMH sono i vari *pattern* scoperti negli anni. Un *pattern* è sostanzialmente un certo comportamento dei prezzi che ricorre nel tempo. L'analisi tecnica si occupa proprio di individuare i *pattern* nelle serie storiche dei prezzi dei titoli in modo da utilizzarli per prevedere il comportamento di questi nel momento in cui un *pattern* insorge. Uno dei *pattern* più famoso è il cosiddetto *January effect*: si tratta di un *pattern* che individua nel mese di Gennaio una performance migliore dei titoli a ridotti volumi di capitalizzazione rispetto a quelli con elevati volumi di capitalizzazione. La scoperta di questo *pattern* è avvenuta nel 1942 da parte di Sidney Wachtel, almeno una ventina di anni prima della formalizzazione della EMH di Samuelson e Fama. Il *January effect* non è l'unico pattern scoperto negli anni; il problema con queste "anomalie" è che minano uno dei capisaldi della EMH, ossia quello che presuppone la non prevedibilità dei prezzi sul mercato. Quindi se

questi *pattern* effettivamente esistono, come possono essere conciliati con la Teoria dei Mercati Efficienti?

Un'ultima ma non meno importante critica mossa alla EMH riguarda il comportamento dei partecipanti nel mercato. La Teoria dei Mercati Efficienti presuppone individui perfettamente razionali in grado di prendere sempre la decisione ottimale che massimizza la propria utilità. La realtà però è ben diversa: esistono dei *bias* comportamentali che inducono gli individui ad agire in un modo non ottimale e che va contro la razionalità economica. Per capire meglio questo aspetto vediamo un famoso esperimento condotto da due psicologi, il premio Nobel Kahneman e Tversky, nel 1979. Supponiamo che ci vengano offerte due opportunità di investimento tra cui scegliere, A e B:

- A. Garantisce un profitto sicuro di 240.000 €;
- B. Consiste in un biglietto della lotteria che permette di ottenere 1.000.000 € con una probabilità del 25% e 0 € con una probabilità del 75%.

Se calcoliamo il valore atteso di entrambe le alternative vediamo che A ha un valore atteso di 240.000 € mentre B ha un valore atteso di 250.000 €. Secondo la classica teoria economica un soggetto razionale sceglierebbe l'investimento B in quanto presenta il valore atteso maggiore; il risultato che hanno ottenuto i due psicologi sottoponendo questo quesito ad un gruppo di soggetti non ha però ottenuto una risposta pienamente in linea con la teoria: la maggior parte dei soggetti ha infatti scelto l'opzione A, cioè l'opzione che garantiva un profitto sicuro.

L'esperimento però continua. Dopo aver effettuato la prima scelta tra A e B, ci vengono sottoposte altre due alternative, C e D:

- C. Si tratta di una perdita sicura di 750.000 €;
- D. Consiste in un biglietto della lotteria che frutta 0 € con una probabilità del 25% e una perdita di 1.000.000 € con una probabilità del 75%.

In questo caso molti dei soggetti analizzati hanno scelto l'opzione D nonostante sia l'opzione più rischiosa. I soggetti che prima si erano dimostrati avversi al rischio scegliendo l'opzione A, di fronte ad una perdita quasi certa diventano amanti del rischio scegliendo l'opzione D. Questo esperimento evidenzia la tendenza di alcuni individui ad essere avversi al rischio quando sono di fronte ad una opportunità di profitto e ad essere invece amanti del rischio di fronte ad una scelta tra opportunità che implicano una perdita.

Questo esempio evidenzia il fatto che il comportamento dei soggetti gioca un ruolo fondamentale nelle scelte che questi ultimi prendono. È per questo che sempre più in ambito accademico si cerca di analizzare la realtà economica sotto la lente della scienza

comportamentale. Di *bias* cognitivi, i quali sono una sorta di distorsione comportamentale, ne esistono molti e tutti contribuiscono a definire uno scenario che si avvicina molto di più alla realtà rispetto all'ambiente descritto dalla EMH.

Dopo questa breve carrellata di critiche alla EMH viene spontaneo chiedersi se il mercato sia davvero efficiente nel modo descritto da Fama e Samuelson. Una possibile risposta ci viene data da Andrew Lo, che nel 1999 pubblica un articolo in cui espone una nuova teoria in grado di conciliare l'efficienza del mercato con la scienza comportamentale. Si tratta di una teoria chiamata Teoria del Mercato Adattivo che applica principi evolutivi alle interazioni finanziarie tra i soggetti attivi sul mercato.

1.3 *La Teoria del Mercato Adattivo*

«In the Adaptive Markets Hypothesis (AMH) intelligent but fallible investors learn from and adapt to changing economic environments.»

Questa citazione [7] è il riassunto di tutto quello che rappresenta la Teoria dei Mercati Adattivi, la quale verrà esaminata nei dettagli nel corso del presente paragrafo.

Il paradigma della razionalità è forse l'elemento della EMH che viene criticato più duramente. Come si può leggere dalla citazione in apertura del paragrafo, la AMH sostiene che gli individui non siano perfettamente razionali ma intelligenti e soprattutto fallibili. In tale *framework* gli investitori non sono né perfettamente razionali né completamente irrazionali, sono in realtà investitori intelligenti, previdenti e competitivi che si adattano ad una realtà economica mutevole. È proprio introducendo il concetto di adattamento degli individui che si arriva ad una conciliazione delle ipotesi della EMH con i principi della scienza comportamentale, dando vita alla AMH.

Il lavoro di Lo trae spunto dalla *evolutionary psychology*, cioè la psicologia evolutiva, introdotta da E.O. Wilson nel 1975. Questa disciplina propone l'applicazione di principi evolutivi quali la competizione, la riproduzione e la selezione naturale alle interazioni sociali: in questo modo è possibile spiegare numerosi comportamenti umani. L'idea di Lo è quella di applicare questi stessi principi evolutivi ai mercati finanziari in modo tale da riconciliare la EMH con le idee dell'economia comportamentale. In quest'ottica, anche i comportamenti che apparivano irrazionali sotto la luce della EMH diventano ora giustificabili ed esplicabili.

Per rendere questa prospettiva evolutivistica operativa in un contesto economico occorre analizzare innanzitutto il concetto di *bounded rationality* esposto per la prima volta da Herbert Simons nel 1955. L'economista suggerì che gli individui non sono assolutamente in grado di eseguire la massimizzazione della propria utilità, che è quanto postulato dalla teoria economica neoclassica in tema di scelte del consumatore; questo tipo di ottimizzazione richiede capacità di calcolo e un costo difficili da sostenere per un soggetto. Quello che invece fanno i partecipanti nel mercato è ricercare una soluzione definita da Simons stesso *satisficing* – letteralmente tradotto con “soddisfacente” – che non necessariamente coincide con la soluzione ottimale.

Nonostante quest'idea valse a Simons il Premio Nobel, questa ebbe un impatto assai ridotto all'interno dell'ambiente accademico e professionale. Alla teoria di Simons mancava infatti un tassello fondamentale: come sa un individuo quando fermare il processo di ottimizzazione poiché viene raggiunto un punto per lui soddisfacente? La risposta a questo quesito è intuibile: utilizzando una prospettiva evolutivistica. Il punto soddisfacente non viene trovato utilizzando un metodo analitico, bensì attraverso un processo *trial and error*, cioè per tentativi, combinato con la selezione naturale. Gli individui basano quindi le loro scelte sull'esperienza passata e su quella che secondo loro è stata la miglior scelta fatta in passato, ricevendo in cambio una risposta positiva o negativa dall'ambiente circostante. A seconda del *feedback* ricevuto, un soggetto è in grado di capire quali scelte conviene ripetere e quali invece non hanno portato a buoni risultati, permettendogli di costruirsi un modello di comportamento che funziona in quel determinato ambiente. In sostanza, un individuo sviluppa euristiche⁵ sulla base del proprio bagaglio di esperienze e che funzionano fintanto che l'ambiente di riferimento rimane immutato. Ma cosa accade se l'ambiente circostante muta? Il modello di comportamento costruito sulla conformazione dell'ambiente precedente al cambiamento potrebbe risultare incompatibile con il nuovo ambiente: l'individuo si trova quindi a prendere scelte secondo un modello che prima funzionava ma che ora non è più adatto all'ambiente in cui opera. In questo caso osserviamo i famosi *bias* comportamentali, cioè azioni all'apparenza irrazionali ma che in realtà sono azioni che vengono viste al di fuori del loro contesto abituale. Lo definisce questo comportamento non tanto irrazionale ma *maladaptive*, che può essere tradotto “mal adattato”. Di seguito una citazione che rende chiarissimo il concetto di comportamento mal adattato [7]:

⁵ Una euristica è un insieme di regole intuitive che permette ad un individuo di elaborare un grosso quantitativo di informazioni riducendo la complessità del problema. Una euristica porta a risolvere un problema, alle volte complesso, in poco tempo ma con un margine di errore. Gli errori osservabili in questi casi sono i più volte citati *bias* o distorsioni cognitive.

«The flopping of a fish on dry land may seem strange and unproductive, but underwater the same motions are capable of propelling the fish away from its predators.»

Arrivati a questo punto è doveroso fare una precisazione: la AMH non è una teoria che si presta a sostituire completamente la EMH. Come dice lo stesso Lo «the EMH is not wrong; it is merely incomplete» [7]. I prezzi sul mercato riflettono effettivamente le informazioni disponibili, ma in che modo e in che misura lo fanno? Le informazioni vengono incorporate dai prezzi in base alla combinazione di due elementi: le condizioni dell'ambiente circostante e il numero e la natura delle "specie" presenti sul mercato. In pratica, il mercato finanziario può essere associato ad un ecosistema in cui diverse specie vivono. In questa analogia, le diverse specie cui si fa riferimento sono i diversi tipi di investitori: fondi comuni, piccoli risparmiatori, investitori istituzionali, sono tutti esempi di specie che popolano il mercato finanziario. In un ecosistema le diverse specie sopravvivono grazie alle risorse disponibili, nel caso dei mercati finanziari le risorse sono le opportunità di profitto. Se in un mercato molti investitori competono per accaparrarsi una parte di risorse relativamente scarse, allora il grado di efficienza di questo mercato sarà elevato; viceversa, in un mercato in cui le opportunità di profitto abbondano e sono gli investitori che vi partecipano ad essere scarsi il grado di efficienza sarà più basso.

Inoltre, proprio come in un ecosistema, i mercati finanziari sono in costante evoluzione e mutamento. Gli investitori che riescono a reinventarsi e ad adattare le proprie strategie di investimento in base alle condizioni del mercato in cui operano sono quelli che sopravvivono. Chi non è in grado di stare al passo con l'evoluzione del mercato non riesce più a sfruttare le opportunità di profitto e di conseguenza verrà espulso dal mercato. Si tratta di pura e semplice selezione naturale.

Il nuovo *framework* descritto da Lo con la sua Teoria dei Mercati Adattivi è ancora in fase di studio e di sviluppo, soprattutto dal punto di vista operativo e prettamente matematico, ma non si può negare che questa teoria sia in grado di riconciliare molti degli aspetti contraddittori nella EMH.

1.3.1 Un nuovo ordine mondiale: le implicazioni della AMH

Il *framework* appena descritto ci obbliga a rivedere alcuni dei punti che con la EMH venivano dati ormai per scontati.

Tanto per iniziare, la relazione tra rischio e rendimento esiste ma è tutt'altro che lineare e stabile nel corso del tempo. Questa relazione viene determinata attraverso le interazioni, le categorie di investitori che partecipano attivamente al mercato finanziario nonché attraverso le preferenze di questi. Nel corso del tempo le combinazioni di questi fattori mutano, conseguentemente anche le relazioni rischio-rendimento si adattano a tali cambiamenti. Ci sono periodi in cui il trade-off tra rischio e rendimento si inverte diventando addirittura negativo. Questo accade in periodi in cui sul mercato finanziario c'è tensione con una dinamica che sembra seguire una logica: nei periodi di tensione finanziaria, gli investitori tendono a smobilizzare i propri investimenti rischiosi per direzionarsi verso *asset* più sicuri. Un comportamento più che logico, si potrebbe pensare. L'effetto che questo comportamento ha sui rendimenti dei diversi titoli va però contro ogni logica finanziaria, poiché il rendimento medio dei titoli rischiosi si abbassa mentre il rendimento medio dei titoli meno rischiosi si alza. Ecco quindi che la relazione rischio-rendimento si interrompe e si inverte, andando contro uno dei pilastri della classica teoria finanziaria. Nei periodi in cui i prezzi sul mercato fluttuano all'interno di un range calcolato in base ai dati storici – in altre parole, quando i mercati non si allontanano dal seminato – allora i paradigmi propri della EMH ritornano in scena e il mercato risulta essere davvero efficiente.

Una seconda implicazione derivabile dalla AMH è che le opportunità di arbitraggio esistono e sono sfruttabili. I primi a gettare le basi a quest'idea furono Grossman e Stiglitz nel 1980 [6], i quali sostenevano che le opportunità di arbitraggio fossero l'unico incentivo per gli investitori a operare sul mercato; in assenza di tali opportunità i mercati sarebbero giunti ad un punto di collasso in quanto il costo per raccogliere informazioni sui prezzi degli *asset* supera di gran lunga i profitti ottenibili con tali informazioni. Le opportunità di arbitraggio, quindi, devono esistere e spariscono non appena queste vengono sfruttate. Nuove opportunità vengono continuamente create, dando origine a fenomeni quali cicli e trend, bolle speculative, crolli finanziari e via dicendo. Queste opportunità sono la vera motivazione e l'incentivo ad una attività di trading e di *active management*, che altrimenti non avrebbe alcun motivo per esistere.

La terza implicazione riguarda invece le strategie di investimento: non esiste una strategia buona e una strategia cattiva. Tutte le strategie funzionano bene in alcuni periodi e

ambientanti mentre performano male se applicate in altri contesti. Come detto poc'anzi, non si tratta di strategie sbagliate, si tratta solo di strategie mal adattate. Solo perché una strategia di investimento ha funzionato bene per anni non vuol dire che continuerà a farlo per l'eternità, i mercati cambiano e di conseguenza anche le strategie di investimento dovrebbero cambiare con loro.

Quest'ultima frase ci collega direttamente a quella che è la quarta e ultima implicazione, riassumibile con una sola parola: innovazione. Sul mercato finanziario sopravvive solo chi è dinamico, chi è in grado di adattarsi velocemente ai cambiamenti del mercato stesso. Questo discorso si collega anche con la prima implicazione, ossia quella legata alla relazione mutevole tra rischio e rendimento: se questa relazione non è stabile nel tempo, il modo migliore per riuscire ad ottenere un soddisfacente livello di rendimento lungo un periodo di tempo è essere pronti ad adattarsi ad ogni minimo cambiamento che si registra sul mercato.

Per concludere, esiste un unico obiettivo sul mercato finanziario, ossia *sopravvivere*. Non si tratta solo di ottenere dei profitti, si tratta di riuscire a farlo per il maggior tempo possibile e questo si realizza solo se si è in grado di non farsi espellere dal mercato. Sviluppare strategie in grado di rendere una fortuna non ha alcun senso se l'orizzonte temporale è limitato. La massimizzazione del profitto e dell'utilità, l'equilibrio generale del mercato, sono tutti aspetti utili e rilevanti sul mercato ma che alla fine dei giochi sono tutti elementi di contorno all'unico scopo, ossia la sopravvivenza.

Il Reinforcement Learning

Quante volte ci siamo sentiti dire dai nonni o dai genitori che “sbagliando s’impara”? Dietro questo semplice proverbio proveniente dalla cultura popolare è radicata un’idea ben più complessa riguardo il metodo di apprendimento degli individui. Il metodo di apprendimento legato all’interazione tra individuo e ambiente è forse una delle idee principali che sovrviene quando si pensa alla natura dell’apprendimento stesso. Detta in parole più semplici, si tratta del meccanismo di apprendimento che applichiamo fin da neonati: un bambino piccolo acquisisce consapevolezza dell’ambiente che lo circonda eseguendo tutte le azioni che gli sono permesse e valutandone le conseguenze, ad esempio capisce che non deve toccare i fornelli accesi dopo essersi scottato. Non si tratta quindi di un’astrusa idea innovativa, l’apprendimento per rinforzo è il metodo di apprendimento più spontaneo e immediato che esista. Quello che è invece innovativo è il *reinforcement learning* (d’ora in avanti abbreviato in RL) applicato in ambito computazionale: il punto di interesse non è più capire come funziona il metodo di apprendimento ma è cercare di creare macchine intelligenti in grado di applicare tali metodi di apprendimento per risolvere problemi di interesse comune in ambito economico-scientifico.

Reinforcement learning significa imparare come comportarsi e quali azioni intraprendere in determinate situazioni grazie all’interazione con l’ambiente in cui si è immersi. L’agente deve quindi scoprire da solo quali sono le azioni che con il tempo lo porteranno a raggiungere il proprio scopo e quali invece non gli sono di alcuna utilità nel raggiungimento del proprio obiettivo. Ad esempio, un giocatore di scacchi professionista valuta la posizione di ogni pedina sulla scacchiera e sceglie la propria mossa sulla base di queste informazioni; esistono azioni che lo porteranno a vincere la partita, altre che invece non avranno conseguenze positive poiché possono portare un vantaggio per la prossima mossa dell’avversario.

Le conseguenze fino ad ora citate vengono definite come *ricompense* a seguito dell’azione intrapresa: si tratta del cosiddetto “rinforzo” che permette all’agente di acquisire esperienza

e formare un modello di comportamento. Sono quindi due le caratteristiche che definiscono e distinguono il RL:

1. Un processo *trial-and-error*, altrimenti detto “per tentativi”. L’agente di fatto non conosce la struttura dell’ambiente circostante e l’unico modo per acquisire esperienza è provare tutte le azioni a sua disposizione per vedere quali sono quelle che gli portano ricompense maggiori;
2. La ricompensa “ritardata”, poiché la ricompensa arriva dopo aver raggiunto il nuovo stato e dopo aver valutato se l’azione che ha portato a quello stato è stata una buona scelta o meno. In sostanza, la ricompensa è quello che permette all’agente di mappare quali sono le azioni migliori (in termini di ricompensa) e quali invece sono azioni che non conviene ripetere.

In sostanza, l’agente ha un obiettivo ben preciso, ossia quello di massimizzare le ricompense scegliendo attentamente le azioni che gli portano le ricompense migliori. Tuttavia è interessante per l’agente massimizzare la ricompensa immediata ma allo stesso tempo cercare di massimizzare la somma, eventualmente pesata, delle ricompense derivanti dalla successione di azioni che deciderà di intraprendere. Sfortunatamente, non esiste un modello comportamentale ottimale da seguire: un agente immerso in un ambiente non conoscendone la struttura e le caratteristiche dovrà da solo scoprire quali azioni sono migliori in termini di ricompensa. Proprio questo fatto ci porta ad una delle sfide del RL: il *trade-off* tra l’esplorazione e lo sfruttamento.

I concetti di esplorazione e sfruttamento sono fondamentali quando si parla di RL. Per ottenere i risultati migliori l’agente deve preferire azioni che in passato si sono già rivelate vincenti; allo stesso tempo, l’agente deve verificare che queste stesse azioni siano effettivamente le migliori a sua disposizione o se effettivamente ne esistono altre in grado di fruttargli ricompense migliori. Nel primo caso l’agente *sfrutta* le conoscenze che già possiede, nel secondo *esplora* stati dell’ambiente non ancora visitati intraprendendo azioni mai intraprese prima. Il dilemma sta nella combinazione dei due elementi, ossia quanto un agente deve sfruttare azioni già conosciute e quanto invece deve esplorare nuove azioni.

2.1 Gli elementi del Reinforcement Learning

Sicuramente alla base di un sistema RL c'è la relazione tra agente e ambiente; oltre a questi elementi imprescindibili esistono quattro importanti subelementi necessari al funzionamento di tutto il sistema: si tratta di un *criterio comportamentale*⁶, una *funzione di ricompensa*, una *funzione di valore* e un *modello dell'ambiente* circostante⁷.

Il *criterio comportamentale*, indicato con π , rappresenta la mappatura sviluppata nel tempo di tutti i possibili stati del mondo e delle azioni che un agente può intraprendere quando si trova in un determinato stato. Questo criterio da solo è in grado di definire lo schema comportamentale dell'agente, in altre parole il criterio π definisce il modo in cui l'agente seleziona le azioni a seconda dello stato in cui si trova. La notazione che viene utilizzata per indicare il criterio comportamentale è $\pi(s|a)$, ovvero la probabilità che venga intrapresa l'azione $a_t = a$ dato lo stato in cui versa l'agente $s_t = s$ per ogni $s \in \mathcal{S}$ e $a \in \mathcal{A}(s)$.

Il *sistema di ricompensa* definisce l'obiettivo dell'agente in un problema RL. A ciascuno stato viene associato un valore numerico, cioè la ricompensa, che indica quanto quello stato è desiderabile per l'agente. La funzione di ricompensa non è in alcun modo alterabile dall'agente: quest'ultimo non può controllare il valore delle ricompense che riceve ma può indirettamente alterarne il valore modificando la propria *policy*, ossia modificando il criterio comportamentale π . Come già anticipato, l'obiettivo dell'agente è di massimizzare l'ammontare, eventualmente pesato, delle ricompense che riceve sul lungo periodo: questo ci introduce al prossimo elemento chiave del RL, che è la *funzione di valore*.

La *funzione di valore* rappresenta il valore totale delle ricompense accumulate sul lungo periodo. Il sistema di ricompensa e la funzione di valore sono in realtà due facce della stessa medaglia: entrambe sono l'espressione della desiderabilità di trovarsi in uno stato piuttosto che in un altro, la differenza tra le due sta nell'orizzonte temporale su cui queste valutazioni vengono fatte. La funzione di ricompensa è un *feedback* immediato sulla desiderabilità dello stato s in cui l'agente è giunto a seguito dell'azione a intrapresa; la funzione di valore utilizza invece un approccio più lungimirante e di fatto si tratta del valore attuale di tutte le ricompense che l'agente si aspetta di ottenere partendo dallo stato in cui si trova. Bisogna prestare attenzione poiché le azioni che portano ad una immediata ricompensa elevata

⁶ Il termine inglese utilizzato è *policy*.

⁷ Quest'ultimo elemento non è sempre presente ma verrà comunque esposto per completare il *framework* teorico del RL.

potrebbero in realtà avere una funzione di valore poco soddisfacente sul lungo periodo e ovviamente può essere valido anche il caso contrario.

L'obiettivo dell'agente è quindi quello di massimizzare questa funzione di valore e per farlo è necessario ricorrere a delle stime di questo valore, sulla base delle quali l'agente prenderà poi le proprie decisioni. L'utilizzo di stime della funzione di valore si rende necessario poiché si lavora effettivamente con valori futuri ancora sconosciuti: come detto in precedenza, la funzione di valore si basa sul valore delle ricompense *attese*, si tratta quindi di ricompense che l'agente ancora non ha ottenuto ma che si aspetta di ottenere nel corso del tempo. In questo modo il valore che si ottiene è una stima della funzione di valore, pertanto potrebbe anche differire dall'effettivo valore che questa assume una volta che si potranno osservare le ricompense effettive.

Le ricompense derivano direttamente dall'ambiente in cui l'agente è immerso mentre la funzione di valore è a tutti gli effetti una funzione che necessita di essere calcolata e, se necessario, aggiustata ad ogni passaggio, il che la rende effettivamente la vera protagonista in un problema RL.

In ultima battuta troviamo il *modello ambientale*, un elemento non sempre presente. Questo elemento ha lo scopo di rappresentare il comportamento dell'ambiente di riferimento. Ad esempio, dati uno stato di partenza e un'azione intrapresa, il modello ambientale potrebbe essere in grado di prevedere quale sarà il prossimo stato e quale sarà il valore della ricompensa connessa a quest'ultimo. Si tratta di un elemento non strettamente necessario, in quanto quello che fa è una sorta di *pianificazione*⁸ delle mosse dell'agente e, dato che un processo RL è sostanzialmente un processo per tentativi, sembrerebbe che i due aspetti possano essere in contrasto tra loro. In realtà il modello ambientale diventa utile dopo una fase di esplorazione – e quindi per tentativi – in cui l'agente impara a conoscere l'ambiente e successivamente, grazie al modello ambientale, sceglie quali azioni intraprendere.

Di seguito, nella Figura 1.1, viene proposto uno schema riassuntivo del funzionamento di un sistema RL e della relazione che intercorre tra agente e ambiente:

⁸ Con il termine pianificazione si intende un metodo decisionale riguardo l'ordine delle azioni da intraprendere considerando tutti i possibili stati in cui l'agente si troverebbe prima ancora che queste azioni vengano effettivamente intraprese.

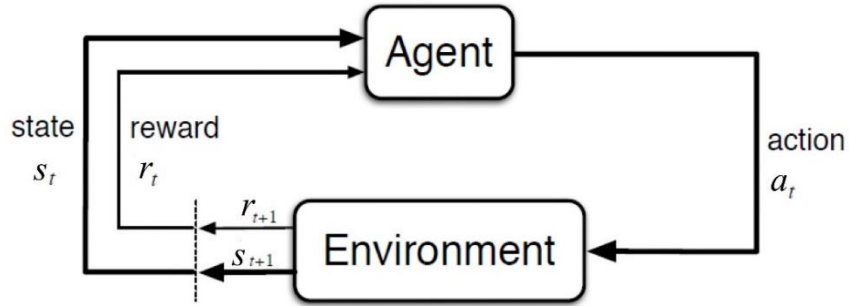


Figura 1.1: Diagramma dell'interazione tra agente e ambiente (Sutton e Barto, 2018)

Ad ogni step temporale t l'agente riceve un input dallo stato dell'ambiente nel quale si trova, denominato come $s_t \in \mathcal{S}$ dove \mathcal{S} rappresenta tutti gli stati del mondo possibili. Sulla base di questo segnale l'agente seleziona l'azione da intraprendere $a_t \in \mathcal{A}(s_t)$, dove $\mathcal{A}(s_t)$ rappresenta l'insieme delle azioni disponibili quando l'agente si trova nello stato s_t . Dopo aver selezionato ed eseguito l'azione, nello step temporale successivo l'agente riceverà una ricompensa numerica r_{t+1} a seconda dello stato s_{t+1} raggiunto. Ad ogni step temporale t l'agente implementa una sorta di mappatura delle probabilità di raggiungere un determinato stato selezionando ogni azione disponibile. Questa mappatura altro non è che il criterio comportamentale π , il quale varia nel tempo a seconda dell'esperienza che l'agente acquisisce. Il ciclo appena descritto continua fino a che l'obiettivo del sistema RL implementato non viene raggiunto.

2.2 I processi decisionali alla Markov

Nel *framework* del Reinforcement Learning l'agente intraprende un processo decisionale basato su un input o segnale fornitogli dallo stato dell'ambiente in cui si trova in un dato momento. L'agente non può in alcun modo modificare il segnale che gli viene fornito, inoltre le informazioni contenute in questo impulso non sono individuate *tutte* le informazioni possibili in quanto esistono stati ambientali che ancora non sono stati visitati dall'agente. Sicuramente sarebbe utile all'agente conoscere la totalità delle informazioni possibili, cioè tutto \mathcal{S} , ma le azioni per visitare certi stati non sono ancora state eseguite o, in alternativa, gli impulsi a visitare taluni stati non sono ancora stati ricevuti. L'ideale per l'agente sarebbe ricevere un impulso in grado di contenere tutte le informazioni sull'esperienza passata, che

può essere utile a fini decisionali, ed evidenziando al contempo quelle più rilevanti per la scelta dell'azione da intraprendere. Con un segnale del genere, chiamato *segnale à la Markov*, l'agente potrebbe essere in grado di scegliere l'azione più adeguata al contesto in cui si trova. Si prenda ad esempio la posizione di una pedina sulla scacchiera: le coordinate che ne definiscono la posizione sono un esempio di segnale alla Markov poiché tutto quello che è necessario per scegliere la prossima mossa è contenuto in quelle informazioni, a prescindere dal percorso che ha portato la pedina in quella posizione. Formalmente la Proprietà Markoviana si sviluppa su due momenti temporali, rispettivamente il momento t e il momento $t + 1$. Nel momento t l'agente esegue un'azione a_t partendo dallo stato s_t e giunge nello stato s_{t+1} nel momento $t + 1$. La dinamica è esattamente quella descritta dal diagramma della Figura 1.1 visto precedentemente. Nel caso più generale, la risposta dell'ambiente in $t + 1$, vale a dire la ricompensa r_{t+1} che l'agente riceverà in seguito all'azione a_t , dipende da tutto l'accaduto precedente a quel preciso momento, ossia:

$$P[s_{t+1} = s', r_{t+1} = r | s_t, a_t, r_t, s_{t-1}, a_{t-1}, \dots, r_1, s_0, a_0] \quad 4$$

Questa formula rappresenta la probabilità di trovarsi in $t + 1$ in uno stato $s_{t+1} = s'$ e di ricevere una ricompensa $r_{t+1} = r$ condizionatamente a tutto quello accaduto tra il momento 0 e il momento t . Tuttavia, se un segnale possiede la Proprietà Markoviana le uniche informazioni necessarie per calcolare la probabilità dell'equazione (6) sono le informazioni del momento temporale t , pertanto la formula diventa:

$$P[s_{t+1} = s', r_{t+1} = r | s_t = s, a_t = a] \quad 7$$

Il sistema passa quindi dallo stato s_t allo stato s_{t+1} basandosi esclusivamente sulle informazioni contenute nello stato e nell'azione precedenti, rendendo non necessarie al fine di tale calcolo tutte le informazioni contenute negli stati e nelle conseguenti azioni precedenti al momento t .

2.3 Obiettivo e sistema di ricompensa

Come più volte ripetuto in precedenza, l'obiettivo dell'agente in un sistema RL è quello di massimizzare la ricompensa cumulata lungo un periodo di tempo e non la massimizzazione della ricompensa immediata conseguente all'azione appena compiuta. Ciascuna di queste ultime ricompense è rappresentata da un singolo numero e, considerando che il processo parte da un momento generico t , è possibile denotare ciascuna ricompensa in questo modo: $r_{t+1}, r_{t+2}, r_{t+3}, \dots$. È possibile pensare a queste ricompense come ad un qualcosa di più concreto, come ad esempio il rendimento di un'attività finanziaria: sommando tutte queste componenti r_t immaginandole come i rendimenti giornalieri di un qualsiasi titolo si ottiene R_t , il quale rappresenta il rendimento che si ottiene al termine dell'*holding period* di un titolo. Dato che l'agente persegue lo scopo della massimizzazione delle ricompense cumulate e la durata dell'orizzonte temporale è sconosciuta – e di conseguenza potenzialmente infinita –, deve selezionare le azioni tenendo conto anche del fattore temporale scontando le ricompense. L'agente pertanto seleziona l'azione a_t al momento t tenendo conto di dover massimizzare il valore attuale dei rendimenti futuri, ossia:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad 8$$

dove γ è il *tasso di sconto* ed ha un valore compreso tra 0 e 1. Più il valore del tasso di sconto è vicino a 1 è più il comportamento dell'agente sarà “lungimirante”; al contrario un valore del tasso di sconto vicino a 0 determina un comportamento dell'agente focalizzato principalmente sugli effetti immediati delle proprie decisioni. Anche se un tasso di sconto vicino a 0 sembra portare a ricompense migliori nell'immediato, con un criterio comportamentale del genere l'agente potrebbe precludersi la possibilità di ottenere rendimenti cumulati più alti sul lungo periodo. È necessario quindi che il tasso di sconto sia ben bilanciato in modo tale da garantire all'agente il giusto equilibrio tra un comportamento lungimirante ed un comportamento concentrato sull'immediato futuro.

2.4 La funzione di valore

Come già anticipato in uno dei paragrafi precedenti la funzione di valore è il vero cuore di un sistema RL: essa rappresenta la desiderabilità per l'agente di trovarsi in un determinato stato s_t o, alternativamente, la performance partendo da una determinata azione a_t e di quelle ad essa successive. Il valore di uno stato (o di una coppia stato-azione) viene calcolato in termini di ricompense attese: tali ricompense dipendono da quali azioni verranno intraprese dall'agente, in altre parole dipendono dal criterio comportamentale π seguito dall'agente. Il valore di uno stato viene quindi indicato con $V^\pi(s)$, notazione che indica la *funzione di valore* definita nel modo seguente:

$$V^\pi(s) = \mathbb{E}_\pi\{R_t | s_t = s\} = \mathbb{E}_\pi\left\{\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s\right\} \quad 9$$

dove \mathbb{E}_π rappresenta il valore atteso dei rendimenti, in cui il pedice π sta ad indicare il fatto che si assume che l'agente segua il criterio comportamentale π . Tale funzione viene definita *state-value function*, che è possibile tradurre con funzione di valore di stato.

Analogamente, è possibile calcolare la *action-value function*, cioè la funzione di valore dell'azione a per il criterio π . Tale funzione viene indicata con $Q^\pi(s, a)$ ed è definita nel modo seguente:

$$Q^\pi(s, a) = \mathbb{E}_\pi\{R_t | s_t = s, a_t = a\} = \mathbb{E}_\pi\left\{\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s, a_t = a\right\}. \quad 10$$

La funzione appena presentata indica il valore qualora l'agente, seguendo ovviamente il criterio comportamentale π , scelga di perseguire l'azione $a_t = a$ partendo dallo stato $s_t = s$. Entrambe le equazioni vengono stimate durante il processo attraverso l'esperienza che l'agente accumula sfruttando l'interazione continua con l'ambiente.

Le funzioni di valore appena presentate rispettano una particolare relazione ricorsiva tra lo stato attuale $s_t = s$ ed ogni possibile stato futuro, valida per ogni *policy*. L'espressione della funzione di valore di stato $V^\pi(s)$ può essere definita attraverso quella che viene chiamata *Equazione di Bellman*:

$$\begin{aligned}
V^\pi(s) &= \mathbb{E}_\pi\{R_t | s_t = s\} & 11 \\
&= \mathbb{E}_\pi\left\{\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s\right\} \\
&= \mathbb{E}_\pi\left\{r_{t+1} + \gamma \sum_{k=0}^T \gamma^k r_{t+k+2} | s_t = s\right\} \\
&= \mathbb{E}_\pi\{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s\}
\end{aligned}$$

Questa relazione stabilisce che il valore dello stato iniziale deve essere uguale al valore attuale atteso dello stato successivo più il valore della rispettiva ricompensa ottenuta dopo il passaggio dallo stato s_t allo stato s_{t+1} .

L'equazione di Bellman può essere applicata in maniera analoga anche al caso della funzione di valore dell'azione $Q^\pi(s, a)$:

$$Q^\pi(s, a) = \mathbb{E}_\pi\{r_{t+1} + \gamma Q^\pi(s_{t+1}, a_t = a) | s_t = s, a_t = a\} \quad 12$$

Le soluzioni possibili alle equazioni (11) e (12) sono solo ed esclusivamente i valori $V^\pi(s)$ e $Q^\pi(s, a)$, con s ed a specificati, e partendo da questo elemento è possibile ora introdurre l'ottimizzazione della funzione di valore.

2.4.1 L'ottimizzazione della funzione di valore

Risolvere un problema RL consiste nel ricercare un modello di comportamento π che porti ad ottenere il rendimento più alto possibile. Si tratta quindi di un problema di ottimizzazione volto a massimizzare il rendimento esplorando nuovi stati e/o azioni sfruttando l'esperienza acquisita. Innanzitutto, è opportuno definire quando un criterio π' è migliore del criterio π : un criterio comportamentale π' è da considerare migliore di un criterio π se il rendimento atteso di π' è maggiore o uguale del rendimento atteso di π per ogni $s \in \mathcal{S}$. Come illustrato precedentemente il rendimento atteso di un dato criterio comportamentale è la definizione della funzione di valore, è quindi possibile tradurre in linguaggio matematico quanto appena scritto nel modo seguente: $\pi' \geq \pi$ se e solo se $V^{\pi'}(s) \geq V^\pi(s)$ per ogni stato $s \in \mathcal{S}$ o, in alternativa, $\pi' \geq \pi$ se e solo se $Q^{\pi'}(s, a) \geq Q^\pi(s, a)$ per ogni $s \in \mathcal{S}$ ed $a \in \mathcal{A}(s)$.

Ovviamente vale la stessa cosa nel caso in cui si utilizzi la funzione di valore dell'azione $Q^\pi(s, a)$. Esiste sempre almeno un *criterio ottimo* π^* che è migliore o tutt'al più uguale agli altri criteri comportamentali, al quale viene associata la *funzione di valore ottima* V^* (o in alternativa Q^*) definita nel modo seguente:

$$V^*(s) = \max_{\pi} V^\pi(s) \quad \text{per ogni } s \in \mathcal{S} \quad 13$$

$$Q^*(s, a) = \max_{\pi} Q^\pi(s, a) \quad \text{per ogni } s \in \mathcal{S} \text{ e } a \in \mathcal{A} \quad 14$$

Dato che V^* e Q^* sono funzioni di valore allora rispettano anche stavolta la relazione ricorsiva dettata dall'equazione di Bellman. In questo caso essendo le due funzioni ottime tale relazione prende il nome di *equazione ottima di Bellman* e, come nel caso precedente, esprime la relazione tra il valore di uno stato ed il valore dello stato successivo tenendo conto del rendimento che si ottiene passando dallo stato iniziale a quello successivo, il tutto a condizione che venga seguita la *policy* ottima π^* . L'equazione ottimale di Bellman per V^* viene rappresentata nel modo seguente:

$$\begin{aligned} V^*(s) &= \max_{a \in \mathcal{A}} Q^{\pi^*}(s, a) & 15 \\ &= \max_{a \in \mathcal{A}} \mathbb{E}_{\pi^*} \{R_t | s_t = s, a_t = a\} \\ &= \max_{a \in \mathcal{A}} \mathbb{E}_{\pi^*} \left\{ \sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s, a_t = a \right\} \\ &= \max_{a \in \mathcal{A}} \mathbb{E}_{\pi^*} \left\{ r_{t+1} + \gamma \sum_{k=0}^T \gamma^k r_{t+k+2} | s_t = s, a_t = a \right\} \\ &= \max_{a \in \mathcal{A}} \mathbb{E}_{\pi^*} \left\{ r_{t+1} + \gamma \sum_{k=0}^T \gamma^k V^{\pi^*}(s_{t+1}) | s_t = s, a_t = a \right\} \end{aligned}$$

mentre per Q^* sarà:

$$Q^*(s, a) = \mathbb{E}_{\pi} \left\{ r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a') | s_t = s, a_t = a \right\} \quad 16$$

L'equazione ottima di Bellman consiste in realtà in un sistema di equazioni composto da tante equazioni quanti sono gli N stati in N incognite, pertanto è necessario risolvere tale sistema per $V^*(s)$ o $Q^*(s, a)$ per poter procedere.

Una volta ottenuta $V^*(s)$ bisogna determinare il criterio ottimo π^* associato a tale funzione andando a scegliere l'azione che garantisce il rendimento maggiore tra tutte le azioni disponibili in quello stato. Potrebbe sembrare sbagliato un approccio che guardi solo alle conseguenze immediate dell'azione poiché nei paragrafi precedenti si è detto che un comportamento dell'agente orientato al profitto sul lungo periodo è da preferire; in questo caso, tuttavia, si sta considerando il valore ottimo della funzione di valore V^* pertanto selezionare le azioni con il migliore rendimento immediato al contempo permette di massimizzare il rendimento sul lungo periodo. Un criterio di selezione come quello appena descritto viene definito *greedy*, letteralmente traducibile con *ingordo* o *famelico*⁹, il quale valuta la scelta dell'azione sulla base della ricompensa attesa che ne deriva.

2.4.2 Valutazione e perfezionamento del criterio π

Un approccio RL per raggiungere il proprio obiettivo deve stimare diverse funzioni di valore di stato – o equivalentemente funzioni di valore dell'azione – con lo scopo di determinare criteri comportamentali ottimi. Tale processo di stima si sviluppa in modo iterativo e si articola in diversi step.

Il primo passo in questo processo iterativo è la *valutazione* del criterio π , la quale consiste nell'approssimare attraverso stime consecutive una funzione di valore $V^\pi(s)$ per un criterio π scelto arbitrariamente. Tale procedimento comincia da un valore iniziale $V_0^\pi(s_t)$ scelto in modo arbitrario; ogni sua successiva approssimazione si ottiene utilizzando l'equazione di Bellman come regola di aggiornamento, la quale aggiorna il valore $V_k^\pi(s_t)$ in $V_{k+1}^\pi(s_t)$ ¹⁰ nel modo seguente:

$$V_{k+1}^\pi(s_t) = \mathbb{E}_\pi\{r_{t+1} + \gamma V_k^\pi(s_{t+1}) | s_t = s\} \quad 17$$

⁹ Si preferirà mantenere comunque il termine inglese originale in quanto considerato più appropriato.

¹⁰ Il pedice k e $k + 1$ indicano il numero dell'iterazione durante il processo di stima del valore dello stato s_t .

In questo modo la stima viene aggiornata applicando la stessa regola ad ogni valore di s in modo da calcolare l'approssimazione successiva $V_{k+1}^\pi(s_t)$. Un procedimento come quello appena descritto assicura la convergenza della sequenza $\{V_k^\pi\}$ al suo valore ottimo V^* se $k \rightarrow \infty$ e $\gamma < 1$. In genere le iterazioni della formula (17) vengono effettuate finché non viene raggiunta una condizione finale che tipicamente si ottiene osservando la quantità $|V_{k+1}^\pi(s) - V_k^\pi(s)|$ al termine di ogni iterazione, terminando l'esecuzione dell'algoritmo se questa dovesse essere prossima a zero.

Lo step successivo è il *perfezionamento* del criterio π , il quale consiste nella ricerca di un criterio π' migliore del precedente π al fine di incrementare il valore atteso delle ricompense future. Per alcuni stati può essere interessante scegliere un'azione $a \notin \pi(s)$, che quindi non appartiene al criterio π in esame, per vedere quali effetti produce sulla funzione di valore. Il processo di perfezionamento calcola il valore di tutte le azioni attraverso la funzione di valore dell'azione $Q^\pi(s, a)$ e successivamente la confronta con $V^\pi(s)$ per verificare se esistono azioni $a \in \mathcal{A}$ per cui $Q^\pi(s, a) > V^\pi(s)$. Se quest'ultima condizione si verifica, allora può essere conveniente selezionare un'azione $a \notin \pi(s)$ in s e successivamente seguire il criterio π . L'azione da intraprendere ad ogni step viene selezionata in base al sistema seguente:

$$a = \begin{cases} \pi'(s) \text{ con probabilità } 1 - \varepsilon \\ a \in \mathcal{A}(s) \text{ con probabilità } \varepsilon \end{cases} \quad 18$$

dove $\varepsilon \in (0,1)$ e π' è il criterio *greedy* che massimizza la funzione $Q^\pi(s, a)$:

$$\begin{aligned} \pi'(s) &= \arg \max_a Q^\pi(s, a) \\ &= \arg \max_a \mathbb{E}_\pi \{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s, a_t = a\} \end{aligned} \quad 19$$

in cui $\arg \max_a(\cdot)$ indica l'azione a che massimizza l'espressione (19). Questo approccio di selezione si basa in sostanza su due *policies*:

- Una *greedy policy*, selezionata con probabilità $1 - \varepsilon$. Selezionare un'azione secondo questo criterio significa sfruttare l'esperienza già acquisita sul valore di ciascuna azione;

- Una *non-greedy policy*, selezionata con probabilità ε . Selezionare un'azione tra quelle disponibili seguendo tale criterio significa andare ad esplorare azioni mai selezionate prima al fine di verificare se queste sono migliori di quelle proprie della *greedy policy*.

Combinando il processo di valutazione con il processo di perfezionamento si genera un algoritmo definito *iterazione generalizzata del criterio π* . Questo procedimento si articola in due fasi che si susseguono continuamente: la prima fase è ovviamente la valutazione del criterio, a seguito della quale l'algoritmo viene fermato per procedere immediatamente al perfezionamento del criterio appena valutato. Questo genere di processo iterativo viene rappresentato nel modo seguente:

$$V_{k+1}^\pi(s_t) = \max_a \mathbb{E}_\pi \{r_{t+1} + \gamma V_k^\pi(s_{t+1}) | s_t = s, a_t = a\} \quad 20$$

dove $V_{k+1}^\pi(s_t)$ rappresenta la stima aggiornata grazie al perfezionamento del criterio allo step $k + 1$, tenuta in considerazione la stima del criterio fatta allo step k .

In sintesi, è possibile rappresentare la combinazione dei due processi con uno schema semplificato:

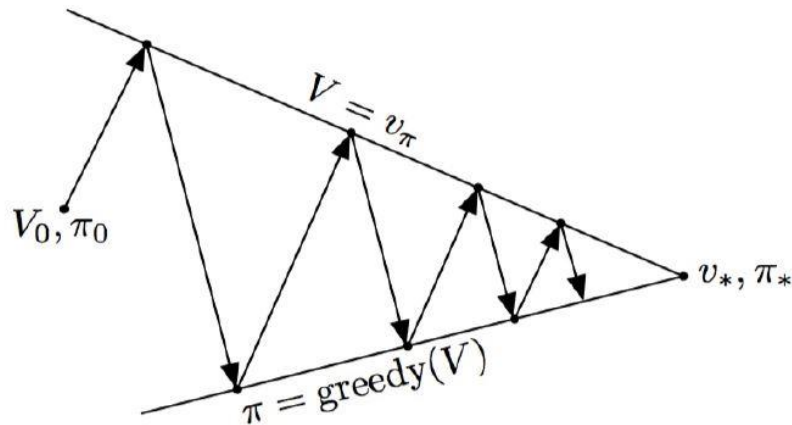


Figura 2.1: Rappresentazione grafica del processo di iterazione generalizzata del criterio π (Barto e Sutton, 2018)

Partendo da un valore iniziale di V_0 e da un criterio iniziale π_0 il processo porta l'uno o l'altro verso il rispettivo valore ottimo. Le frecce indicano il percorso mentre il processo di

iterazione tenta di raggiungere una delle due soluzioni ottime. Inevitabilmente, il processo finirà per convergere ad una soluzione comune.

2.5 I metodi TD: apprendimento per differenze temporali

Quando si parla di *Reinforcement Learning* si fa riferimento ad una categoria di metodi di apprendimento specifica. All'interno di questa categoria è possibile trovare diversi approcci all'apprendimento, come i metodi Monte Carlo o i metodi per differenze temporali.

Questi ultimi – d'ora in avanti abbreviati con TD¹¹ – sono l'oggetto del presente lavoro e verranno illustrati in questa sede. Tali metodologie utilizzano le differenze tra le funzioni di valore di due step temporali consecutivi per aggiornare le stime delle funzioni di valore stesse, senza la necessità che il processo raggiunga il risultato finale e impiegando in parte le stime già esistenti. Questa metodologia viene definita *bootstrapping*¹² e non richiede l'utilizzo di un modello ambientale. I metodi TD fanno quindi uso di tutta l'esperienza a loro disposizione per aggiornare passo dopo passo le stime dei valori durante il processo di valutazione. Per ogni stato s_t non terminale i metodi TD calcolano una stima V della funzione di valore ottima V^* basandosi sul rendimento r ottenuto all'istante temporale $t + 1$.

Il metodo TD più semplice, definito con $TD(0)$, considera un solo step temporale in avanti rispetto allo step in cui viene effettuata la stima e utilizza la seguente regola di aggiornamento:

$$V_{k+1}(s_t) = V_k(s_t) + \alpha[r_{t+1} + \gamma V_k(s_{t+1}) - V_k(s_t)] \quad 21$$

dove $\alpha \in (0,1]$ rappresenta il *tasso di apprendimento*, detto anche *parametro step-size*, il quale misura il grado di aggiornamento della stima. La differenza tra le parentesi viene definita *errore TD* ed esprime la distanza che all'iterazione k – *esima* intercorre tra la stima $V_k(s_t)$ e il target $r_{t+1} + \gamma V_k(s_{t+1})$: indicando questa quantità con δ_t è possibile sapere a quanto ammonta l'errore di stima in quell'istante temporale indicando di conseguenza la possibile direzione per il comportamento futuro. I metodi della classe TD cercano di minimizzare la

¹¹ L'abbreviazione TD deriva dalla denominazione inglese di questa classe di metodi, ossia *Temporal-Difference learning*.

¹² In generale, il termine *bootstrapping* si riferisce ad un processo in grado procedere senza bisogno di alcun input esterno dopo che è stato avviato.

differenza tra stime consecutive e conseguentemente aggiustando e perfezionando ad ogni istante temporale il criterio comportamentale. Il valore di δ_t dipende dal valore dello stato successivo a quello di riferimento e dal valore del rendimento che si ottiene nel passaggio dallo stato attuale a quello successivo; questo significa che l'errore TD può essere calcolato dopo aver osservato tali realizzazioni e solo poi l'algoritmo procede all'aggiornamento delle stime. Come già anticipato, gli algoritmi TD non necessitano di un modello ambientale né della distribuzione delle probabilità di transizione di cui si è parlato precedentemente. Di fatto questi algoritmi producono un segnale di apprendimento ad ogni step eseguito indipendentemente dall'azione selezionata, si aggiornano al termine di ogni step e non è necessario che venga raggiunto lo stato terminale affinché le stime vengano aggiornate. Questa metodologia viene definita *incrementale on-line* e risulta essere particolarmente utile per risolvere problemi di controllo dinamico legati al mondo dei mercati finanziari.

A tal proposito, è ben noto che i mercati finanziari non seguono di certo un andamento stazionario, il che significa che le serie dei prezzi hanno almeno la media e la varianza che variano nel tempo. In questi casi risulta essere più appropriato dare più peso ai rendimenti recenti rispetto ai rendimenti più lontani nel tempo. Un risultato del genere si può ottenere selezionando un tasso di apprendimento α costante. Prima di dimostrare quanto appena affermato, è necessario ricordare che $R_{t+1} = r_{t+1} + \gamma V_k(s_{t+1})$; fatta questa premessa è possibile riscrivere l'equazione (21) nel modo seguente:

$$V_{k+1}(s_t) = V_k(s_t) + \alpha[R_{t+1} - V_k(s_t)] . \quad 22$$

Andando poi a sostituire ricorsivamente i valori nell'equazione (22) si ottiene la seguente relazione:

$$\begin{aligned} V_k(s_t) &= V_k(s_t) + \alpha[R_{t+1} - V_k(s_t)] & 23 \\ &= \alpha R_{t+1} + (1 - \alpha)V_k(s_t) \\ &= \alpha R_{t+1} + \alpha(1 - \alpha)R_t + (1 - \alpha)V_k(s_{t-1}) \\ &= \alpha R_{t+1} + \alpha(1 - \alpha)R_t + \alpha(1 - \alpha)^2 R_{t-1} + \dots \\ &+ \alpha(1 - \alpha)^{T-1} R_1 + (1 - \alpha)^T V_k(s_0) \\ &= (1 - \alpha)^T V_k(s_0) + \sum_{t=1}^T \alpha(1 - \alpha)^{T-1} R_t . \end{aligned}$$

L'equazione (23) è una media pesata nella quale la somma dei pesi $(1 - \alpha)^T + \sum_{t=1}^T (1 - \alpha)^{T-t} = 1$. Inoltre, dato che $(1 - \alpha) < 1$, il peso che viene attribuito a ciascun rendimento decresce esponenzialmente all'aumentare del numero di ricompense.

Al riguardo, esistono due condizioni di convergenza in ambiente stazionario che devono essere rispettate se si vuole assicurare la convergenza della stima della funzione di valore $V_k(s_t)$ al suo valore ottimo V^* :

$$\sum_{t=1}^{\infty} \alpha_t = \infty \quad e \quad \sum_{t=1}^{\infty} \alpha_t^2 < \infty . \quad 24$$

Con la prima condizione viene assicurata l'elevata numerosità degli step volta a superare le possibili condizioni di instabilità iniziali; la seconda condizione serve invece a garantire step sufficientemente piccoli in modo da convergere al valore ottimo V^* della funzione di valore. Queste condizioni vengono sicuramente rispettate nel caso in cui venga selezionato un tasso di apprendimento non costante nel tempo, per la precisione un tasso di apprendimento che decresce nel tempo, il che ci pone di fronte ad un problema dato che nel caso in esame il tasso di apprendimento deve essere costante a causa della non stazionarietà delle serie dei prezzi. Per superare l'ostacolo è necessario selezionare un parametro *step-size* sufficientemente piccolo: in questo modo la convergenza di $V_k(s_t)$ al suo valore ottimo V^* viene raggiunta in media anche se l'ambiente non è stazionario. Bisogna comunque fare attenzione nella scelta del tasso di apprendimento tenendo conto del fatto che più piccolo questo viene selezionato tanto più lento sarà il processo di convergenza.

2.6 Algoritmo Q-Learning

I metodi TD presentati poco sopra si occupano del processo di valutazione delle stime della funzione di valore V_k dato un qualsiasi criterio comportamentale π . Il passo successivo è quello di ricercare la *policy* ottima π^* sfruttando il processo di iterazione generalizzata del criterio π di cui si è discusso nei paragrafi precedenti. In questa fase possiamo distinguere due approcci diversi: ci sono i metodi definiti *on-policy* e i metodi *off-policy*. L'algoritmo sul quale si basa il presente lavoro è chiamato *Q-Learning* (d'ora in avanti abbreviato in QL) e si

tratta di un algoritmo per differenze temporali appartenente alla categoria dei metodi *off-policy*.

Ci sono tre caratteristiche principali in un algoritmo QL, ovvero:

1. Si tratta di un algoritmo *off-policy* il che significa che approssima direttamente la funzione di valore dell'azione $Q_k(s, a)$ al suo valore ottimo $Q^*(s, a)$ indipendentemente dal criterio comportamentale π . Più precisamente, l'algoritmo fa uso di due criteri durante il processo di perfezionamento: un criterio utilizzato per la stima della funzione di valore e un criterio improntato più sull'esplorazione. L'algoritmo si aggiorna sulla base delle azioni che massimizzano la funzione di valore, indipendentemente che questa venga scelta tra le azioni di uno o dell'altro criterio;
2. È un algoritmo *online*, ossia un algoritmo che aggiorna le stime della funzione di valore ad ogni istante temporale senza che sia necessario raggiungere lo stato terminale;
3. Viene definito algoritmo *di controllo* in quanto il suo scopo è l'ottimizzazione della funzione di valore e per raggiungerlo intraprende ogni azione a sua disposizione.

In termini qualitativi, questo algoritmo calcola una stima di una funzione di valore dell'azione $Q_k(s, a)$ in uno stato s e successivamente esegue iterativamente tutte le azioni disponibili a seguendo un criterio π . In questo modo, l'agente accumula esperienza eseguendo tutte le azioni disponibili e impara quali sono le migliori in termini di rendimento; il passo successivo sarà quello di aggiornare la stima precedente della funzione di valore in un'ottica *one-step*. Seguendo tale procedimento vengono ottimizzate sia la funzione di valore che il criterio comportamentale.

Un algoritmo QL segue una regola di aggiornamento così strutturata:

$$Q_{k+1}(s_t, a_t) = Q_k(s_t, a_t) + \alpha[r_{t+1} + \gamma Q_k(s_{t+1}, a) - Q_k(s_t, a_t)]. \quad 25$$

Come si può notare la struttura della regola di aggiornamento è la medesima illustrata per i metodi TD.

Fino ad ora è stata fatta un'importante assunzione riguardo gli stati del mondo possibili e i valori che questi possono assumere, ossia che fossero un insieme di variabili discrete. La realtà dei mercati finanziari è però ben diversa e lontana da tali assunzioni, dato che sono caratterizzati da stati del mondo (ad esempio, i prezzi dei titoli) continui nel tempo e che possono assumere un numero infinito di possibili valori. Operando in un ambiente con simili caratteristiche non si può sperare di arrivare ad una soluzione di ottimo assoluto: per ovviare a tale problema si andrà alla ricerca non tanto della soluzione ottima ma di una buona approssimazione della stessa utilizzando risorse computazionali limitate.

2.6.1 Approssimazione e ottimizzazione della funzione di valore

Al fine di ottenere un'approssimazione accettabile della stima della funzione di valore è necessario scegliere una struttura di approssimazione adeguata. Due sono i prerequisiti necessari affinché un'approssimazione possa essere definita efficace: il primo è una struttura della funzione di approssimazione sufficientemente articolata in modo da garantire un'approssimazione accettabile della funzione di valore; il secondo è un algoritmo in grado di calibrare in maniera efficace i parametri selezionati dalla funzione di approssimazione strutturata. Utilizzando una funzione di approssimazione parametrica è possibile rappresentare la funzione di valore $Q(s_t, a_t)$ attraverso un vettore di pesi $\theta_t \in \mathbb{R}^n$ composto da un numero fisso di numeri reali:

$$\theta_t = [\theta_1 \quad \theta_2 \quad \theta_3 \quad \dots \quad \theta_n]' . \quad 26$$

Utilizzando tale vettore è possibile scrivere la funzione di approssimazione della funzione di valore:

$$Q(s_t, a_t) = Q(s_t, a_t, \theta_t), \quad 27$$

la quale dipende totalmente dal vettore dei pesi (26) che può variare ad ogni istante temporale, influenzando di conseguenza le stime. Normalmente il numero dei parametri che compongono il vettore dei pesi è inferiore al numero di stati del mondo, di conseguenza una variazione di anche un solo parametro può comportare variazioni nella stima del valore di più stati contemporaneamente. Questa caratteristica rende necessario il calcolo di una misura in grado di stabilire quanta distanza c'è tra la funzione di valore effettiva $Q(s_t, a_t)$ e la stima della sua approssimazione $Q(s_t, a_t, \theta_t)$; questo passaggio è necessario in quanto bisogna capire quale sia la stima dell'approssimazione che più si avvicina al valore effettivo dopo che un parametro del vettore dei pesi è stato modificato. Un indicatore utilizzato per svolgere tale compito è l'*errore quadratico medio* indicato con la sigla MSE^{13} , il quale andrà minimizzato utilizzando la formula seguente:

¹³ Dall'inglese *Mean-Squared Error*.

$$\min_{\boldsymbol{\theta}_t} MSE(\boldsymbol{\theta}_t) = \sum_{s \in \mathcal{S}, a \in \mathcal{A}(s)} [Q^*(s, a) - Q(s_t, a_t, \boldsymbol{\theta}_t)]^2 . \quad 28$$

L'obiettivo consiste nell'individuare un vettore $\boldsymbol{\theta}^*$ tale per cui $MSE(\boldsymbol{\theta}^*) \leq MSE(\boldsymbol{\theta})$ per ogni possibile settaggio di $\boldsymbol{\theta}$. In sostanza, quello che si sta cercando è un punto di *minimo globale*. Può accadere però che un punto di minimo globale non sia sempre raggiungibile se la funzione di approssimazione non è lineare: durante il processo di ottimizzazione non è raro che il processo si interrompa dopo aver individuato un punto che all'apparenza sembra essere un punto di minimo globale ma che in realtà si rivela essere un punto di *minimo locale*. In un punto di minimo locale si avrà comunque $MSE(\boldsymbol{\theta}^*) \leq MSE(\boldsymbol{\theta})$, tuttavia questo accade per settaggi di $\boldsymbol{\theta}$ nell'intorno di $\boldsymbol{\theta}^*$ e non per ogni settaggio di $\boldsymbol{\theta}$ possibile.

L'unica soluzione accettabile per superare il problema è reiterare più volte il processo partendo da diverse inizializzazioni dei parametri della funzione di approssimazione in modo da trovare un insieme di ottimi locali tra cui cercare poi il punto che può essere identificato anche come un ottimo globale. Esistono tuttavia dei metodi in grado di generare soluzioni ottime sotto le opportune condizioni, uno di questi metodi è il *metodo della discesa del gradiente* e verrà illustrato nel prossimo paragrafo.

2.6.2 Il metodo della discesa del gradiente

Per risolvere il problema di minimizzazione della funzione di approssimazione viene ampiamente utilizzato il *metodo della discesa del gradiente*, il quale viene applicato alla struttura di approssimazione della funzione di valore $Q(s_t, a_t, \boldsymbol{\theta}_t)$ al fine di trovare il vettore $\boldsymbol{\theta}^*$ che minimizza la distanza tra l'effettiva funzione di valore e la stima della sua approssimazione.

La strategia che utilizza questo metodo consiste nel minimizzare il MSE tra l'incognita funzione di valore e la stima della sua approssimazione, aggiustando di volta in volta il vettore dei parametri nella direzione verso cui l'errore stesso si riduce. La direzione verso cui procedere viene fornita dal *gradiente* $\nabla_{\boldsymbol{\theta}_t} f(\boldsymbol{\theta})$, cioè il vettore delle derivate parziali delle

componenti di $\boldsymbol{\theta}_t$. La regola di aggiornamento utilizzata per l'aggiustamento delle componenti del vettore dei parametri è la seguente:

$$\begin{aligned}\boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t - \frac{1}{2}\alpha\nabla_{\boldsymbol{\theta}_t}[Q^*(s, a) - Q(s_t, a_t, \boldsymbol{\theta}_t)]^2 \\ &= \boldsymbol{\theta}_t - \alpha[Q^*(s, a) - Q(s_t, a_t, \boldsymbol{\theta}_t)]\nabla_{\boldsymbol{\theta}_t}Q(s_t, a_t, \boldsymbol{\theta}_t).\end{aligned}\tag{29}$$

Tale metodo procede a piccoli passi e non in un unico step verso la direzione indicata dall'opposto del gradiente: questo accade poiché una soluzione esatta sarà generalmente impossibile da trovare, come più volte ripetuto nei paragrafi precedenti. Quello che invece si cerca di fare con questo metodo è trovare l'approssimazione della funzione di valore tale per cui gli errori (cioè la differenza tra la funzione di valore effettiva e la sua approssimazione) siano bilanciati in tutti gli stati possibili.

Il metodo della discesa del gradiente garantisce la convergenza ad un punto di ottimo locale nel caso in cui le condizioni di convergenza (24) siano rispettate in un ambiente stazionario. In questo caso, essendo l'ambiente di riferimento non stazionario la convergenza viene raggiunta in media reiterando più volte il processo e selezionando un valore del tasso di apprendimento sufficientemente piccolo.

Dato che la forma della funzione di valore ottima $Q^*(s, a)$ è sconosciuta, i punti di minimo che il processo appena descritto individua potrebbero essere punti di minimo globale o punti di minimo locale. Se la funzione di valore ottima non rispetta le condizioni di convessità¹⁴ – il che accade molto spesso – il processo di minimizzazione del $MSE(\boldsymbol{\theta})$ verrà inevitabilmente attirato verso uno dei numerosi punti di minimo locale, causando l'interruzione del processo. Questo problema è uno dei principali punti di debolezza del metodo della discesa del gradiente, a cui è possibile far capo eseguendo più e più volte il processo di ottimizzazione selezionando arbitrariamente di volta in volta un punto di partenza diverso. Così facendo il processo porta ad ottenere un nutrito insieme di punti di minimo locale tra i quali è probabile che sia incluso anche il punto di minimo globale a cui si è maggiormente interessati. Ovviamente non è garantito che in questo insieme sia incluso

¹⁴ Una funzione f definita su un intervallo I si definisce convessa se $\forall x_1, x_2 \in I$ con $x_1 < x_2$ e $\forall t \in [0,1]$ risulta valida la disuguaglianza $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$. Se la medesima disuguaglianza risulta essere valida in senso stretto allora si parla di funzione convessa in senso stretto.

L'ottimo globale, tuttavia procedere in questo modo nella maggior parte dei casi risulta essere la miglior scelta a disposizione.

Il fatto che la funzione di valore ottima $Q^*(s, a)$ sia sconosciuta implica il fatto che in tutte le metodologie appena esposte sia necessario utilizzare una sua approssimazione:

$$q_t = r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \boldsymbol{\theta}_t) \quad 30$$

Sostituendo questa espressione nell'equazione (29) si ottiene:

$$\begin{aligned} \boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \boldsymbol{\theta}_t) - Q(s_t, a_t, \boldsymbol{\theta}_t) \right] \nabla_{\boldsymbol{\theta}_t} Q(s_t, a_t, \boldsymbol{\theta}_t) \quad 31 \\ &= \boldsymbol{\theta}_t + \alpha \delta_t \nabla_{\boldsymbol{\theta}_t} Q(s_t, a_t, \boldsymbol{\theta}_t) \end{aligned}$$

dove δ_t è l'errore TD visto precedentemente. Se $\mathbb{E}(q_t) = Q^*(s, a) \forall t$, e quindi se q_t è una stima corretta di $Q^*(s, a)$, allora $\boldsymbol{\theta}_t$ convergerà ad un punto di minimo locale a patto che le condizioni di convergenza per il parametro *step-size* α vengano rispettate.

2.6.3 Approssimazione lineare della funzione di valore

Allo scopo di trovare un'approssimazione della funzione di valore $Q(s_t, a_t, \boldsymbol{\theta}_t)$, tale per cui il parametro dei vettori $\boldsymbol{\theta}_t$ converga al suo ottimo $\boldsymbol{\theta}^*$, molto spesso si utilizza una approssimazione di tipo lineare del vettore dei pesi. L'idea è che per ogni stato s_t esista un vettore composto da tanti elementi quanti sono quelli del vettore dei pesi $\boldsymbol{\theta}_t$. Tale vettore, definito $\boldsymbol{\phi}_{s_t}$, moltiplicato per il vettore dei pesi $\boldsymbol{\theta}_t$ permette di ottenere l'approssimazione lineare della funzione di valore:

$$Q(s_t, a_t, \boldsymbol{\theta}_t) = \boldsymbol{\theta}_t' \boldsymbol{\phi}_{s_t} = \sum_{n=1}^N \boldsymbol{\theta}_t(n) \boldsymbol{\phi}_{s_t}(n). \quad 32$$

La funzione appena presentata viene chiamata funzione di *squashing* dello stato n -esimo. Tale funzione deve rispettare necessariamente una proprietà fondamentale in questo contesto,

ossia deve essere una funzione *monotona crescente*¹⁵ affinché l'ordine degli elementi trasformati mediante l'approssimazione rimanga inalterato.

¹⁵ Una funzione f si definisce monotona crescente se per ogni x_1, x_2 con $x_1 < x_2$ risulta che $f(x_1) < f(x_2)$.

Struttura dell'algoritmo

In questo capitolo viene presentata la struttura dell'algoritmo Q-Learning utilizzato nel presente elaborato. Tale algoritmo è stato sviluppato utilizzando il software MatLab ed è stato applicato a cinque diverse serie di prezzi di titoli reali. Per presentare l'operatività dell'algoritmo è necessario analizzare singolarmente ciascun elemento che lo caratterizza, di seguito verranno quindi presentati i vari elementi che lo compongono e le relative configurazioni. Per alcuni parametri sono stati selezionati diversi valori da utilizzare: in questo modo è stato possibile testare l'operatività del sistema di trading automatico sotto diverse condizioni. Ciascuna combinazione di questi parametri è stata testata per un numero elevato di volte, precisamente fissato pari a $K = 500$. Come già anticipato nei capitoli precedenti, una elevata numerosità delle iterazioni è necessaria affinché venga individuato il vettore dei pesi ottimo θ^* : il vettore θ viene inizializzato arbitrariamente ad ogni iterazione per permettere all'algoritmo di trovare la soluzione ottima, e quindi è fondamentale che il processo venga ripetuto più e più volte per permettere all'algoritmo di "spaziare" durante la ricerca della miglior soluzione possibile.

Verranno infine presentate brevemente le serie storiche dei prezzi alle quali è stato applicato l'algoritmo in modo da fornire un'idea del contesto in cui ha operato il sistema di trading automatico.

3.1 I descrittori dello stato

Uno degli aspetti che si intende investigare nel presente elaborato è testare l'operatività del sistema di trading utilizzando due diversi insiemi di descrittori di stato. Il primo insieme di descrittori è costituito dai rendimenti logaritmici giornalieri mentre il secondo è costituito da alcuni indicatori di Analisi Tecnica, i quali verranno presentati nelle pagine successive. La necessità di una duplice struttura come questa nasce dall'esigenza di confrontare l'utilizzo di un vettore di stato costituito da elementi semplici – i rendimenti logaritmici – con l'utilizzo

di un vettore costituito da descrittori un po' più raffinati – gli indicatori AT – allo scopo di verificare se questi ultimi possono portare ad ottenere una redditività ed una operatività maggiori.

Come già anticipato, nella prima parte dell'applicazione sono state utilizzate informazioni semplici per descrivere l'ambiente: tali informazioni sono i rendimenti logaritmici giornalieri. Si è scelto quindi di utilizzare come input gli ultimi $N - 1$ rendimenti logaritmici giornalieri di ciascun titolo, il rendimento logaritmico corrente e l'ultima azione che è stata selezionata dall'algoritmo. Ricordando brevemente che il rendimento logaritmico di un titolo si calcola secondo la formula seguente

$$e_t = \log\left(\frac{p_t}{p_{t-1}}\right) \quad 33$$

dove con p_t si indica il prezzo di un titolo all'istante t , allora all'istante t lo stato s_t sarà descritto dal seguente vettore:

$$[e_{t-N+1} \ e_{t-N+2} \ \dots \ e_t \ a_{t-1}]. \quad 34$$

Al fine di testare la risposta del sistema di trading a questi input sono stati settati due diversi valori di N , rispettivamente $N = 2$ e $N = 5$.

La seconda parte dell'applicazione ha uno scopo molto preciso, ossia verificare se un input un po' più raffinato sia in grado di portare a risultati migliori di quelli che si sono ottenuti utilizzando i soli rendimenti logaritmici. A tale scopo entra in gioco l'*analisi tecnica*, cioè una serie di tecniche che permette di individuare un trend in corso o, alle volte, di prevederne l'andamento in modo tale da orientare in maniera ottimale le proprie scelte di trading. L'analisi tecnica si avvale di strumenti quali la lettura dei grafici e il calcolo di indicatori e oscillatori al fine di individuare dei *pattern* o dei trend che possano guidare le scelte in ambito finanziario. L'ipotesi è che l'utilizzo di indicatori e oscillatori come input possa portare a risultati migliori, in termini di rendimento, di quelli ottenuti utilizzando solamente i rendimenti logaritmici per descrivere lo stato. Di fatto, in presenza di titoli con un andamento dei prezzi molto volatile e caratterizzato da continui *break* strutturali l'algoritmo potrebbe non essere in grado di adattarsi velocemente ai repentini cambi di prezzo; conseguentemente il sistema di trading potrebbe non essere in grado di individuare un comportamento ottimale da seguire poiché dovrebbe frequentemente "eliminare" *pattern*

comportamentali da ultimo appresi per cominciare l'apprendimento di *pattern* completamente diversi. Tutto questo potrebbe portare alla selezione di azioni non propriamente conformi allo schema comportamentale richiesto dall'ambiente finanziario in quel preciso momento.

Premesso ciò, sono stati selezionati alcuni indicatori da utilizzare come input in alternativa ai rendimenti logaritmici. Nel dettaglio, si è scelto di utilizzare due medie mobili, una corta e una lunga, e l'oscillatore *Relative Strength Index* (d'ora in poi abbreviato con *RSI*).

In generale, la media mobile è un indicatore molto semplice ma che può rappresentare un valido aiuto per le scelte di trading se utilizzato nel modo corretto. Una media mobile si calcola utilizzando le osservazioni di una *rolling windows* di ampiezza prefissata. Il risultato è un indicatore che segue l'andamento del mercato, anche se con un certo ritardo che dipende dall'ampiezza della *rolling window*, ed è in grado di segnalare l'inizio di un nuovo trend o la fine di uno in corso. La formula per il calcolo della media mobile è quindi molto semplice, ossia:

$$MA(L) = \frac{\sum_{t=1}^L e_t}{L} \quad 35$$

dove L rappresenta l'ampiezza della *rolling window* ed è quindi il numero di rendimenti da includere nel calcolo.

Come già detto sono state scelte due medie mobili: una corta $MA(5)$ calcolata su una finestra di 5 giorni, una lunga $MA(21)$ calcolata su una finestra di 21 giorni. La scelta di tali finestre temporali non è casuale: in ambito finanziario, queste due grandezze rappresentano rispettivamente una settimana e un mese di rendimenti. Il motivo per cui è necessario combinare una media mobile corta con una media mobile più lunga è presto detto: una media mobile calcolata su un breve lasso temporale traccia abbastanza bene l'andamento dei rendimenti ma produce al contempo molti falsi segnali che potrebbero depistare il sistema di trading. Questo accade perché la quantità di rendimenti utilizzati per il calcolo dell'indicatore è ridotta, se una serie è caratterizzata da una volatilità dei prezzi notevole allora una media mobile corta avrà un andamento altrettanto volatile, seppur ritardato a causa della natura dell'indicatore. Affiancandola con una media mobile calcolata su un lasso temporale più lungo viene effettuato uno *smoothing* maggiore che elimina eventuali falsi segnali, portando ad una riduzione delle transazioni potenzialmente inutili con relativo risparmio in termini di costi di transazione.

L'altro elemento scelto come input è l'oscillatore *RSI*, uno degli oscillatori più famosi nel mondo dell'analisi tecnica. L'*RSI* viene calcolato nel modo seguente:

$$RSI = 100 - \frac{100}{1 + RS} \quad 36$$

dove RS indica la *Relative strenght*, la quale viene calcolata attraverso la formula seguente:

$$RS = \frac{\text{Media degli } L \text{ giorni con chiusura positiva}}{\text{Media degli } L \text{ giorni con chiusura negativa}} \quad 37$$

Per il presente elaborato si è scelto di utilizzare un periodo di 14 giorni per il calcolo di tale oscillatore, dato che questo è il numero di giorni che viene normalmente utilizzato in gran parte della letteratura in materia [8] [9].

In questa seconda applicazione che utilizza gli indicatori di analisi tecnica verranno quindi utilizzati un numero di input $N = 3$, e cioè la media mobile corta $MA(5)$, la media mobile lunga $MA(21)$, l'indice di forza relativa $RSI(14)$.

3.2 La struttura di selezione dell'azione

Ad ogni step temporale t l'algoritmo testa tutte le azioni disponibili per decidere quale intraprendere effettivamente. Tale scelta avviene sulla base della ricompensa r_{t+1} che ciascuna azione garantisce, portando l'algoritmo a selezionare l'azione che gli dovrebbe permettere di ricevere la ricompensa maggiore. Le azioni disponibili ad ogni istante temporale sono rappresentate dal seguente sistema [10]:

$$a_t = \begin{cases} -1 \\ 0 \\ 1 \end{cases} \quad 38$$

dove con $a_t = -1$ si intende “vendi o posizione short sul mercato”, con $a_t = 0$ si intende “star fuori dal mercato” – il che implica la chiusura di qualsiasi precedente posizione eventualmente aperta –, con $a_t = 1$ si intende “compra o posizione long sul mercato”.

Una volta che l'agente avrà selezionato l'azione riceverà all'istante temporale $t + 1$ una ricompensa r_{t+1} che gli permetterà di fare una valutazione della scelta appena effettuata in termini di performance. Tuttavia, come visto nei capitoli precedenti non è sufficiente che l'agente scelga sempre e soltanto l'azione che massimizza la ricompensa per ottenere sul lungo periodo la performance migliore. È assolutamente necessario che l'algoritmo esplori anche nuovi scenari selezionando le azioni secondo un criterio diverso da quello *greedy*, in questo modo è in grado di esplorare nuovi stati che potenzialmente potrebbero condurre a performance migliori di quelli fino ad ora conosciuti. In sostanza, l'algoritmo seleziona l'azione seguendo il sistema seguente:

$$a = \begin{cases} \mathop{\text{arg max}}_a Q(s_t, a_t, \theta_t) & \text{con probabilità } 1 - \varepsilon \\ a_t & \text{con probabilità } \varepsilon \end{cases} \quad 39$$

Il parametro ε determina la frequenza con cui l'algoritmo seleziona un'azione esplorativa, pertanto il valore di ε rappresenta la probabilità che ad ogni step temporale t venga selezionata un'azione di tipo esplorativo. Le azioni esplorative si distribuiscono uniformemente; in questo modo ciascuna azione esplorativa ha la medesima probabilità di tutte le altre di essere selezionata qualora l'algoritmo scelga di intraprendere il percorso esplorativo. Al fine di testare se sia più utile un algoritmo più o meno esplorativo sono stati selezionati due diversi valori del parametro ε , ovvero $\varepsilon \in \{5\%, 15\%\}$. Aumentando il valore di ε l'algoritmo selezionerà più spesso un'azione esplorativa da intraprendere; inoltre, dato che la selezione di una di queste è del tutto casuale potrebbe accadere che le performance del sistema di trading peggiorino. Questo procedimento è tuttavia necessario all'algoritmo per imparare quali nuove azioni possono migliorarlo e quali invece è meglio non selezionare più.

A fini computazionali, ciascuna delle azioni del sistema (38) non viene rappresentata attraverso un singolo valore, bensì attraverso un vettore:

- L'azione -1 , cioè l'azione "*vendi o posizione short sul mercato*" viene rappresentata con il vettore

$$\mathbf{a}_{-1} = [1 \quad 0 \quad 0]; \quad 40$$

- l'azione 0, cioè l'azione “*star fuori dal mercato*” viene rappresentata con il vettore

$$\mathbf{a}_0 = [0 \quad 1 \quad 0]; \quad 41$$

- l'azione 1, cioè l'azione “*compra o posizione long sul mercato*” viene rappresentata con il vettore

$$\mathbf{a}_1 = [0 \quad 0 \quad 1]. \quad 42$$

Il motivo per cui è necessaria questa struttura è presto detto: durante il processo di stima della funzione di valore, alcune azioni potrebbero penalizzate rispetto ad altre se si utilizza il sistema (38) in cui l'azione \mathbf{a}_t assume un valore diverso a seconda del caso selezionato. Dato che il vettore dei pesi $\boldsymbol{\theta}_t$ e il vettore contenente i descrittori dello stato \mathbf{s}_t sono identici per ciascuna diversa azione \mathbf{a}_t , il valore effettivo della funzione di valore viene determinato grazie all'azione che viene selezionata. Trattandosi però di un processo di massimizzazione, è naturale che il processo selezioni con molta più frequenza l'azione $\mathbf{a}_t = 1$ a discapito dell'azione $\mathbf{a}_t = -1$: la funzione di valore calcolata utilizzando il valore della prima azione risulterebbe essere maggiore rispetto alla funzione di valore calcolata utilizzando la seconda, causando uno sbilanciamento del sistema di trading verso la posizione di acquisto. Invece, utilizzando i vettori (40), (41) e (42) le azioni sono differenziate; con questa impostazione le funzioni di valore possono essere rappresentate nel modo seguente:

- funzione di valore associata all'azione “*vendi*”:

$$Q(s_t, \mathbf{a}_{-1}, \boldsymbol{\theta}_t) = \boldsymbol{\theta}_t \boldsymbol{\phi}_{s_t, \mathbf{a}_{-1}} = \theta(0) + \sum_{n=1}^{N-1} \theta_t(n) + \phi_{s_t}(n) + \theta(N) \phi_t \mathbf{a}_{-1}; \quad 43$$

- funzione di valore associata all'azione “*star fuori dal mercato*”:

$$Q(s_t, \mathbf{a}_0, \boldsymbol{\theta}_t) = \boldsymbol{\theta}_t \boldsymbol{\phi}_{s_t, \mathbf{a}_0} = \theta(0) + \sum_{n=1}^{N-1} \theta_t(n) + \phi_{s_t}(n) + \theta(N) \phi_t \mathbf{a}_0; \quad 44$$

- funzione di valore associata all'azione "compra":

$$Q(s_t, \mathbf{a}_1, \boldsymbol{\theta}_t) = \boldsymbol{\theta}_t \boldsymbol{\phi}_{s_t, \mathbf{a}_1} = \theta(0) + \sum_{n=1}^{N-1} \theta_t(n) + \phi_{s_t}(n) + \theta(N) \phi_t \mathbf{a}_1 \quad 45$$

Con questo approccio ciascun vettore (39), (40) e (41) ha il proprio parametro $\boldsymbol{\theta}$, il quale assume diversi valori. Si otterranno pertanto tre diverse funzioni di valore $Q(s, \mathbf{a}, \boldsymbol{\theta})$, una per ciascuna delle tre azioni selezionabili, e tra queste verrà selezionata quella con il valore maggiore. Il valore della funzione $Q(s, \mathbf{a}, \boldsymbol{\theta})$, quindi, non dipende più dal valore che viene attribuito ad ogni azione secondo il sistema (38), ogni funzione di valore dipenderà esclusivamente dal valore della variabile $\boldsymbol{\theta}$. Così facendo si evita che il sistema di trading sia sbilanciato e indotto a selezionare più spesso l'azione "compra" a discapito delle altre due.

3.3 La funzione di ricompensa

Come si è visto nei capitoli precedenti, la funzione di ricompensa fornisce un'indicazione immediata sulla performance del sistema di trading in seguito alla selezione dell'azione. Nel presente elaborato sono state selezionate due diverse funzioni di ricompensa al fine di testare la risposta dell'algoritmo in base al tipo di ricompensa che viene utilizzata.

La prima funzione di ricompensa è la più comunemente utilizzata nella letteratura sul tema, ossia lo Sharpe Ratio (SR) [10] [11]. Si tratta di una misura *risk-adjusted*, tra tutte forse la più famosa, che si ottiene dal seguente rapporto:

$$SR_t = \frac{\mathbb{E}_L(e_t)}{\sqrt{\text{Var}_L(e_t)}} \in \mathbb{R} \quad 46$$

dove SR_t è lo Sharpe ratio calcolato all'istante temporale t , $\mathbb{E}_L(e_t)$ e $\text{Var}_L(e_t)$ sono rispettivamente la media e la varianza campionarie dei rendimenti ottenuti durante gli ultimi L giorni di trading. In sostanza, lo Sharpe ratio misura la quantità di rendimento attribuibile ad un'attività rischiosa per unità di rischio, motivo per cui viene definita una misura *risk-adjusted*.

Una delle principali critiche all'utilizzo dello Sharpe ratio è la presenza al denominatore della deviazione standard dei rendimenti del periodo considerato; utilizzando

tale misura vengono pesati in egual modo i rendimenti al di sotto della media e i rendimenti che invece sono al di sopra di essa. Gli scostamenti negativi dalla media sono quelli che effettivamente penalizzano la performance del sistema di trading, mentre gli scostamenti positivi possono essere invece utili per migliorarla. Un modo per tenere conto di questo aspetto è l'utilizzo di una diversa misura *risk-adjusted* chiamata Sortino Ratio (*SOR*), la quale si calcola attraverso il seguente rapporto:

$$SOR_t = \frac{\mathbb{E}_L(e_t)}{\sqrt{\text{Var}_L[\min(e_t - \bar{e}_t), 0]}} \in \mathbb{R}. \quad 47$$

A differenza dello Sharpe ratio, il Sortino ratio ha al denominatore una misura chiamata *downside risk*, ossia la deviazione standard campionaria che però tiene conto esclusivamente degli scostamenti negativi dalla media lungo il periodo L considerato. Utilizzando tale misura la performance del sistema di trading viene penalizzata solamente in caso di rendimenti che siano inferiori alla media, il che la rende perfetta per l'utilizzo in ambito finanziario.

Così come per gli elementi dell'algoritmo già presentati, è utile avere un confronto utilizzando diversi settaggi anche per quanto riguarda la funzione di ricompensa. Per il calcolo sia dello Sharpe che del Sortino ratio verranno infatti utilizzati due diversi valori di L , ossia $L = 5$ e $L = 21$, che come già accennato precedentemente rappresentano un orizzonte temporale di una settimana e di un mese in termini di giorni di contrattazione sul mercato finanziario.

3.4 I costi di transazione

Per rendere l'applicazione il più realistica possibile è stata prevista l'applicazione di costi di transazione. I costi di transazione sono un elemento molto importante di cui bisogna necessariamente tenere conto dato che, se vengono eseguite troppe transazioni da parte dell'agente artificiale, questi possono erodere considerevolmente il profitto netto. Tenere conto di questo aspetto è estremamente importante per capire quali strategie risultano essere profittevoli e quali invece sono da considerarsi cattive; il sistema di trading dovrà quindi essere in grado di valutare attentamente quali transazioni conviene eseguire e quali invece no al fine di preservare nel tempo il capitale ottenuto.

I costi di transazione vengono applicati quando l'agente cambia la propria posizione su un determinato titolo; in parole povere vengono applicati se due azioni consecutive risultano essere diverse e quindi se $a_t \neq a_{t-1}$. Nel presente lavoro i costi di transazione vengono considerati in termini percentuali e sono applicati sia che la posizione venga aperta sia che venga chiusa, pertanto i costi di transazione δ si ottengono dimezzando i costi di transazione totali che l'algoritmo calcola:

$$\delta = \frac{\text{costi di transazione totali}}{2}. \quad 48$$

Nei test eseguiti, la percentuale dei costi di transazione è stata fissata pari allo 0,12%, percentuale basata sulla media dei costi di transazione applicati da diverse compagnie di brokeraggio italiane.

Il rendimento al netto dei costi di transazione all'istante temporale $t + 1$ – se l'azione che lo ha generato è diversa da quella selezionata nell'istante temporale precedente – può essere definito nel modo seguente:

$$r_{t+1}^{net} = a_t e_{t+1} - \delta |a_{t+1} - a_t| \quad 49$$

dove $a_t e_{t+1}$ rappresenta il rendimento lordo all'istante temporale $t + 1$. L'utilizzo del modulo $|a_{t+1} - a_t|$ fa sì che la percentuale dei costi di transazione si applichi al rendimento solo se le due azioni consecutive sono diverse tra loro. Al termine di ogni step viene calcolata l'*equity line* dello step temporale $t + 1$ attraverso la formula seguente:

$$E_{t+1} = E_t (1 + r_{t+1}^{net}) \quad 50$$

dove E_{t+1} e E_t rappresentano le *equity lines* dell'istante temporale $t + 1$ e t , cioè i grafici che rappresentano l'andamento della performance del sistema di trading.

3.5 La funzione di *squashing*

La funzione di *squashing* si utilizza per trasformare i descrittori di stato, ossia gli input contenuti nel vettore \mathbf{s} . Lo scopo di tale trasformazione è di aumentare la sensibilità del

processo di valutazione della funzione di valore e di selezione dell'azione *greedy*. Si può pensare alla funzione di *squashing* come ad una sorta di *clusterizzazione* di dati con valori molto vicini tra loro. L'utilizzo di una funzione di *squashing* è importante data la peculiarità dei mercati finanziari: i rendimenti finanziari sono caratterizzati da valori estremi ravvicinati e si concentrano per lo più attorno allo 0; rendere più sensibile l'algoritmo con la funzione di *squashing* significa differenziare le realizzazioni che presentano valori molto vicini tra di loro. In questo lavoro si è scelto di utilizzare una funzione logistica definita nel modo seguente [11]:

$$\phi(x) = \frac{a}{1 + be^{-cx}} - d \quad 51$$

dove $a = 2$, $b = 1$, $c = 10^{15}$ e $d = -1$. La scelta di un valore di c così elevato è dettata dalla necessità di distanziare il più possibile valori molto vicini tra di loro.

A tale scopo può essere utile una rappresentazione grafica di tale funzione utilizzando diversi valori per il parametro c . Come si può notare dalla Figura 3.1, all'aumentare del valore del parametro c la funzione assume una forma "a gradino", indice del fatto che valori che prima erano molto vicini tra loro dopo la trasformazione risultano essere più differenziati.

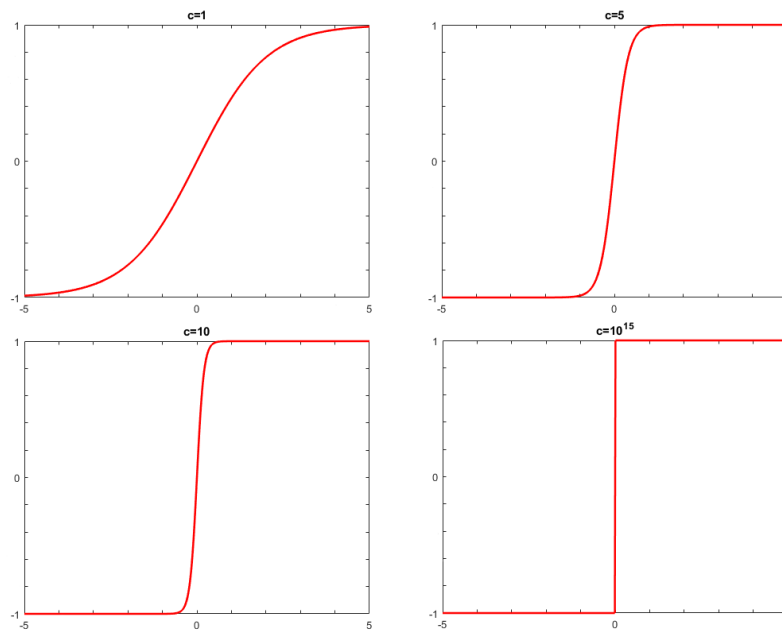


Figura 3.1: Confronto tra funzioni logistiche a seconda del valore di c .

3.6 Presentazione delle serie storiche

Per l'applicazione dell'algoritmo sono stati selezionati cinque titoli quotati sul mercato italiano: Davide Campari-Milano (CPR.MI), Enel S.p.A. (ENEL.MI), Fiat Chrysler Automobiles (FCA.MI), Intesa Sanpaolo S.p.A. (ISP.MI) e Telecom Italia S.p.A. (TIT.MI). Tali titoli sono stati scelti per diversi motivi, il principale dei quali è che sono tra i titoli a maggior capitalizzazione sul mercato italiano. Un altro motivo che ha portato alla scelta di questi titoli è l'andamento delle serie dei prezzi; alcune serie hanno andamenti dei prezzi particolari, lo scopo dell'elaborato è verificare l'efficacia degli indicatori di Analisi Tecnica utilizzati come descrittori dello stato e applicare i diversi sistemi di trading a serie storiche dei prezzi con un andamento particolare può aiutare a capire in quali casi questi sono effettivamente efficaci e in quali invece non lo sono.

Tutte le serie storiche sono state scaricate dal sito Yahoo Finance e ricoprono un periodo di 15 anni di contrattazione, rispettivamente dal 3 Gennaio 2005 al 30 Dicembre 2020, per un totale di 4066 osservazioni dei prezzi di chiusura giornalieri. Di seguito è possibile osservare i grafici relativi ai prezzi e ai rendimenti giornalieri di ciascun titolo, seguiti da una tabella riassuntiva delle principali statistiche descrittive di ciascun titolo.

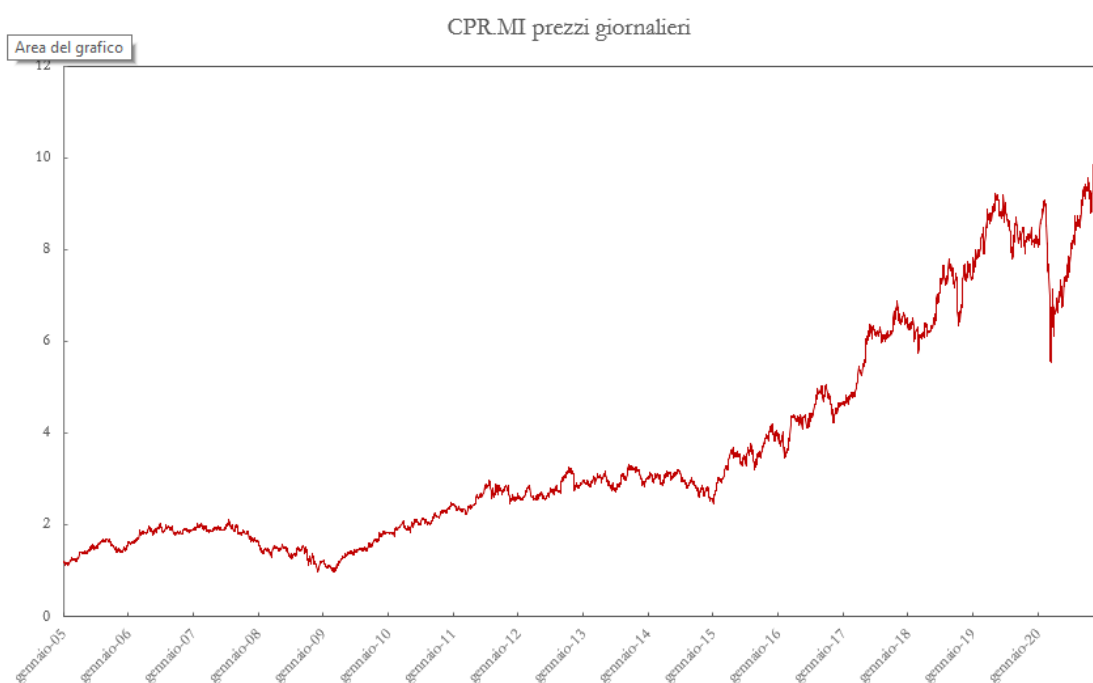


Figura 3.2: Grafico dei prezzi giornalieri di Davide Campari – Milano.

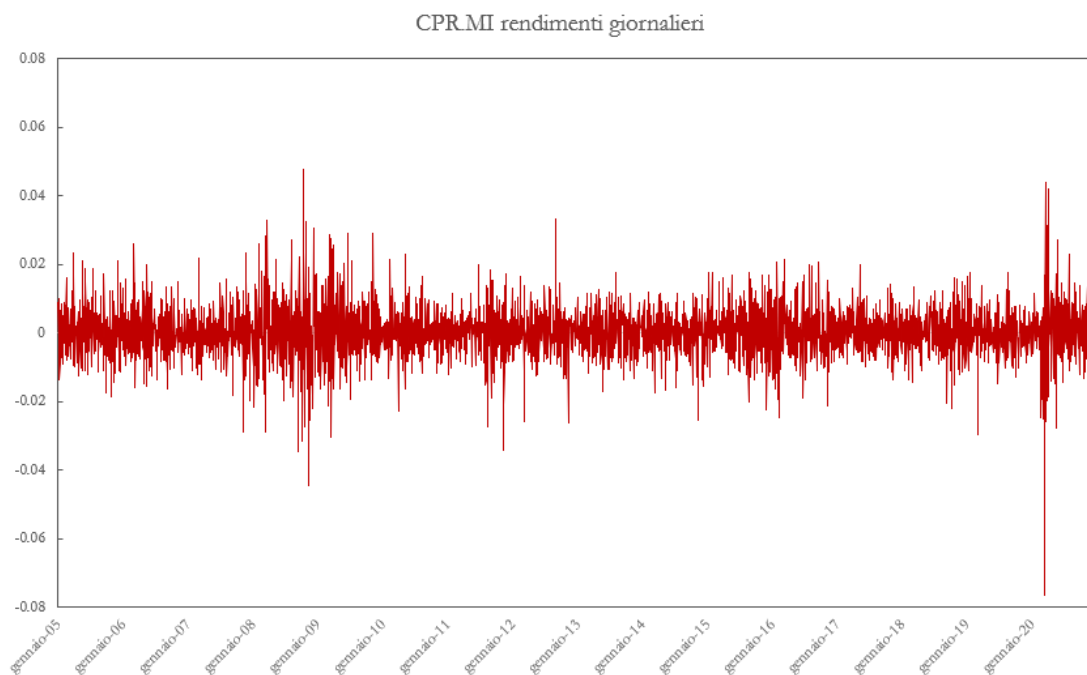


Figura 3.3: Grafico dei rendimenti giornalieri di Davide Campari – Milano.

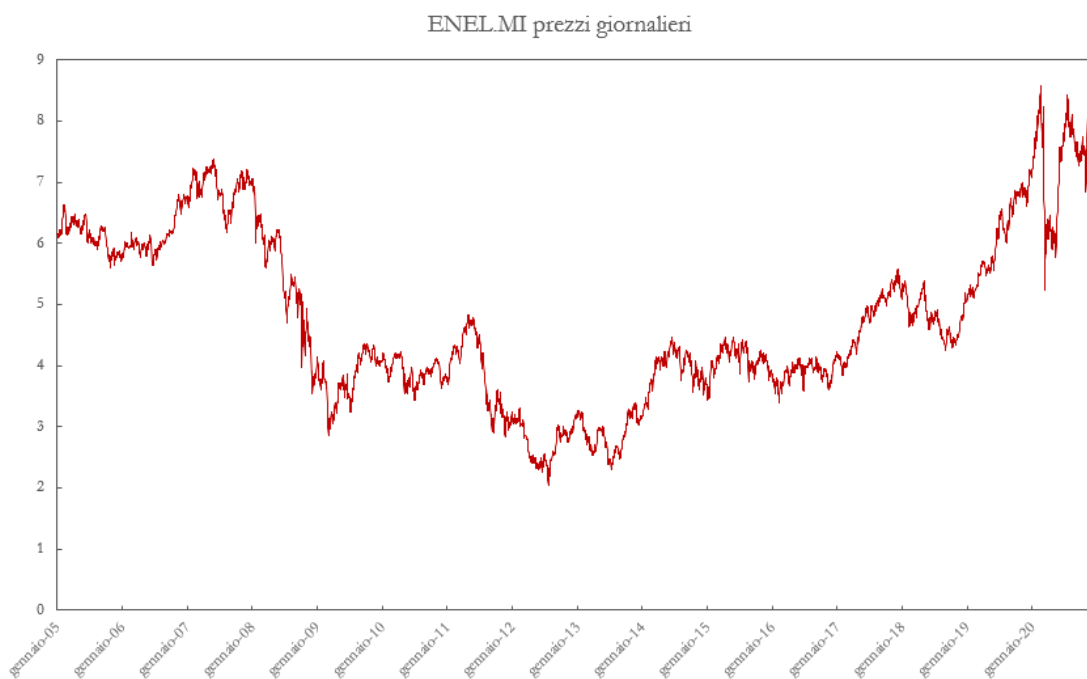


Figura 3.4: Grafico dei prezzi giornalieri di Enel S.p.A

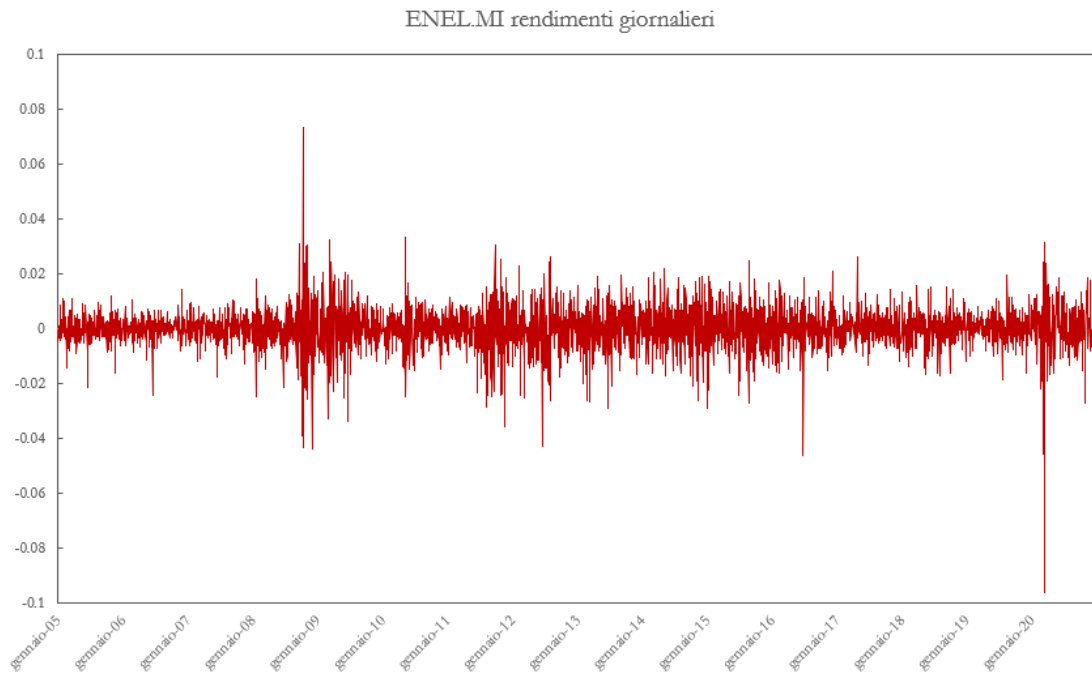


Figura 3.5: Grafico dei rendimenti giornalieri di Enel S.p.A.

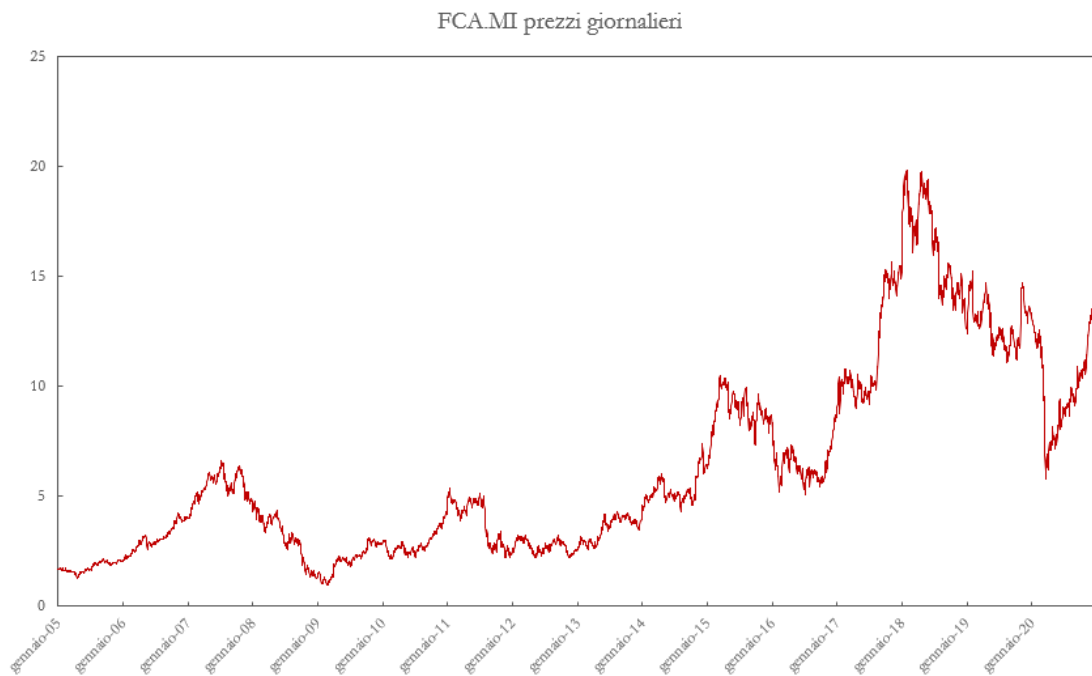


Figura 3.6: Grafico dei prezzi giornalieri di Fiat Chrysler Automobiles.

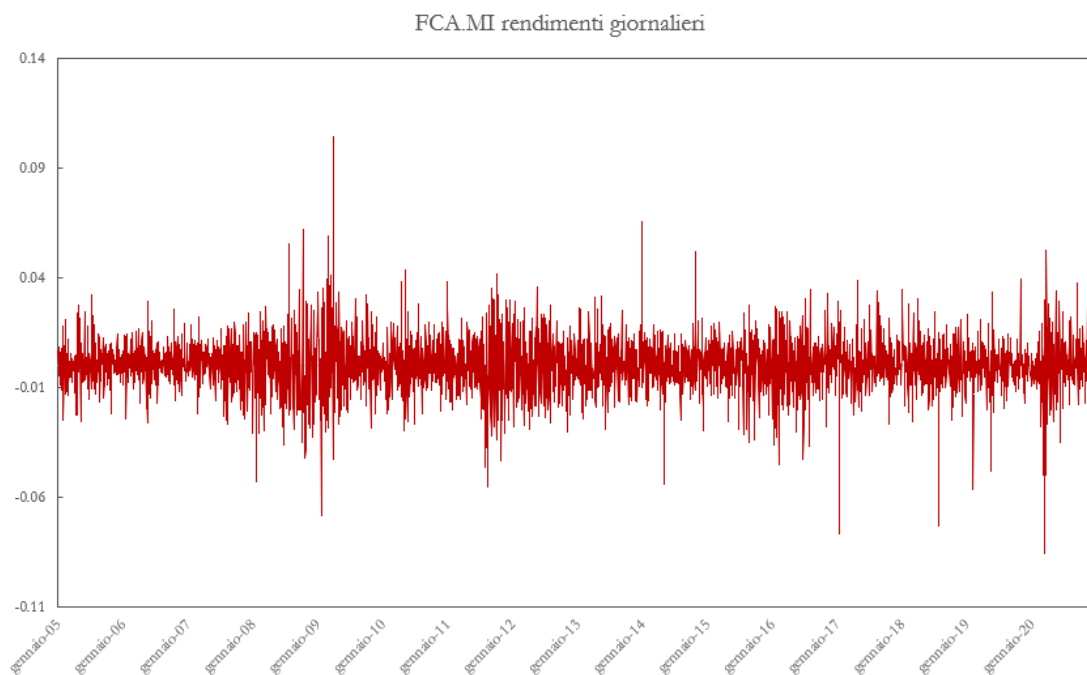


Figura 3.7: Grafico dei rendimenti giornalieri di Fiat Chrysler Automobiles NV

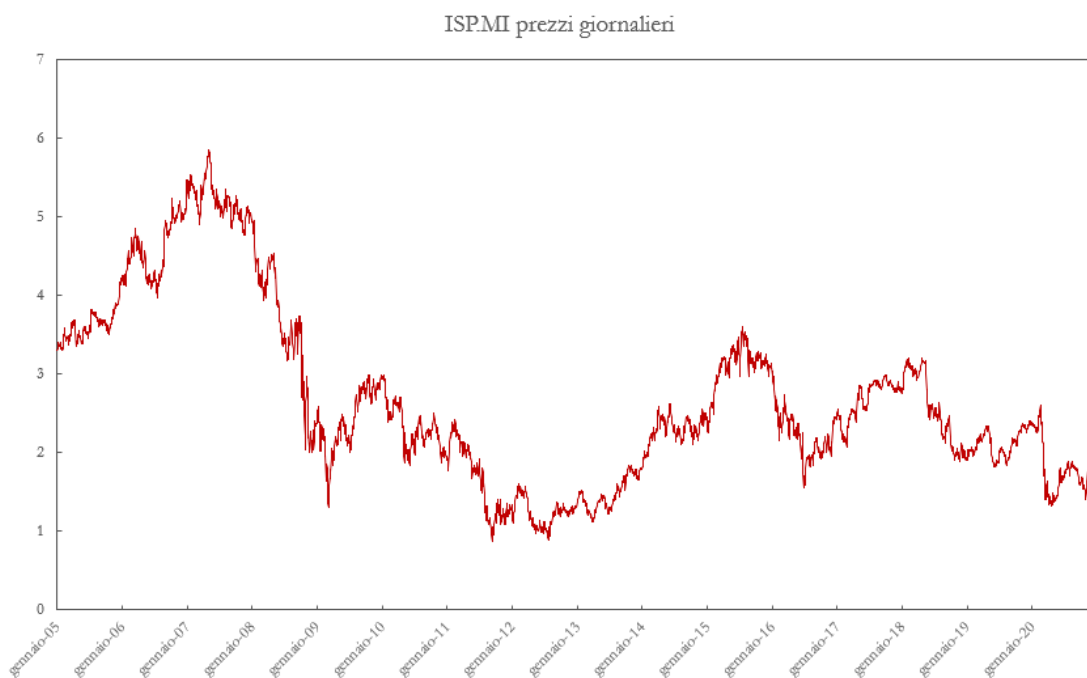


Figura 3.8: Grafico dei prezzi giornalieri di Intesa Sanpaolo S.p.A.

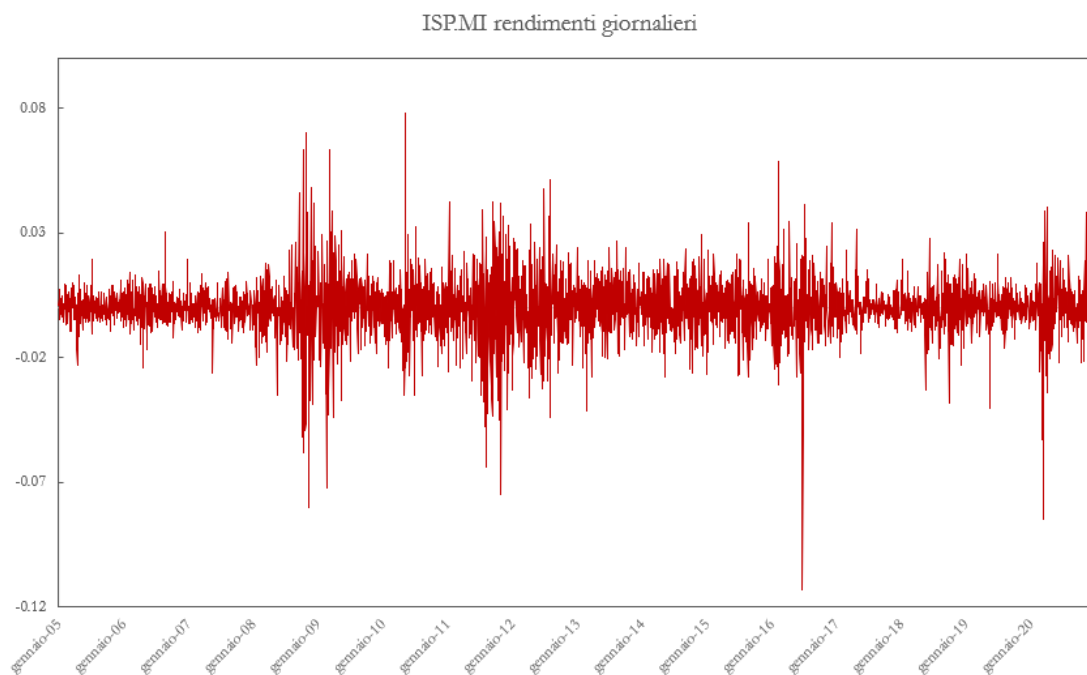


Figura 3.9: Grafico dei rendimenti giornalieri di Intesa Sanpaolo S.p.A.

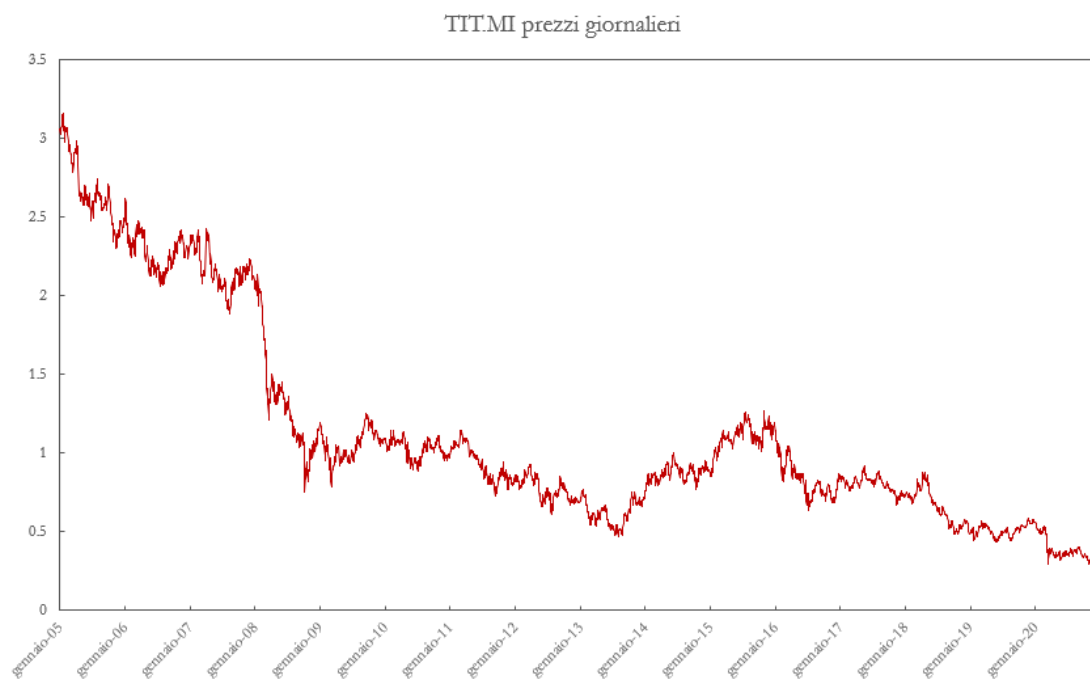


Figura 3.10: Grafico dei prezzi giornalieri di Telecom Italia S.p.A.

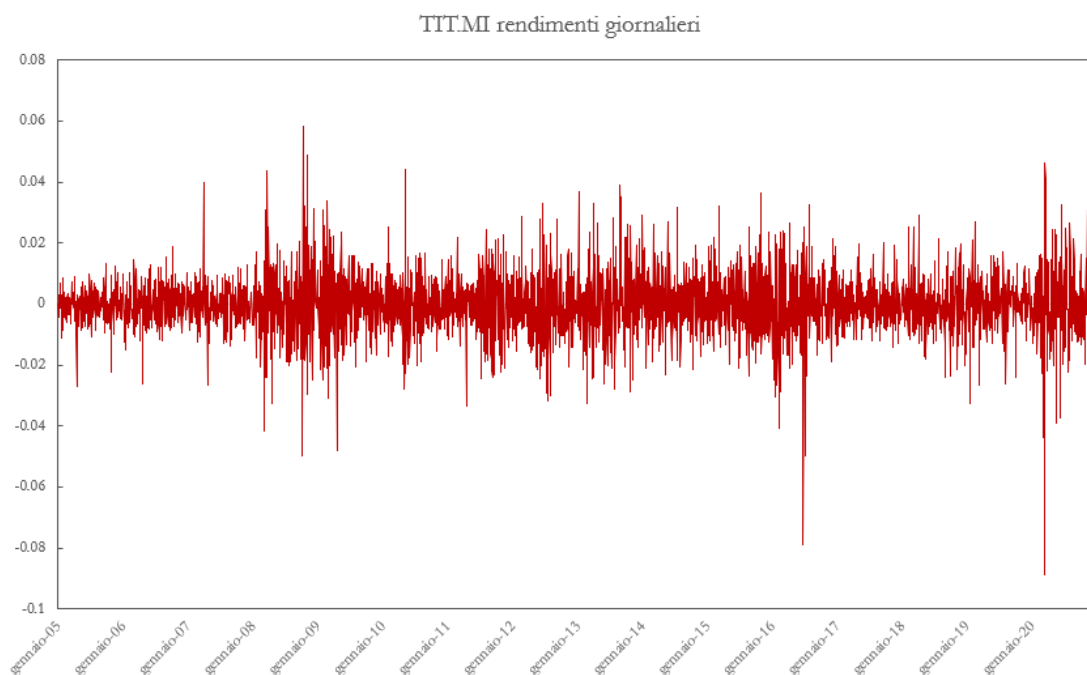


Figura 4: Grafico dei rendimenti giornalieri di Telecom Italia S.p.A.

Come si può rilevare osservando i grafici dei prezzi di ciascuno dei titoli, il periodo 2006-2008 in cui si è verificata la grande crisi finanziaria è caratterizzato da crolli verticali e una volatilità marcata. Particolare risulta però essere il caso relativo al titolo Davide Campari-Milano, in quanto nel periodo citato non presenta particolari oscillazioni di prezzo. Probabilmente, il settore alimentare in cui l'azienda in questione opera ha risentito in maniera minore degli effetti della crisi a differenza delle altre aziende qui analizzate. Un altro periodo interessante rilevabile dal grafico dei prezzi dei diversi titoli è un periodo molto più recente rispetto a quello evidenziato poco sopra, si tratta infatti del periodo 2019-2020. In questo periodo le economie mondiali sono state – e sono attualmente – minate dallo scoppio della pandemia del tristemente noto virus Covid-19. In questo periodo, infatti, si possono rilevare dei crolli dei prezzi notevoli e una volatilità estrema in tutte le serie storiche presentate. Per completare questa breve analisi delle serie storiche dei prezzi utilizzate, si riportano le principali statistiche descrittive di ciascun titolo riassunte nella Tabella 3.1. Come si può notare, il rendimento medio di ciascun titolo si attesta intorno allo zero e addirittura in due casi risulta essere negativo. I due titoli che riportano un rendimento negativo sono rispettivamente Intesa Sanpaolo S.p.A. e Telecom Italia S.p.A.

	CPR.MI	ENEL.MI	FCA.MI	ISP.MI	TIT.MI
Media	0.000221047	0.000030294	0.000233359	-0.000057962	-0.000223596
Varianza	0.000052691	0.000055038	0.000138078	0.000120596	0.000090832
Deviazione Standard	0.007258834	0.007418745	0.011750657	0.010981631	0.009530579
Massimo	0.047653976	0.073239779	0.104211339	0.078008211	0.058172865
Minimo	-0.076312884	-0.09607819	-0.085465891	-0.113174689	-0.088682572
Curtosi	6.374674118	12.60914311	5.393551097	9.356789418	5.730652786
Asimmetria	-0.159632557	-0.753798301	-0.139265975	-0.559259864	-0.29284293

Tabella 3.1: Principali statistiche descrittive calcolate utilizzando i rendimenti giornalieri.

Richiamando le Figure 3.8 e 3.10, si nota che l'andamento dei prezzi di Intesa Sanpaolo S.p.A. è estremamente volatile con frequenti crolli, anche verticali, della quotazione del titolo, mentre l'andamento dei prezzi di Telecom Italia S.p.A. è invece palesemente orientato al ribasso. Queste particolari caratteristiche fanno sì che il calcolo del rendimento medio porti ad ottenere un valore negativo, anche se entrambi si posizionano molto vicini allo zero. Un altro aspetto interessante rilevabile dalla lettura dei dati della Tabella 3.1 è il valore dell'asimmetria e della curtosi. In genere, un elevato valore della curtosi e un'asimmetria con valori come quelli riportati sono indice di non normalità nella distribuzione delle osservazioni di una serie storica. Tuttavia, per poter affermare con esattezza che le serie si allontanano dall'ipotesi di normalità bisognerebbe eseguire alcuni test specifici, i quali però non sono oggetto di questo lavoro; verrà pertanto ipotizzata la non normalità della distribuzione delle osservazioni, ipotesi sostenuta dai valori dei diversi momenti presentati nella Tabella 3.1.

3.7 Fase di calcolo

In questa sezione viene presentata una sintesi delle fasi che verranno attraversate dall'algoritmo, descrivendo brevemente quali sono gli elementi che in ciascuna verranno calcolati e/o approssimati. È possibile, quindi, riassumere il processo computazionale dell'algoritmo in tre fasi:

1. Selezione dei valori del *learning rate* α e del tasso di sconto γ , i quali vengono fissati pari a $\alpha = 5\%$ e $\gamma = 95\%$. I valori così selezionati vengono mantenuti per ciascuna delle $K = 500$ iterazioni al fine di garantire una soddisfacente approssimazione della funzione di valore. Come già anticipato, la numerosità delle iterazioni è necessaria affinché il processo possa convergere almeno in media ad una buona approssimazione della soluzione ottima;
2. Inizializzazione del vettore dei pesi θ_0 in modo casuale e sulla base dei settaggi impostati;
3. Ad ogni step temporale vengono ripetute le seguenti operazioni per ciascuno dei titoli:
 - Stima della funzione di valore $Q_t(s_t, a_t, \theta_t)$ per ogni azione a_t disponibile nello stato s_t . In questa fase vengono stimate tre diverse funzioni di valore: una per l'azione $a_t = -1$, una per l'azione $a_t = 0$ ed infine una per l'azione $a_t = 1$;
 - Scelta dell'azione a_t da eseguire partendo dallo stato s_t seguendo il criterio ε -greedy, il che significa che verrà selezionata l'azione a_t che porta alla stima della funzione di valore $Q_t = (s_t, a_t, \theta_t)$ maggiore;
 - Una volta scelta l'azione questa viene eseguita. Viene quindi fatta la rilevazione dello stato s_{t+1} , con conseguente ricompensa a seconda della bontà della precedente azione;
 - Viene stimata e selezionata la funzione di valore $\max_a Q(s_{t+1}, a_t, \theta_t)$;
 - Si procede all'aggiornamento della stima del parametro θ_t utilizzando tutti gli elementi ottenuti con i passaggi precedenti e utilizzando la regola di aggiornamento:

$$\theta_{t+1} = \theta_t + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a_t, \theta_t) - Q(s_t, a_t, \theta_t) \right] \nabla_{\theta_t} Q(s_t, a_t, \theta_t)$$

- Lo stato s_{t+1} diventa ora lo stato da cui ripartire con tutto il procedimento, pertanto $s_t = s_{t+1}$ e il processo comincia nuovamente partendo dalla fase 1.

Rimane un solo problema da affrontare, ossia come rendere effettivamente operativo il segnale che viene fornito dall' algoritmo. Dato che vengono eseguite 500 simulazioni ad ogni step temporale e che l'inizializzazione di ciascuna simulazione è casuale, ci si ritrova ad avere

per ogni step temporale un insieme di 500 possibili azioni. L'obiettivo è aggregare questi 500 segnali individuali per ricavarne uno solo operativo, il quale sarà effettivamente l'azione che verrà poi intrapresa. Per fare ciò si calcola la media di tutte le 500 simulazioni calcolate in un dato istante temporale nel modo seguente [10]:

$$\bar{a}_t = \frac{\sum_{k=1}^K a_{t,k}}{K} \quad 52$$

dove \bar{a}_t è il valore medio delle 500 simulazioni eseguite all'istante temporale t , $a_{t,k}$ rappresenta l'azione selezionata nello step temporale t durante l'esecuzione della k - *esima* simulazione rispetto al totale K . Per trasformare il numero che si ottiene in un segnale operativo a tutti gli effetti si utilizza il seguente sistema:

$$a_t = \begin{cases} -1 & \text{vendi o posizione corta sul mercato se } \bar{a}_t \in [-1, -1/3] \\ 0 & \text{star fuori dal mercato se } \bar{a}_t \in [-1/3, 1/3] \\ 1 & \text{compra o posizione lunga sul mercato se } \bar{a}_t \in [1/3, 1] \end{cases} \quad 53$$

In questo modo le performance di tutte le 500 *equity lines* vengono riassunte attraverso un'unica *equity line*.

Applicazione dell'algoritmo e risultati

In questo capitolo vengono presentati ed analizzati i risultati ottenuti dall'applicazione dell'algoritmo in tutte le possibili configurazioni illustrate nel capitolo precedente. Come già anticipato, l'algoritmo è stato applicato ad un *pool* di cinque titoli quotati sul mercato italiano. I risultati ottenuti da tali applicazioni sono stati raggruppati in tabelle in modo tale da garantire una visione sull'andamento generale del sistema di trading nel contesto di uno specifico titolo. Inoltre, data la numerosità delle configurazioni da analizzare, ne verranno fornite le rappresentazioni grafiche per alcuni particolari tipi di output.

In particolare, al fine di valutare l'operatività e le performance del sistema di trading implementato sono state selezionate alcune statistiche utili a comprendere se i risultati ottenuti con una specifica configurazione dell'algoritmo applicata ad uno dei titoli siano accettabili o meno. Per considerare una performance accettabile, in generale questa deve possedere alcune caratteristiche quali un rendimento positivo, una permanenza del capitale al di sopra del livello iniziale per almeno la metà del periodo osservato e un'operatività non eccessiva ma nemmeno troppo ridotta. Ovviamente, questi parametri di valutazione sono generali, ciò non toglie che possano configurarsi casi in cui tali caratteristiche non sono pienamente rispettate ma che comunque sono da considerare come accettabili. Le statistiche selezionate per effettuare una valutazione delle performance ottenute sono: il capitale finale al lordo e al netto dei costi di transazione; il rendimento medio annuo al lordo e al netto dei costi di transazione; l'indice di incremento del capitale anch'esso al lordo e al netto dei costi di transazione; la percentuale di volte in cui l'*equity line* netta è maggiore o uguale al capitale iniziale; il numero medio annuale di transazioni eseguite.

4.1 Presentazione delle statistiche

Il punto di partenza per il calcolo delle statistiche utili alla valutazione delle performance del sistema di trading è il segnale operativo generato attraverso l'equazione (52). Di fatto, una volta che tale valore viene calcolato e tradotto in un segnale operativo attraverso il sistema (53) è possibile calcolare l'*equity line* al lordo e al netto dei costi di transazione come già visto con l'equazione (50). Al termine del periodo di trading è possibile determinare il capitale finale lordo e netto osservando le *equity lines* che sono state prodotte. Partendo dall'ammontare del capitale finale viene calcolato il rendimento medio giornaliero e successivamente il rendimento medio annuo secondo le formule seguenti:

$$\text{Rendimento medio giornaliero} = \text{daily} = \left(\frac{\text{Capitale finale}}{\text{Capitale iniziale}} \right)^{\frac{1}{T}} - 1 \quad 54$$

$$\text{Rendimento medio annuo} = \bar{g} = (1 + \text{daily})^{252} - 1 \quad 55$$

dove T corrisponde al totale dei giorni di trading considerati – in questo caso $T = 4065$ – e 252 corrisponde ai giorni di contrattazione in un anno borsistico. Ovviamente a seconda che venga utilizzato il capitale finale al lordo o al netto dei costi di transazione si ottiene di conseguenza il rendimento lordo o netto.

In generale, il rendimento medio annuo ottenuto dall'applicazione dei diversi settaggi dell'algoritmo non basta a definire la bontà o meno di un risultato. In alcuni casi, valutare una performance limitandosi solamente al rendimento netto ottenuto può essere fuorviante: può accadere che il capitale finale sia infatti inferiore al capitale inizialmente investito, il che potrebbe indurre a pensare che la performance non sia da considerare buona; per poter dare una valutazione più completa è necessario osservare, contestualmente al rendimento medio netto, l'andamento che ha avuto l'*equity line* lungo tutto il periodo osservato.

È utile quindi calcolare un ulteriore indicatore da affiancare alla lettura del rendimento per capire quante volte l'*equity line* è maggiore o uguale livello del capitale iniziale, vale a dire l'indice di preservazione del capitale. Tale indicatore viene espresso in termini percentuali e viene calcolato attraverso la formula seguente:

$$\% = \frac{\Sigma}{T} * 100 \quad 56$$

dove Σ rappresenta la somma dei periodi in cui l'*equity line* è maggiore o uguale al livello del capitale iniziale e T è, come già visto, il numero totale dei giorni di contrattazione considerati. Il rendimento medio annuo fornisce una indicazione relativamente al livello del capitale finale, si potrebbe definire come un indicatore statico di performance; al contrario, l'indicatore % permette una lettura più dinamica dell'evoluzione del capitale durante tutto il periodo di contrattazione.

La terza statistica che viene presa in considerazione per la valutazione delle performance dell'algorithm è l'indice di incremento del capitale. Si tratta di un indicatore molto usato negli ultimi tempi dai *traders* e dai gestori di fondi in quanto offre una chiara visione riguardo la performance degli *asset*. Tale statistica permette di capire se il rendimento generato da un'attività – sia essa un singolo titolo, un fondo comune o qualsiasi altro tipo di attività– segua un trend crescente nel tempo. Una performance che genera una *equity line* che rimane positiva per la maggior parte del periodo di contrattazione e che segue un trend sostanzialmente crescente – anche se in maniera non regolare – è da preferire ad una *equity line* che è maggiore del capitale iniziale lungo tutto il periodo di *trading* ma che rimane piatta e costante. Per questo torna utile calcolare l'indice di incremento del capitale, il quale con frequenza prefissata verifica quante volte il livello del capitale attuale è maggiore o uguale al livello del capitale relativo al periodo precedente. Viene dunque fissato un *range* temporale che determina la frequenza di calcolo di questo indicatore, in questo caso verrà preso in considerazione un periodo di 21 giorni, corrispondenti ad un mese di contrattazioni sul mercato. La formula di calcolo di questo indicatore, espresso in termini percentuali, è la seguente:

$$Capital\ Incrementation = \frac{\Gamma}{T/21} * 100 \quad 57$$

dove Γ indica la somma delle volte in cui l'*equity line* $E_t \geq E_{t-21}$.

L'ultima statistica che verrà presa in considerazione in questa parte dedicata alla valutazione delle performance è l'operatività media annua, ossia il numero medio annuo di transazioni effettuate dal sistema di trading. Essa viene calcolata utilizzando la formula che segue:

$$\# = \frac{\Pi}{T/252}$$

dove Π rappresenta il numero totale delle transazioni effettuate dal sistema di trading. Questa statistica è utile per verificare se il sistema apre e chiude la propria posizione troppo frequentemente; se questo accadesse si tradurrebbe in costi di transazione elevati, i quali andrebbero inevitabilmente ad erodere l'eventuale *capital gain* ottenuto.

Al fine della valutazione delle performance, si terrà in particolare considerazione la redditività, quindi la statistica riportante il rendimento medio annuo, e la preservazione del capitale, cioè la statistica $\%$. Per considerare una performance accettabile è necessario che abbia ottenuto un rendimento positivo ma anche che il capitale permanga al di sopra del livello del capitale inizialmente investito per almeno la metà dell'orizzonte temporale osservato. In altri termini, il rendimento medio annuo deve essere positivo mentre la statistica $\%$ che indica la preservazione del capitale deve essere pari ad almeno il 50%. Le altre statistiche citate in questo stesso paragrafo, vale a dire il numero medio di operazioni annue $\#$ e l'indice di incremento del capitale, verranno tenute in considerazione nella valutazione di un risultato in subordine alle altre due; questi ultimi indicatori sono utili ad inquadrare meglio le *policies* seguite dall'algorithm fornendo un supporto all'analisi del rendimento medio annuo e della preservazione del capitale.

La rappresentazione delle performance verrà suddivisa a seconda del titolo: per ciascuno di essi verranno presentati e analizzati i dati dei risultati – raccolti in tabelle – derivanti dalla prima applicazione dell'algorithm, cioè la versione che utilizza i rendimenti logaritmici come input, e successivamente dalla seconda applicazione, la quale utilizza gli indicatori AT come input. Successivamente, i risultati delle due applicazioni verranno messi a confronto al fine di individuarne eventuali differenze, siano esse in miglioramento o in peggioramento. Al termine del confronto, verranno presentate alcune particolari configurazioni, relative a entrambe le applicazioni del sistema di trading, in modo tale da comprendere meglio quali sono state le scelte perseguite dal sistema di trading automatico.

4.2 Davide Campari-Milano

N=2			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	153.14	143.12	2.68	2.25	67.88	67.36	27.93	7.01
SR	5	15%	209.97	192.02	4.71	4.13	68.39	66.84	59.42	9.24
SR	21	5%	115.91	105.48	0.92	0.33	68.91	66.84	23.38	9.74
SR	21	15%	112.26	100.59	0.72	0.04	72.54	71.50	36.07	11.35
SOR	5	5%	104.32	93.71	0.26	-0.40	61.66	61.14	4.58	11.10
SOR	5	15%	28.96	22.40	-7.40	-8.86	64.77	62.69	4.38	26.54
SOR	21	5%	90.22	74.91	-0.64	-1.78	74.61	73.58	7.53	19.22
SOR	21	15%	77.09	72.07	-1.60	-2.01	93.26	91.71	8.46	6.94

N=5			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	109.44	98.41	0.56	-0.10	65.80	65.80	11.65	10.98
SR	5	15%	129.17	113.94	1.60	0.81	66.84	63.73	14.18	12.97
SR	21	5%	130.71	118.23	1.68	1.04	68.39	67.36	23.10	10.36
SR	21	15%	75.03	63.12	-1.77	-2.81	69.95	69.43	8.22	17.87
SOR	5	5%	111.94	102.54	0.70	0.16	63.21	63.21	33.56	9.06
SOR	5	15%	37.41	30.39	-5.92	-7.12	67.88	67.36	8.27	21.47
SOR	21	5%	74.13	63.79	-1.84	-2.75	73.06	71.50	4.63	15.51
SOR	21	15%	96.71	93.63	-0.21	-0.41	98.45	97.93	4.46	3.35

Tabella 4.1: Riepilogo delle statistiche relative a Davide Campari-Milano (input: rendimenti logaritmici).

I dati riportati nella Tabella 4.1 sono i risultati delle configurazioni dell'algorithm che utilizza i rendimenti logaritmici come input nel vettore di stato. Le performance del titolo in questione possono definirsi mediamente accettabili. È possibile, infatti, notare che il capitale finale in un buon numero di casi risulta essere maggiore del capitale inizialmente investito. Una cosa che salta immediatamente all'occhio è l'inadeguatezza del Sortino Ratio utilizzato come funzione di ricompensa per le configurazioni dell'algorithm applicate a questo titolo specifico. Come si può facilmente ricavare dalla Tabella 4.1, i risultati ottenuti con le configurazioni che prevedono l'uso del Sortino ratio nella quasi totalità dei casi sono negative. Altro fatto rilevabile dalla lettura dei dati in tabella è un indicatore di *capital preservation* – la statistica % – che solamente in un caso si presta ad essere considerato accettabile. Come si ricorderà, una statistica % è da considerarsi sufficiente se è pari ad almeno 50%, che tradotto in termini qualitativi significa che per almeno la metà dell'orizzonte temporale il capitale è

maggiore del capitale inizialmente investito. Le configurazioni che utilizzano il Sortino ratio come funzione di *reward* sono anche quelle con l'indice di preservazione del capitale tra i più bassi in assoluto, il che conferma l'inadeguatezza di tale funzione di ricompensa nell'ambito di questo titolo. Alcune delle configurazioni che utilizzano il Sortino ratio presentano anche un'operatività media annua elevata. Valutare le performance solamente sulla base di questa informazione è poco corretto, tuttavia in questo caso eseguire un numero elevato di transazioni non ha portato ad ottenere un rendimento positivo, anzi l'unico effetto ottenuto con un'operatività come questa è stata l'erosione del capitale finale. Un'altra tendenza ricavabile dai dati nella Tabella 4.1 è quella di ottenere rendimenti e una preservazione del capitale migliori in caso venga utilizzato un vettore di stato composto da $N = 2$ elementi. In generale, le performance di Davide Campari-Milano sono mediamente accettabili, anche se va evidenziato un limite dell'algorithm nel produrre delle *equity lines* che permangano al di sopra del livello del capitale iniziale sufficientemente a lungo.

Vengono ora presentate le statistiche relative alla seconda applicazione dell'algorithm, cioè il sistema di trading automatico che utilizza un insieme di indicatori di Analisi Tecnica come descrittori dello stato.

N=3			Capitale Finale		Rendimento annuo		Incremento del capitale			
Reward	L	ε	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	88.94	81.38	-0.72	-1.27	73.58	73.58	3.91	9.18
SR	5	15%	93.15	81.39	-0.44	-1.27	66.84	66.32	4.38	13.96
SR	21	5%	217.32	208.11	4.93	4.65	73.06	73.06	89.52	4.47
SR	21	15%	121.24	112.07	1.20	0.71	74.09	74.09	32.71	8.13
SOR	5	5%	123.89	114.46	1.34	0.84	69.95	69.43	63.18	8.19
SOR	5	15%	49.77	41.42	-4.24	-5.32	67.36	66.32	18.90	18.98
SOR	21	5%	102.30	90.95	0.14	-0.59	68.91	67.36	21.61	12.16
SOR	21	15%	67.89	55.49	-2.37	-3.59	82.38	78.24	4.82	20.84

Tabella 4.2: Riepilogo delle statistiche relative a Davide Campari-Milano (input: indicatori AT).

Passando in rassegna i dati raccolti nella Tabella 4.2 si può affermare che le performance del titolo non sono mediamente buone. Ci sono alcuni risultati che in realtà sono accettabili – con un rendimento positivo e una statistica % maggiore del 50% – ma che non sono sufficienti a caratterizzare la performance generale come accettabile.

La tendenza che si osserva è un rendimento migliore quando viene utilizzato lo Sharpe Ratio come funzione di *reward*, esattamente come osservato nella prima applicazione dell'algorithm, e una preferenza per le azioni *greedy*, quindi un funzionamento più efficiente con un

parametro $\varepsilon = 5\%$. L'algoritmo anche in questo caso non riesce a produrre un livello del capitale che permanga consistentemente al di sopra del capitale iniziale, di fatto oltre la metà dei settaggi presenta un indice di preservazione del capitale insufficiente. L'inadeguatezza del Sortino ratio come funzione di ricompensa spicca anche in questa seconda applicazione, con risultati che presentano una redditività quasi in ogni caso negativa ed una statistica % quasi sempre al di sotto della soglia del 50%. Le caratteristiche della seconda applicazione del sistema di trading sono molto simili a quelle relative alla prima applicazione, può essere utile un confronto tra le due per ricavarne qualche informazione aggiuntiva.

			Rendimento netto annuo			Statistica %			Statistica #		
Reward	L	ε	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione
SR	5	5%	2.25	-0.10	-1.27	27.93	11.65	3.91	7.01	10.98	9.18
SR	5	15%	4.13	0.81	-1.27	59.42	14.18	4.38	9.24	12.97	13.96
SR	21	5%	0.33	1.04	4.65	23.38	23.10	89.52	9.74	10.36	4.47
SR	21	15%	0.04	-2.81	0.71	36.07	8.22	32.71	11.35	17.87	8.13
SOR	5	5%	-0.40	0.16	0.84	4.58	33.56	63.18	11.10	9.06	8.19
SOR	5	15%	-8.86	-7.12	-5.32	4.38	8.27	18.90	26.54	21.47	18.98
SOR	21	5%	-1.78	-2.75	-0.59	7.53	4.63	21.61	19.22	15.51	12.16
SOR	21	15%	-2.01	-0.41	-3.59	8.46	4.46	4.82	6.94	3.35	20.84

Tabella 4.3: Confronto tra applicazioni dell'algoritmo al titolo Davide Campari-Milano.

Nella Tabella 4.3 è possibile vedere un confronto dei valori di tre degli indicatori usati per valutare una performance: il rendimento netto medio annuo, l'indice di preservazione del capitale (la statistica %) e l'operatività media annua (la statistica #). Osservando i risultati nella Tabella 4.3 in maniera generale, sembrano essere migliori quelli ottenuti con la prima versione del sistema di trading dato che in più occasioni è riuscito ad ottenere un capitale finale maggiore del capitale iniziale investito. Tuttavia, i risultati ottenuti dalla prima applicazione dell'algoritmo avevano un difetto, ossia una statistica % sufficiente in un unico caso – si veda la configurazione $N = 2, L = 5, \varepsilon = 15\%, SR$ nella Tabella 4.1 – e decisamente insufficiente in tutti gli altri. Nella seconda applicazione dell'algoritmo, la configurazione che ha prodotto il miglior risultato – si veda $N = 3, L = 21, \varepsilon = 5\%, SR$ nella Tabella 4.2 – possiede anche una percentuale di permanenza del capitale al di sopra del livello del capitale iniziale migliore, il che permette di classificarla come una performance in grado di produrre dei risultati più bilanciati, con una redditività soddisfacente ed un buon indicatore di *capital preservation*. La tendenza ad ottenere risultati migliori quando viene utilizzato lo Sharpe Ratio come funzione di *reward* è riscontrabile in entrambe le applicazioni, il che induce a ritenere il Sortino ratio inadatto a fornire dei *feedback* corretti nell'ambito di

questo titolo. Alcune configurazioni hanno prodotto risultati disastrosi sia con la prima applicazione che con la seconda; si veda, ad esempio, la configurazione $L = 5, \varepsilon = 15\%, SOR$ nella Tabella 4.3. Utilizzando però un diverso insieme di descrittori dello stato, l'algoritmo è riuscito a migliorare leggermente performance negative come quella in questione. Come si può osservare, l'impiego di indicatori AT come input nel vettore di stato ha portato ad ottenere un miglioramento in termini di redditività e preservazione del capitale della configurazione in questione. Il rendimento medio annuo ottiene un miglioramento del 3% circa mentre la statistica % arriva a superare il 18%. Questa particolare configurazione si caratterizzava anche per un'operatività media annua elevata e con gli indicatori AT questa si riduce, con una minore incidenza dei costi di transazione sul capitale finale. La riduzione dell'operatività media annua sicuramente ha influito sull'aumento della redditività – pur rimanendo questa negativa –, in quanto la mole dei costi di transazione applicati si è ridotta. Un'altra configurazione che sembra aver giovato dell'utilizzo di input diversi è $L = 5, \varepsilon = 5\%, SOR$, la quale presentava un rendimento negativo e una statistica % decisamente pessima con $N = 2$: con il nuovo algoritmo la medesima configurazione ha prodotto una performance positiva con un rendimento netto medio annuo del 0.84% e una statistica % = 63.18% soddisfacente. Al termine dell'analisi, è possibile affermare che il nuovo insieme di descrittori è stato in grado di produrre risultati più consistenti rispetto a quelli ottenuti con la versione di partenza; inoltre, ha migliorato leggermente le performance là dove prima si erano ottenuti risultati non buoni. Si può dunque dire che in questo caso l'utilizzo degli indicatori di Analisi Tecnica come input al posto dei rendimenti logaritmici sia stato fruttuoso.

Per completare l'analisi del titolo in questione, si riportano alcuni settaggi ritenuti degni di nota per ciascuna delle applicazioni. Relativamente alla prima applicazione, interessante è la configurazione $N = 2, L = 5, \varepsilon = 15\%, SR$; questa combinazione di parametri ha portato ad ottenere il miglior risultato tra tutti quelli ottenuti ed è possibile apprezzarne la rappresentazione grafica nella Figura 4.1. Tale configurazione ha prodotto un capitale finale netto di 192.02 € con un rendimento medio annuo pari al 4.71%. L'indice di preservazione del capitale è pari al 59.42%, un valore considerabile accettabile dato che è maggiore del 50%. Osservando nel dettaglio il secondo pannello, attorno all'osservazione $t = 750$ l'algoritmo decide di aprire una posizione *short* per poi chiuderla poco dopo l'osservazione $t = 1000$. Questa particolare decisione dimostra un buon *market timing* dato che nel momento in cui il sistema di trading decide di vendere allo scoperto il titolo il prezzo

stava per giungere al termine di un andamento sostanzialmente piatto. Quando poi la posizione viene chiusa il prezzo risulta essere al suo minimo assoluto.

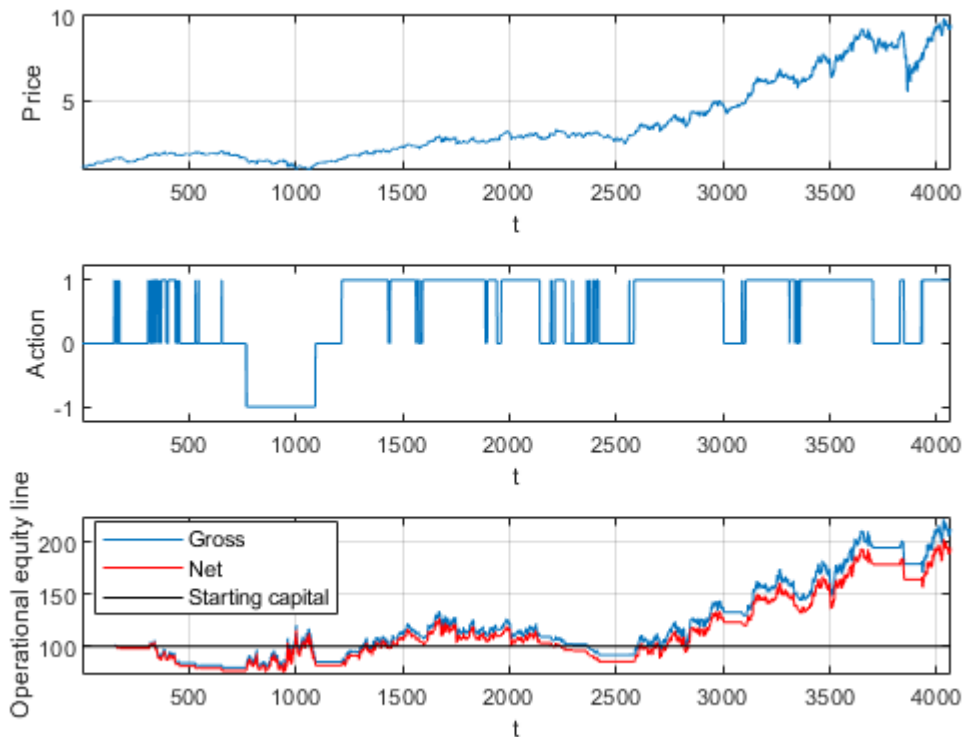


Figura 4.1: Davide Campari-Milano $N = 2, L = 5, \varepsilon = 15\%, SR$ (input: rendimenti logaritmici).

Un altro caso simile si ha tra l'osservazione $t = 2500$ e $t = 3000$ in cui l'algoritmo acquista il titolo proprio quando un trend crescente inizia e lo rivende al primo ribasso del prezzo che incontra. L'*equity line* prodotta dal sistema di trading, osservabile nel terzo pannello della Figura 4.1, presenta un andamento crescente, fatto che indica che l'algoritmo ha appreso correttamente il comportamento della serie storica, agendo di conseguenza.

Per quanto riguarda la seconda applicazione dell'algoritmo, vengono riportati due settaggi interessanti utili a comprendere quali *policies* da seguito l'algoritmo. Il primo ad essere riportato è il miglior risultato ottenuto, vale a dire la configurazione $N = 3, L = 21, \varepsilon = 5\%, SR$ di cui si trova la rappresentazione grafica nella Figura 4.2, mentre il secondo output riportato è una performance in perdita relativa alla configurazione $N = 3, L = 5, \varepsilon = 5\%, SR$, rappresentata graficamente nella Figura 4.3.

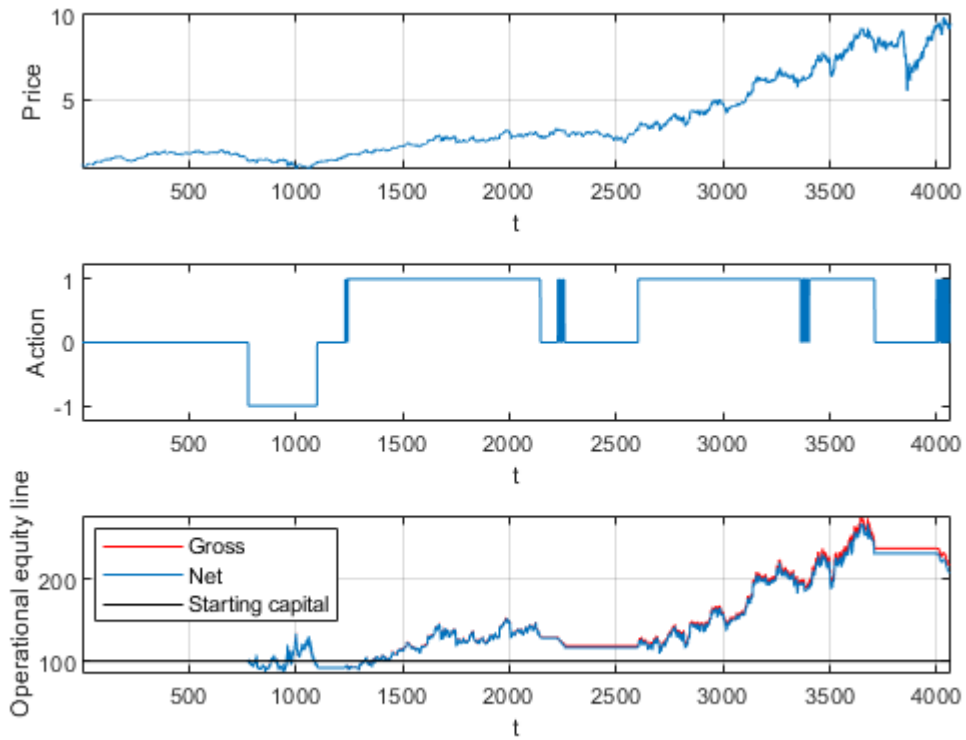


Figura 4.2: Davide Campari-Milano $N = 3, L = 21, \varepsilon = 5\%, SR$ (input: indicatori AT).

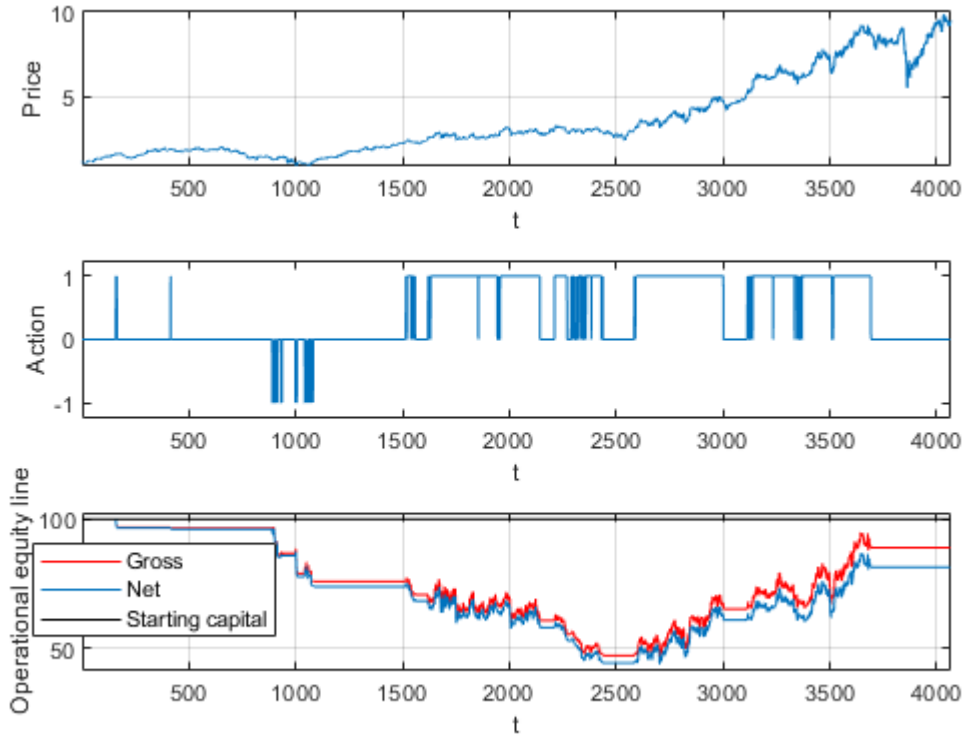


Figura 4.3: Davide Campari-Milano $N = 3, L = 5, \varepsilon = 5\%, SR$ (input: indicatori AT)

Facendo riferimento alla prima configurazione, ossia la Figura 4.2, l'*equity line* ottenuta è da considerarsi buona dato che cresce per tutta la durata del periodo osservato. Seppur con qualche oscillazione, l'algoritmo è stato in grado di costruire il capitale finale gradualmente e in maniera efficiente raggiungendo l'ammontare di 208.11 € con un rendimento netto medio annuo pari al 4.65%. Anche la permanenza dell'*equity line* netta al di sopra del capitale iniziale presenta un risultato notevole con una statistica $\% = 89.52\%$, fatto deducibile anche solo dando uno sguardo al terzo *panel* della Figura 4.2. Ad ulteriore conferma della qualità dei risultati ottenuti con questa configurazione, c'è l'indice di incremento del capitale netto pari al 73.06% - si veda la Tabella 4.2 -. Dato che in questo caso specifico non ci sono notevoli periodi di inattività che possano "distorcere" la valutazione di questo indice - giusto all'inizio dell'orizzonte temporale osservato l'algoritmo "tentenna" ad entrare nel mercato - si può considerare un risultato interessante. Confrontando il grafico dell'andamento del prezzo (primo pannello) con il grafico dell'operatività (secondo pannello) si può constatare che il sistema di trading abbia identificato con successo i principali trend in crescita della serie storica, aprendo una posizione lunga al momento giusto e mantenendola aperta per tutta la durata della tendenza - si veda ad esempio attorno all'osservazione $t = 1300$ e $t = 2500$ - . Dato che lo scopo degli indicatori di Analisi Tecnica utilizzati come input è quello di agevolare l'individuazione delle tendenze nella serie dei prezzi ed agire di conseguenza, si può dunque affermare che in questo caso l'obiettivo è stato raggiunto con risultati soddisfacenti. Per quanto riguarda la seconda configurazione riportata in Figura 4.3, si può vedere come questo sia un output in perdita. Fatta questa premessa, se si osserva l'andamento dell'*equity line* nel terzo pannello della Figura 4.3 si può notare come l'algoritmo ottenga una performance estremamente negativa nel periodo in cui la tendenza del prezzo non è molto evidente. Al contrario, quando il trend comincia a diventare più marcato il sistema di trading segna una ripresa e riguadagna quanto perduto nel periodo precedente. Evidentemente, nel primo periodo l'algoritmo non è stato in grado di cogliere appieno i segnali operativi a causa della lenta crescita dei prezzi del titolo, il trend è quindi risultato di difficile individuazione per il sistema di trading. Si può arrivare a concludere che, nell'ambito del presente titolo, l'utilizzo di Medie Mobili e del RSI come descrittori dello stato risulta essere efficace sicuramente nel caso in cui il trend sia piuttosto marcato, mentre quando questo tende ad essere meno evidente i risultati spesso non sono buoni.

4.3 Enel S.p.A.

N=2			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	108.66	103.77	0.52	0.23	87.05	86.53	12.25	4.77
SR	5	15%	111.07	101.82	0.65	0.11	73.58	71.50	26.18	8.99
SR	21	5%	48.44	46.64	-4.40	-4.62	75.13	74.61	18.75	3.91
SR	21	15%	82.18	75.96	-1.21	-1.69	75.65	75.13	18.90	8.12
SOR	5	5%	67.34	63.00	-2.42	-2.82	75.13	75.13	19.39	6.88
SOR	5	15%	100.99	91.85	0.06	-0.53	77.20	76.68	23.67	9.80
SOR	21	5%	81.58	76.41	-1.25	-1.65	83.94	82.38	16.04	6.76
SOR	21	15%	69.20	65.24	-2.26	-2.61	93.26	92.75	12.08	6.08

N=5			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	90.78	86.36	-0.60	-0.91	86.01	86.01	19.23	5.15
SR	5	15%	95.19	87.58	-0.31	-0.82	73.06	72.54	19.75	8.63
SR	21	5%	81.31	79.14	-1.28	-1.44	78.76	78.76	18.94	2.79
SR	21	15%	67.05	60.95	-2.45	-3.03	77.72	74.61	19.08	9.87
SOR	5	5%	104.25	97.53	0.26	-0.16	79.27	78.76	18.96	6.89
SOR	5	15%	70.85	62.98	-2.12	-2.83	78.76	77.72	17.93	12.16
SOR	21	5%	85.34	79.55	-0.98	-1.41	88.60	87.05	15.56	7.26
SOR	21	15%	87.18	83.70	-0.85	-1.10	95.85	95.34	13.20	4.22

Tabella 4.4: Riepilogo delle statistiche relative al titolo Enel S.p.A. (input: rendimenti logaritmici).

I dati riportati nella Tabella 4.4 sono relativi alla prima applicazione dell'algoritmo al titolo in questione. Le performance del titolo Enel S.p.A. non sono buone; solamente in due casi il sistema di trading riesce ad ottenere un rendimento medio annuo positivo. Come si può vedere dalla Figura 4.4, la quale riporta tutte le *equity lines* prodotte in fase di simulazione operativa con la configurazione $N = 5, L = 5, \epsilon = 15\%, SOR$, queste sono molto disperse e caratterizzate da una tendenza al ribasso¹⁶. Un grafico come questo indica che l'algoritmo ha avuto difficoltà nell'individuare le operazioni di trading più adeguate da eseguire. Osservando i dati nella tabella, non è possibile individuare con quale funzione di *reward* il sistema operi meglio nell'ambito di questo titolo. Per entrambi i valori di N , le due funzioni di ricompensa producono perdite più o meno contenute, non è possibile definire quale delle due funzioni porti ad ottenere i risultati peggiori. Allo stesso modo, non si individua alcun elemento che induca a pensare che l'algoritmo operi meglio con un preciso valore del parametro ϵ . D'altra parte, la dispersione delle *equity lines* in Figura 4.4 indica una certa difficoltà di apprendimento della serie storica, il che si riflette sicuramente nei dati in tabella.

¹⁶ Questa considerazione vale per quasi tutti i settaggi relativi al titolo ENEL.MI

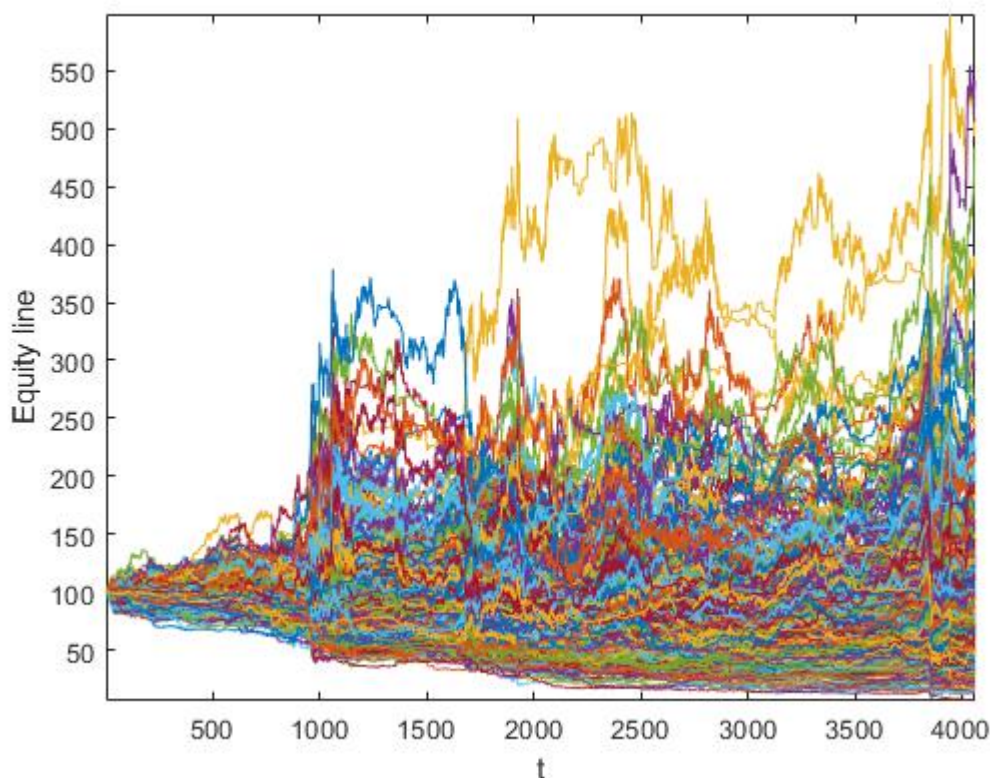


Figura 4.4: *Equity lines* relative al titolo Enel S.p.A. $N = 5, L = 5, \varepsilon = 5\%, SOR$.

In definitiva, i risultati del sistema di trading applicato al titolo Enel S.p.A. non sono brillanti, con una redditività quasi in ogni caso negativa ed una preservazione del capitale che non arriva nemmeno a sfiorare la soglia del 30% in nessun caso.

Nel proseguo verranno presentate, nella Tabella 4.5, le statistiche relative alla seconda applicazione dell'algoritmo al titolo in questione.

N=3			Capitale Finale		Rendimento annuo		Incremento del capitale			
Reward	L	ε	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	113.90	108.49	0.81	0.51	91.71	91.19	21.61	5.02
SR	5	15%	107.31	98.19	0.44	-0.11	77.20	75.65	31.06	9.18
SR	21	5%	45.81	42.49	-4.73	-5.17	76.68	76.17	15.01	7.75
SR	21	15%	144.96	136.52	2.33	1.95	80.83	79.79	88.90	6.20
SOR	5	5%	90.58	84.13	-0.61	-1.07	80.83	80.31	11.22	7.63
SOR	5	15%	84.62	76.86	-1.03	-1.62	78.24	77.20	17.25	9.92
SOR	21	5%	105.49	97.28	0.33	-0.17	78.24	77.20	24.74	8.37
SOR	21	15%	100.15	89.58	0.01	-0.68	86.53	84.97	52.40	11.54

Tabella 4.5: Riepilogo delle statistiche relative al titolo Enel S.p.A. (input: indicatori AT).

Anche in questo caso le performance del sistema di trading non sono buone. In molti casi il rendimento lordo è positivo ma non è sufficientemente elevato da garantire anche un rendimento netto altrettanto positivo. Non è possibile individuare una particolare tendenza nel comportamento dell'algorithm: inizialmente sembrerebbe operare meglio quando viene utilizzato lo Sharpe ratio come funzione di ricompensa; la configurazione $L = 21, \varepsilon = 5\%$, SR ha però prodotto il risultato peggiore in assoluto, mentre i risultati in perdita relativi alle configurazioni che utilizzano il Sortino ratio sono più contenuti e portano a riconsiderare l'effettiva efficacia dello Sharpe ratio. Per quanto riguarda la preservazione del capitale, è possibile osservare che la statistica % in un paio di casi arriva a superare la soglia del 50% – uno di questi casi è anche la configurazione che ha ottenuto il miglior risultato tra tutti –. Nonostante ciò, il valore dell'indice di preservazione del capitale di tutte le restanti configurazioni è talmente basso da non permettere di dare un giudizio positivo a tutta la performance generale.

Interessante è però il confronto tra le due applicazioni al titolo Enel S.p.A.. La Tabella 4.6 riporta una comparazione tra tre degli indicatori già analizzati nelle Tabelle 4.4 e 4.5, rispettivamente il rendimento netto medio annuo, l'indice di preservazione del capitale e l'operatività media annua. Si può apprezzare un generale miglioramento della redditività dato che nella maggior parte dei casi della seconda applicazione il rendimento lordo è positivo, e anche se questo viene poi eroso dai costi di transazione, risulta comunque un salto di qualità rispetto ai risultati ottenuti applicando l'algorithm originale alla serie storica dei prezzi del titolo. Un limite riscontrabile nella performance della seconda applicazione dell'algorithm è un peggioramento del risultato nel caso della configurazione $L = 21, \varepsilon = 5\%$, SR : questa particolare combinazione dei parametri è risultata essere problematica anche nel caso della prima applicazione dell'algorithm, producendo l'output peggiore in assoluto nel caso di $N = 2$. In questo caso, l'utilizzo degli indicatori AT non ha migliorato la situazione, portando anzi ad un peggioramento e ottenendo di fatto il peggior risultato tra tutti quelli analizzati. Un altro miglioramento riscontrabile confrontando la prima e la seconda applicazione è un leggero miglioramento della preservazione del capitale, vale a dire della statistica %. In molti casi, la percentuale di volte in cui il capitale è maggiore del livello del capitale inizialmente investito aumenta; questo fatto indica che l'algorithm ha prodotto dei risultati con *equity lines* che permangono per più tempo al di sopra del capitale iniziale lungo tutto l'orizzonte osservato. In definitiva, nonostante non si tratti di un miglioramento decisivo, l'utilizzo degli indicatori AT ha permesso di migliorare sia la redditività che la permanenza del capitale al di

sopra del capitale iniziale, il che porta anche in questo caso a definire fruttuoso l'utilizzo di questo genere di input all'interno del vettore di stato.

Reward	L	ε	Rendimento netto annuo			Statistica %			Statistica #		
			Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione
SR	5	5%	0.23	-0.91	0.51	12.25	19.23	21.61	4.77	5.15	5.02
SR	5	15%	0.11	-0.82	-0.11	26.18	19.75	31.06	8.99	8.63	9.18
SR	21	5%	-4.62	-1.44	-5.17	18.75	18.94	15.01	3.91	2.79	7.75
SR	21	15%	-1.69	-3.03	1.95	18.90	19.08	88.90	8.12	9.87	6.20
SOR	5	5%	-2.82	-0.16	-1.07	19.39	18.96	11.22	6.88	6.89	7.63
SOR	5	15%	-0.53	-2.83	-1.62	23.67	17.93	17.25	9.80	12.16	9.92
SOR	21	5%	-1.65	-1.41	-0.17	16.04	15.56	24.74	6.76	7.26	8.37
SOR	21	15%	-2.61	-1.10	-0.68	12.08	13.20	52.40	6.08	4.22	11.54

Tabella 4.6: Confronto tra applicazioni dell'algoritmo al titolo Enel S.p.A.

Per analizzare più approfonditamente le performance dell'algoritmo, vengono riportati alcune delle configurazioni che hanno prodotto output particolari sia per la prima applicazione che per la seconda. Nell'ambito della prima applicazione dell'algoritmo, si riporta nella Figura 4.5 la rappresentazione grafica della configurazione $N = 2, L = 5, \varepsilon = 5\%, SR$, la quale è la configurazione che ha ottenuto il miglior risultato tra tutti (si veda la Tabella 4.4).

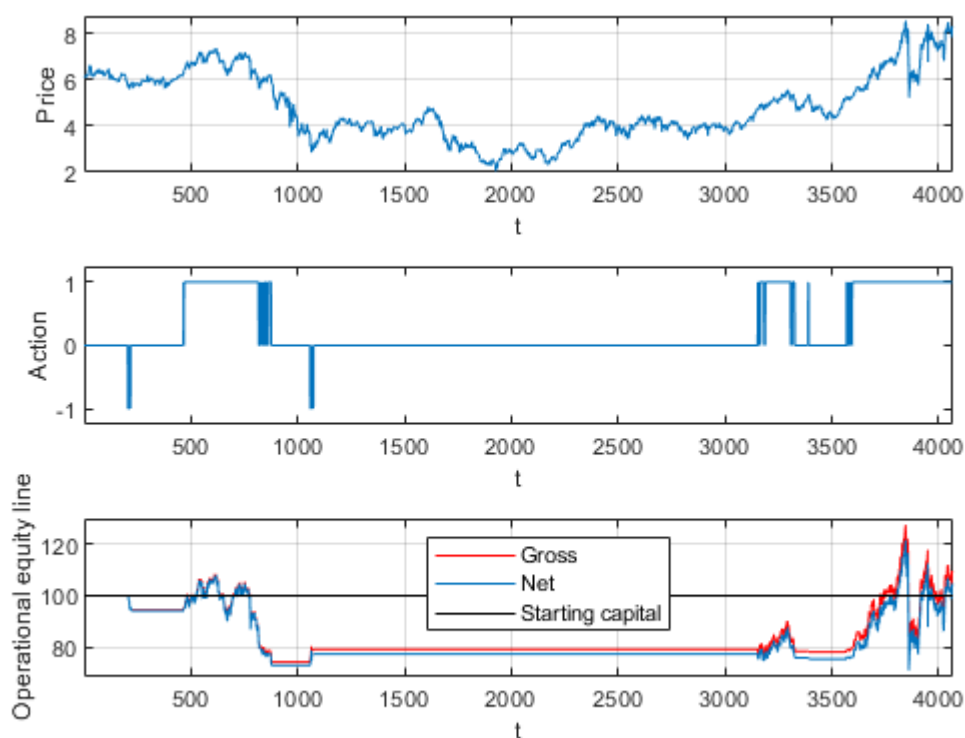


Figura 4.5: Enel S.p.A. $N = 2, L = 5, \varepsilon = 5\%, SR$ (input: rendimenti logaritmici).

Va precisato che, nonostante questa configurazione abbia ottenuto un risultato migliore delle altre non è assolutamente da considerarsi una performance accettabile per una serie di motivi. Come già anticipato, nonostante questa sia la performance migliore tra tutte quelle ottenute, dalla Figura 4.5 si può facilmente vedere quanto essa sia una performance per nulla accettabile. L'attenzione viene immediatamente attirata dal secondo pannello in cui vi è riportata l'operatività del sistema di trading: si può infatti individuare un lungo periodo di tempo, che va circa dall'osservazione $t = 1000$ all'osservazione $t = 3200$, in cui l'algoritmo sceglie sistematicamente di rimanere fuori dal mercato. Se si osserva l'andamento del prezzo del titolo in quel lasso di tempo si può vedere che questo attraversa una fase caratterizzata da una evidente volatilità e che non segue una tendenza ben precisa e definita. A causa di ciò, l'algoritmo preferisce evitare di prendere una posizione e di rimanere fuori dal mercato, con due principali conseguenze dovute a questa scelta: innanzitutto rimanendo fuori dal mercato per un periodo così prolungato vengono perse eventuali opportunità di profitto; in secondo luogo, essendo la *equity line* già al di sotto del livello del capitale iniziale quando comincia questo periodo di inattività, rimanere fuori dal mercato non fa altro che rendere tale situazione persistente. Solo in ultima battuta l'algoritmo decide di reagire in qualche modo e riesce ad ottenere un capitale finale appena superiore a quello iniziale. Intuibilmente, la percentuale di volte in cui l'*equity line* è maggiore o uguale al livello del capitale iniziale non può che essere insufficiente, di fatto è pari a $\% = 12.25$.

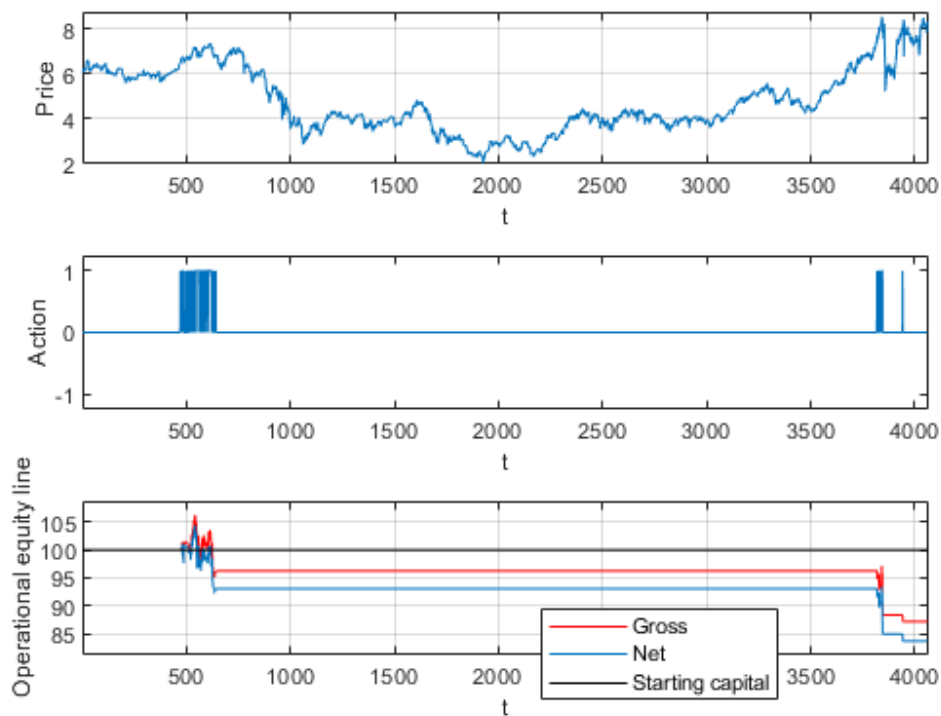


Figura 4.6: Enel S.p.A. $N = 5, L = 21, \varepsilon = 15\%, SOR$ (input: rendimenti logaritmici).

Per approfondire l'analisi, si è deciso di riportare il grafico di un'altra configurazione in perdita relativa ai parametri $N = 5, L = 21, \varepsilon = 15\%, SOR$, osservabile nella Figura 4.6. Si può notare una somiglianza spiccata con il settaggio analizzato poc'anzi: il grafico dell'operatività è pressoché piatto così come lo è, di conseguenza, l'*equity line*. Anche in questo caso, dunque, si rileva un problema di operatività dell'algoritmo per quanto concerne il titolo in questione, il che porta a pensare che il sistema di trading necessiti di qualche modifica per riuscire a cogliere il comportamento del titolo. A tal proposito, si è deciso di implementare una modifica al sistema di trading andando ad intervenire sul meccanismo di selezione dell'azione. Più dettagliatamente, è stata apportata una piccola modifica strutturata come una regola di aggiornamento dell'algoritmo, la quale permette di verificare quali sono state le ultime venti azioni selezionate e di aggiornare di conseguenza il sistema (53). Dato che in questo caso il problema riscontrato è una inattività prolungata, l'algoritmo è stato riapplicato in modo tale da restringere l'intervallo di selezione dell'azione $a_t = 0$ – passando da $[-1/3, 1/3]$ a $[-1/4, 1/4]$ – se l'azione in questione è stata selezionata per almeno venti volte consecutive. In termini più precisi, l'algoritmo, dopo aver selezionato l'azione da eseguire allo step temporale t , controlla quali sono state le azioni selezionate nei venti istanti temporali precedenti. Viene sostanzialmente creato un vettore $\mathbf{r} = [a_{t-19} \ a_{t-18} \ \dots \ a_t]$ che contiene le ultime venti azioni selezionate. Questo vettore viene poi confrontato con un vettore di zeri, contenente a sua volta venti elementi. Se i due vettori risultano essere uguali, e quindi se ogni elemento del vettore \mathbf{r} è uguale a zero, allora il sistema di selezione dell'azione (53) viene modificato, riducendo l'intervallo di selezione dell'azione “*stay out of the market*” $a_t = 0$. Il sistema di selezione dell'azione così aggiornato viene poi mantenuto per tutta la durata dell'orizzonte temporale osservato. Nella Figura 4.7 è possibile osservare il risultato dell'esperimento condotto. La differenza tra questo e il risultato precedente è chiara: innanzitutto, la performance dell'algoritmo è migliorata dato che ha ottenuto un capitale finale netto 106.42€ e un rendimento netto medio annuo pari a 0.39%; in secondo luogo, l'operatività è migliorata, fatto riscontrabile visivamente nel secondo pannello della Figura 4.7 in cui non è più presente un periodo di inattività che perdura per quasi tutto l'orizzonte temporale osservato. In ultima battuta, vale la pena esaminare anche la statistica %, la quale passa da 13.20% a 99.78%. Considerando la performance negativa ottenuta con la medesima configurazione senza però la regola di aggiornamento del sistema di selezione, si può arrivare a dire di aver ottenuto un risultato degno di nota, con un miglioramento generalizzato della redditività, della preservazione del capitale ma soprattutto dell'operatività, dato che il problema dell'inattività prolungata viene superato.

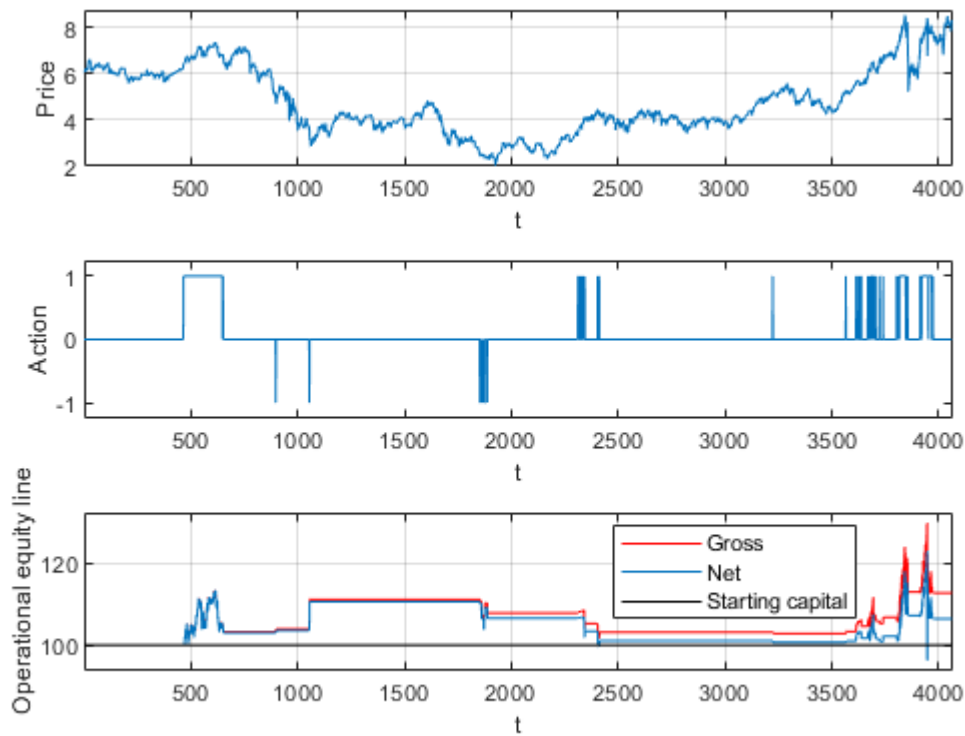


Figura 4.7: Enel S.p.A. $N = 5, L = 21, \varepsilon = 5\%$, SOR , algoritmo modificato con regola di aggiornamento (input: rendimenti logaritmici).

Per quanto riguarda i risultati ottenuti con la seconda applicazione dell'algoritmo, si riporta la performance relativa alla configurazione $N = 3, L = 21, \varepsilon = 15\%$, SR , riportata graficamente nella Figura 4.8. Partendo dall'analisi dell'andamento dell'*equity line*, si può notare come nel periodo centrale del periodo temporale osservato – partendo all'incirca dall'osservazione $t = 1000$ – il sistema non sia riuscito ad individuare una tendenza ben definita nei prezzi, pertanto le performance risultano abbastanza altalenanti. Successivamente, però, inizia un periodo in cui il trend in aumento dei prezzi diventa più marcato e l'algoritmo riesce, infatti, ad ottenere dei risultati decisamente migliori di quelli ottenuti nel periodo precedente. Il sistema di trading sembra quindi essere più sensibile all'individuazione di un trend nell'andamento dei prezzi quando questo si presenta in modo marcato, esattamente la stessa tendenza individuata precedentemente nell'ambito del titolo Davide Campari-Milano. Probabilmente, scegliendo intervalli temporali più brevi per il calcolo delle Medie Mobili utilizzate come input – combinando ad esempio una Media Mobile a 5 giorni con una calcolata a 10 giorni – si potrebbe rendere l'algoritmo più sensibile ai cambi di direzione del prezzo agevolando l'individuazione dei trend.

L'operatività e la permanenza dell'*equity line* al di sopra del livello del capitale iniziale sono entrambe buone, rispettivamente pari a $\% = 88.90\%$ e $\# = 6.20$. Un'operatività media come quella registrata in questo caso sembra essere adeguata, ciò non toglie che l'algoritmo avrebbe potuto anche effettuare più transazioni e ottenere comunque un buon risultato. Un'importante considerazione da fare riguarda il problema di operatività riscontrato durante l'analisi delle configurazioni relative alla prima applicazione dell'algoritmo. Come si ricorderà, è stato rilevato un limite dell'algoritmo per quanto riguarda la selezione dell'azione "*stay out of the market*". Nella prima applicazione, di fatto, tale azione è stata selezionata sistematicamente per molti istanti temporali successivi, causando periodi di inattività in cui l'algoritmo ha perso eventuali opportunità di profitto. Utilizzando però gli indicatori AT all'interno del vettore che descrive lo stato, tale problema è stato superato, rendendo quindi superfluo un intervento sul sistema di selezione dell'azione come si è reso invece necessario in precedenza. Questo fatto è un'ulteriore prova dell'efficacia degli indicatori di Analisi Tecnica utilizzati come descrittori.

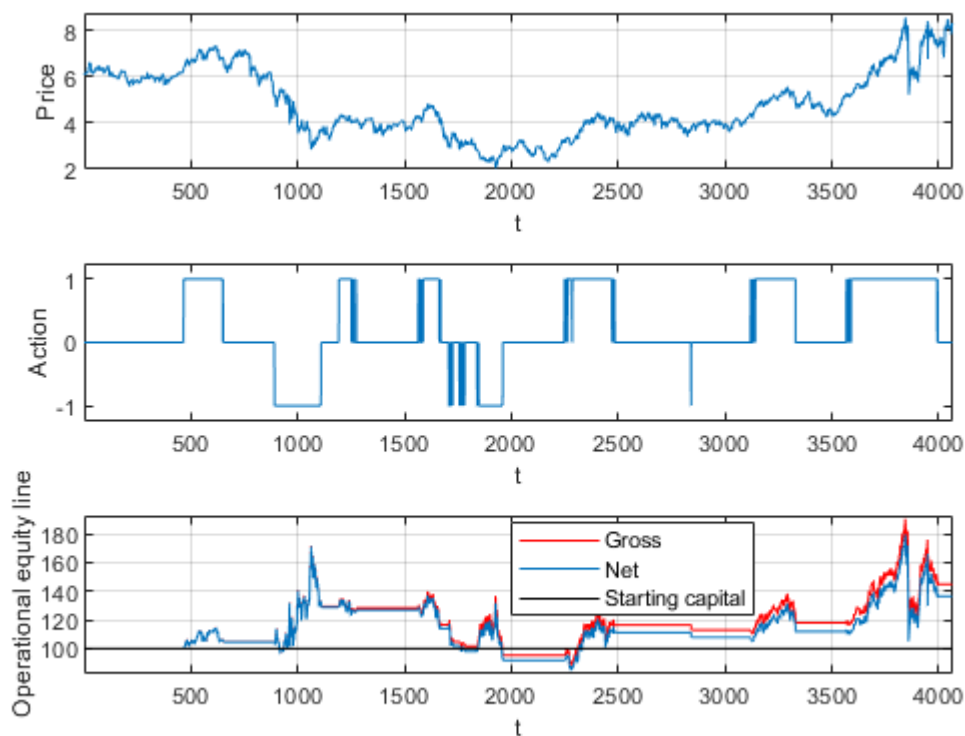


Figura 4.8: Enel S.p.A. $N = 3, L = 21, \varepsilon = 15\%, SR$ (input: indicatori AT).

4.4 Fiat Chrysler Automobiles

N=2			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	71.81	69.52	-2.03	-2.23	83.94	83.94	25.86	3.35
SR	5	15%	246.13	229.73	5.74	5.29	76.68	76.17	89.10	7.13
SR	21	5%	39.96	37.30	-5.53	-5.93	72.02	70.98	41.14	7.13
SR	21	15%	167.46	153.85	3.25	2.71	73.58	73.06	90.16	8.74
SOR	5	5%	47.50	43.75	-4.51	-5.00	80.83	80.83	15.21	8.50
SOR	5	15%	231.66	212.57	5.35	4.79	79.79	79.27	97.54	8.87
SOR	21	5%	330.04	309.57	7.68	7.26	88.60	88.60	99.98	6.63
SOR	21	15%	241.79	229.92	5.63	5.30	92.75	92.23	100.00	5.21

N=5			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	67.77	62.97	-2.39	-2.83	84.46	83.94	9.95	7.57
SR	5	15%	183.70	173.91	3.85	3.49	76.68	76.68	80.77	5.65
SR	21	5%	27.96	26.44	-7.60	-7.92	75.13	75.13	26.15	5.77
SR	21	15%	149.98	136.50	2.55	1.95	78.76	78.24	94.93	9.74
SOR	5	5%	61.59	57.45	-2.96	-3.38	83.94	83.42	8.47	7.20
SOR	5	15%	112.85	102.06	0.75	0.13	82.38	82.38	29.18	10.36
SOR	21	5%	98.83	92.13	-0.07	-0.51	85.49	84.97	94.98	7.26
SOR	21	15%	186.28	181.85	3.94	3.78	94.82	94.30	99.90	2.48

Tabella 4.7: Riepilogo delle statistiche relative al titolo Fiat Chrysler Automobiles (input: rendimenti logaritmici)

La Tabella 4.7 riporta le statistiche relative alle performance della prima applicazione dell'algoritmo al titolo Fiat Chrysler Automobiles. Da una prima analisi si può subito vedere che l'algoritmo opera decisamente meglio quando viene utilizzato un parametro $\epsilon = 15\%$, dimostrando quindi l'utilità dell'azione di tipo esplorativo. La Figura 4.9 rappresenta le $K = 500$ equity lines prodotte con i settaggi $N = 2, L = 5, \epsilon = 15\%, SR$. Queste si presentano compatte nella loro distribuzione, salvo qualche sporadico caso. Un grafico come questo comunica una certa semplicità di apprendimento del comportamento del titolo, il che giustifica i buoni risultati generali ottenuti. Un'altra caratteristica che si può rilevare osservando la Tabella 4.7 è che il sistema di trading riesce ad ottenere performance mediamente migliori con $N = 2$ rispetto a quelle che ottiene con $N = 5$. Per quanto riguarda le funzioni di *reward*, non c'è una prevalenza dell'una sull'altra in termini di risultati: entrambe le funzioni presentano infatti configurazioni con un rendimento positivo e altre invece in perdita, rendendo difficile definire quale delle due operi meglio nel contesto del presente titolo.

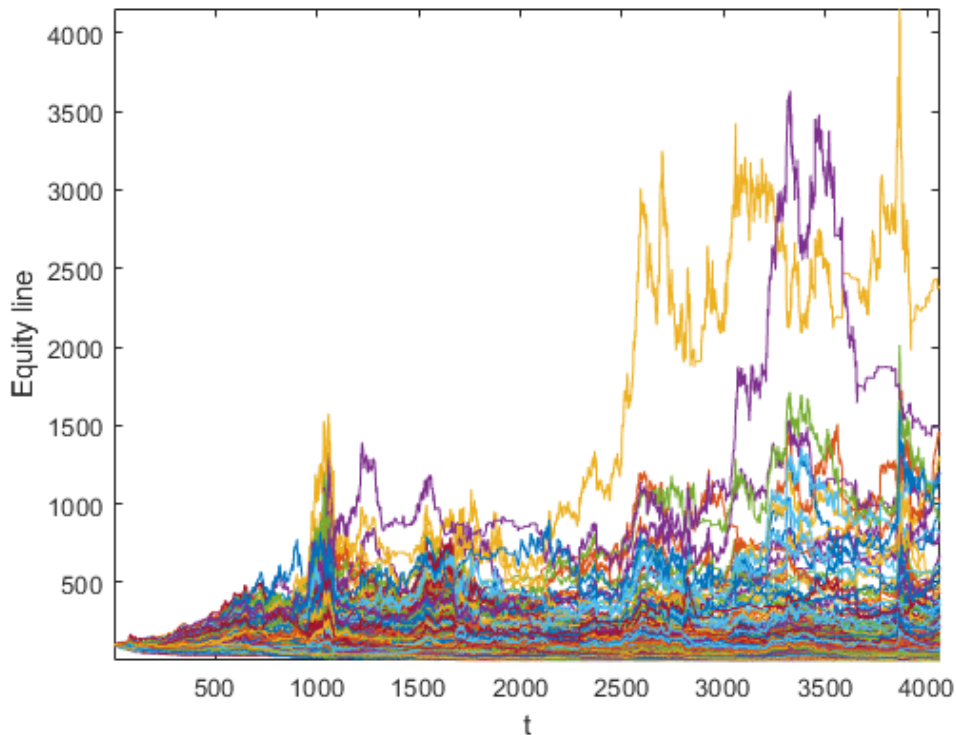


Figura 4.9: *Equity lines* totali relative al titolo Fiat Chrysler Automobiles $N = 2, L = 21, \varepsilon = 5\%$, *SOR* (input: rendimenti logaritmici)

N=3			Capitale Finale		Rendimento annuo		Incremento del capitale			
Reward	L	ε	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	187.85	176.68	3.99	3.59	88.60	88.08	99.70	6.33
SR	5	15%	246.50	227.64	5.76	5.23	78.24	76.68	99.95	8.25
SR	21	5%	125.42	119.34	1.41	1.10	75.13	74.61	70.15	5.15
SR	21	15%	279.13	256.14	6.57	6.01	73.06	72.54	100.00	8.87
SOR	5	5%	108.94	100.47	0.53	0.03	82.38	81.87	70.44	8.37
SOR	5	15%	213.32	192.57	4.81	4.15	77.72	77.72	91.02	10.61
SOR	21	5%	125.83	114.51	1.44	0.84	72.54	72.02	46.76	9.74
SOR	21	15%	122.82	107.69	1.28	0.46	82.38	81.35	96.11	13.58

Tabella 4.8: Riepilogo delle statistiche relative al titolo Fiat Chrysler Automobiles (input: indicatori AT).

Esattamente come per gli altri titoli, si riportano le statistiche relative alla seconda applicazione dell'algoritmo, contenute nella Tabella 4.8. I numeri che si possono leggere nella Tabella 4.8 parlano da soli: non esiste una sola configurazione che abbia prodotto un capitale finale inferiore al livello del capitale iniziale. Che fosse una serie storica facile da apprendere lo si era già detto in fase di presentazione dei risultati ottenuti dalla prima applicazione dell'algoritmo e qui ne abbiamo la chiara conferma. La tendenza qui osservabile è una redditività e un'operatività migliori nel caso venga utilizzato lo Sharpe ratio come funzione

di ricompensa. Inoltre, anche in questo caso le configurazioni con un'incidenza dell'azione esplorativa più alta sono quelle con gli output migliori. Non essendo una serie particolarmente difficoltosa da apprendere per l'algoritmo, non ci sono particolari elementi da evidenziare oltre a quanto già esposto. Interessante però è il confronto tra i risultati delle due applicazioni, osservabile nella Tabella 4.9.

			Rendimento netto annuo			Statistica %			Statistica #		
Reward	L	ε	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione
SR	5	5%	-2.23	-2.83	3.59	25.86	9.95	99.70	3.35	7.57	6.33
SR	5	15%	5.29	3.49	5.23	89.10	80.77	99.95	7.13	5.65	8.25
SR	21	5%	-5.93	-7.92	1.10	41.14	26.15	70.15	7.13	5.77	5.15
SR	21	15%	2.71	1.95	6.01	90.16	94.93	100.00	8.74	9.74	8.87
SOR	5	5%	-5.00	-3.38	0.03	15.21	8.47	70.44	8.50	7.20	8.37
SOR	5	15%	4.79	0.13	4.15	97.54	29.18	91.02	8.87	10.36	10.61
SOR	21	5%	7.26	-0.51	0.84	99.98	94.98	46.76	6.63	7.26	9.74
SOR	21	15%	5.30	3.78	0.46	100.00	99.90	96.11	5.21	2.48	13.58

Tabella 4.9: Confronto tra applicazioni dell'algoritmo al titolo Fiat Chrysler Automobiles.

L'utilizzo degli indicatori di Analisi Tecnica come input ha permesso di migliorare la redditività e l'operatività delle configurazioni che utilizzano un parametro $\varepsilon = 5\%$, come è possibile appurare dal confronto dei dati nella Tabella 4.9. Non solo, in alcuni casi la redditività di alcune delle migliori configurazioni della prima applicazione è addirittura aumentata, come ad esempio nel caso di $L = 21, \varepsilon = 5\%, SR$. Nonostante si possa ancora individuare una tendenza a performare meglio con un'incidenza maggiore dell'azione esplorativa, nella seconda applicazione il sistema è stato in grado di produrre output soddisfacenti anche con un parametro di selezione dell'azione esplorativa più basso. In definitiva, anche in questo caso l'impiego delle Medie Mobili e del RSI all'interno del vettore di stato ha permesso di risolvere il principale problema del sistema di trading applicato alla serie storica del titolo in questione – cioè la tendenza ad ottenere performance negative con un parametro ε più basso –, portando ad un miglioramento generalizzato e notevole nelle performance.

Vale la pena riportare alcuni settaggi sia relativamente alla prima applicazione dell'algoritmo che alla seconda, in modo da comprendere meglio le *policies* seguite dall'algoritmo. Per quanto riguarda la prima applicazione, si riporta con la Figura 4.10 la

configurazione $N = 2, L = 21, \varepsilon = 5\%, SOR$, che è la configurazione che ha prodotto il miglior risultato in termini di redditività e preservazione del capitale.

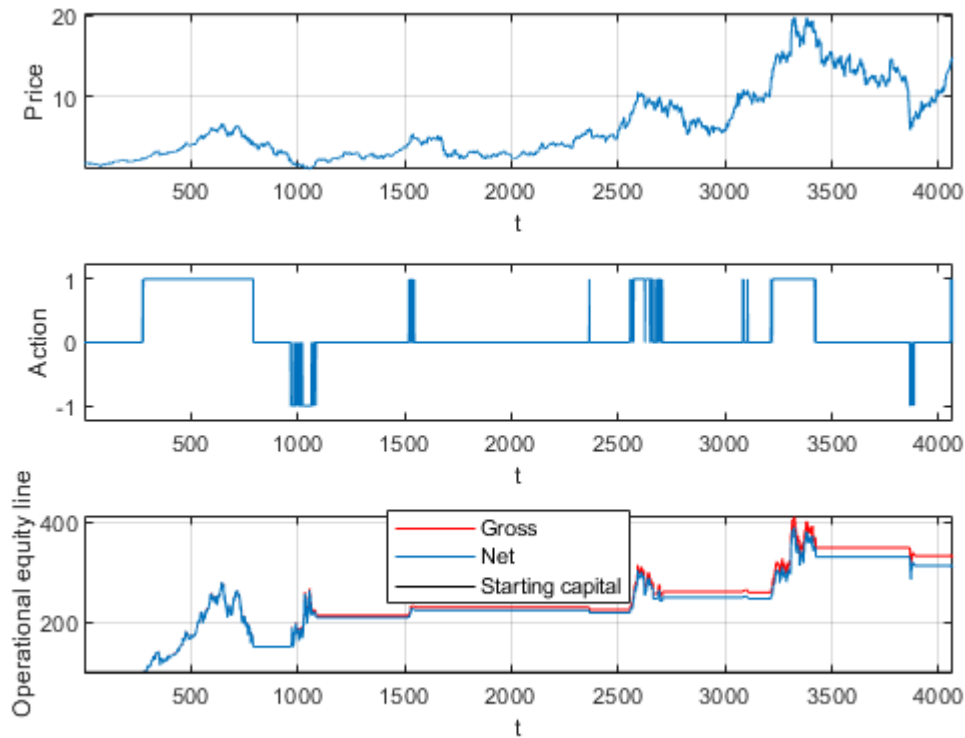


Figura 4.10: Fiat Chrysler Automobiles $N = 2, L = 21, \varepsilon = 5\%, SOR$ (input: rendimenti logaritmici).

Nel caso in esame, la performance è una delle migliori tra tutte quelle ottenute con il presente *pool* di titoli. Il capitale finale netto ottenuto è pari a 309.57€, oltre il triplo rispetto al capitale iniziale, e il rendimento netto medio annuo è 7.26%. L'operatività media è pari a $\# = 6.63$, il che risulta interessante dato che non vengono effettuate troppe transazioni che rischierebbero di andare a gravare il capitale finale con costi di transazione elevati. La statistica % è altrettanto buona, attestandosi ad un livello pari al 99.98%. Accanto a questo risultato, vale la pena analizzare anche l'indice di incremento del capitale netto, pari a 88.60%. Tale indice dimostra quanto la performance ottenuta sia consistente dato che nell'oltre 88% dei casi il livello del capitale risulta essere maggiore o uguale al livello capitale del mese precedente. Si tratta di un risultato importante poiché dimostra che il capitale cresce nel tempo, anche se nel primo periodo – all'incirca fino all'osservazione $t = 1000$ – oscilla notevolmente. Una crescita come questa significa che l'algoritmo non ha prodotto il capitale finale grazie a transazioni fortuite eseguite al momento giusto, si tratta di una costruzione di

valore graduale ma efficace, a dimostrazione del fatto che il sistema di trading ha compreso l'andamento del titolo e adattato di conseguenza le proprie scelte di investimento.

Per quanto riguarda la seconda applicazione del sistema di trading la configurazione che ha prodotto il miglior risultato tra tutti è $N = 3, L = 21, \varepsilon = 15\%, SR$, la cui rappresentazione grafica è riportata nella Figura 4.11. Oltre ad un capitale finale elevato e ad un rendimento netto medio annuo del 6.01%, questa configurazione ha una persistenza dell'*equity line* al di sopra del capitale iniziale del 100%. In generale, la statistica % è buona nella quasi totalità dei casi, con un'unica eccezione in cui questa non è propriamente buona – si veda la configurazione $N = 3, L = 21, \varepsilon = 5\%, SOR$ nella Tabella 4.8 –. Questa configurazione non presenta caratteristiche peculiari da richiedere ulteriori analisi, si tratta di un buon risultato dal punto di vista della redditività e della permanenza dell'*equity line* al di sopra del livello del capitale iniziale e in definitiva si può dire che il titolo Fiat Chrysler Automobiles è il titolo che ha permesso di ottenere con il sistema di trading i risultati più soddisfacenti.

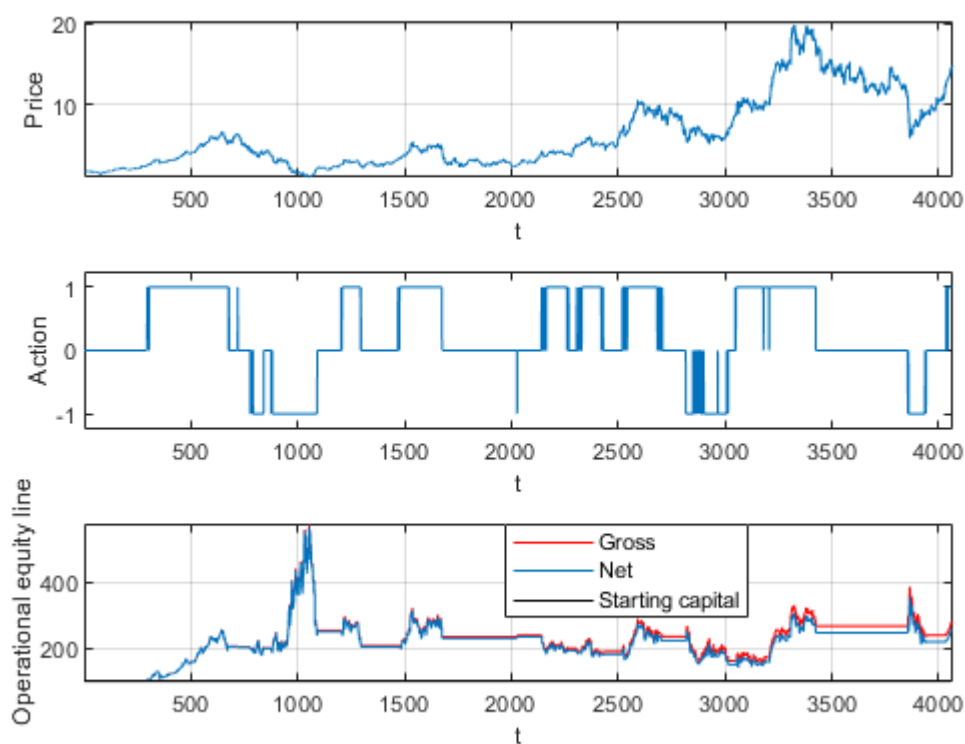


Figura 4.11: Fiat Chrysler Automobiles $N = 3, L = 21, \varepsilon = 15\%, SR$ (input: indicatori AT).

4.5 Intesa Sanpaolo S.p.A.

N=2			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	73.48	70.38	-1.89	-2.15	87.56	87.56	10.70	4.46
SR	5	15%	74.70	66.73	-1.79	-2.48	68.91	67.36	28.62	11.66
SR	21	5%	36.34	33.77	-6.08	-6.51	70.47	69.43	28.72	7.56
SR	21	15%	49.39	43.49	-4.28	-5.03	70.47	68.91	37.65	13.15
SOR	5	5%	51.57	47.51	-4.02	-4.51	78.76	78.76	10.51	8.43
SOR	5	15%	69.79	63.08	-2.21	-2.82	80.31	80.31	14.84	10.42
SOR	21	5%	86.17	80.17	-0.92	-1.36	82.38	80.83	42.67	7.44
SOR	21	15%	107.35	100.49	0.44	0.03	92.75	92.23	65.99	6.82

N=5			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	86.28	80.77	-0.91	-1.32	86.53	86.53	50.46	6.83
SR	5	15%	60.05	52.75	-3.12	-3.89	70.98	69.95	10.15	13.40
SR	21	5%	55.76	51.13	-3.56	-4.08	74.61	74.09	22.09	8.94
SR	21	15%	50.55	44.61	-4.15	-4.89	74.61	73.06	11.67	12.91
SOR	5	5%	149.46	141.80	2.52	2.19	84.97	84.46	71.85	5.46
SOR	5	15%	100.74	93.83	0.05	-0.39	87.56	87.05	69.61	7.32
SOR	21	5%	50.73	47.77	-4.12	-4.48	87.05	86.53	23.39	6.21
SOR	21	15%	91.72	88.15	-0.53	-0.78	96.37	96.37	7.26	4.10

Tabella 4.10: Riepilogo delle statistiche relative al titolo Intesa Sanpaolo S.p.A. (input: rendimenti logaritmici).

La Tabella 4.10 riporta il riepilogo delle statistiche relative alla prima applicazione del sistema di trading applicato al titolo Intesa Sanpaolo S.p.A. Le performance dell'algoritmo in questo contesto non sono buone. Nella tabella si può vedere che in un unico caso l'algoritmo riesce ad ottenere un capitale finale netto ed un rendimento netto annuo positivi, il che non è sufficiente a classificare la performance generale come accettabile. Tendenzialmente, il sistema di trading funziona leggermente meglio quando utilizza il Sortino Ratio come funzione di ricompensa. In questo caso, per proseguire l'analisi, è necessario riportare il grafico dell'andamento del prezzo del titolo; si riporta quindi il grafico di una delle configurazioni nella Figura 4.12, precisamente $N = 5, L = 5, \epsilon = 5\%, SOR$. Osservando la serie dei prezzi del titolo *ex-post* nel primo pannello della figura, si osserva che la serie è caratterizzata da continui *break* strutturali e una volatilità notevole. Evidentemente, un comportamento del genere è difficile da inquadrare per l'algoritmo, pertanto le performance negative sono più che comprensibili. In un ambiente come questo, dove i cambi di direzione del prezzo sono rapidi e consistenti, il sistema di trading non è in grado di apprendere il comportamento della serie storica alla velocità con cui i prezzi si muovono; il risultato è un

comportamento costantemente “mal adattato” che non riesce a produrre performance brillanti.

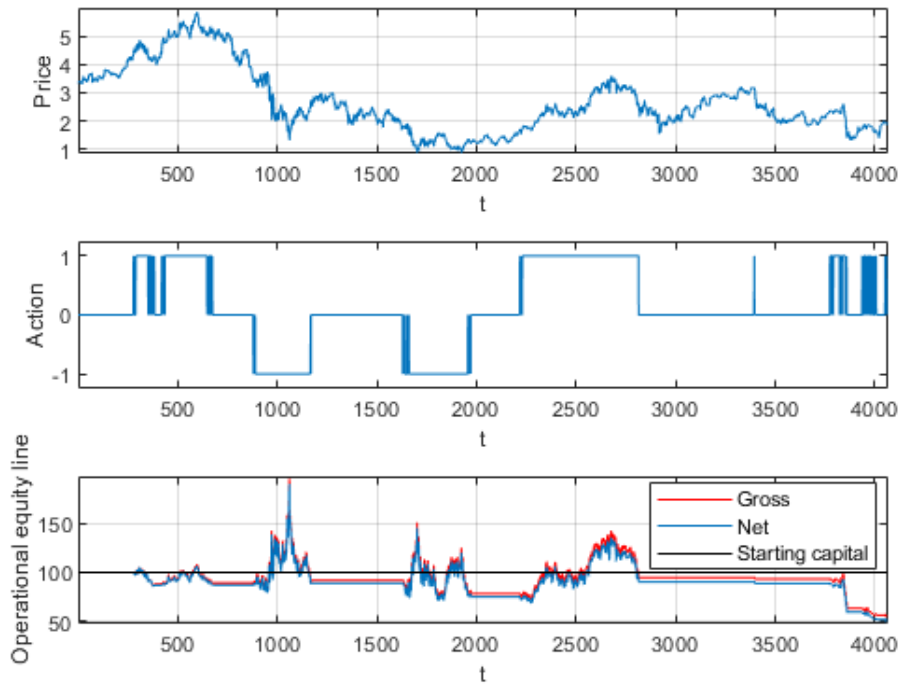


Figura 4.12: Intesa Sanpaolo S.p.A. $N = 5, L = 5, \varepsilon = 5\%$, *SOR* (input: rendimenti logaritmici).

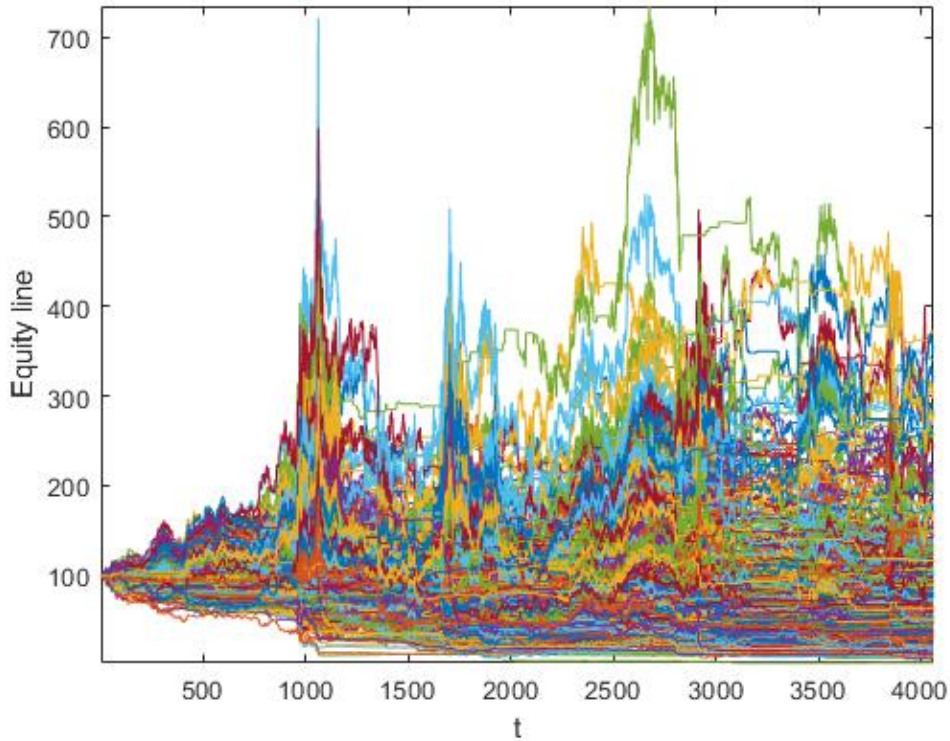


Figura 4.13: *Equity lines* totali relative al titolo Intesa Sanpaolo S.p.A. $N = 5, L = 5, \varepsilon = 5\%$, *SOR* (input: rendimenti logaritmici).

A ulteriore dimostrazione di quanto già detto, si possono osservare nella Figura 4.13 le $K = 500$ *equity lines* prodotte dall'algoritmo. In questo grafico si rileva una tendenza al ribasso e una certa dispersione delle diverse linee, indice del fatto che in fase di operatività simulata le *equity lines* che vengono prodotte sono diverse ad ogni k – *esima* iterazione. Questo accade proprio per i motivi elencati sopra, il sistema di trading si trova in difficoltà a causa dei continui e repentini cambi di direzione del prezzo producendo ad ogni iterazione un output diverso. Il risultato è un grafico come quello della Figura 4.13 in cui le linee non seguono lo stesso andamento.

N=3			Capitale Finale		Rendimento annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	108.18	102.86	0.49	0.18	91.19	90.67	83.07	5.21
SR	5	15%	84.55	75.43	-1.04	-1.73	72.54	70.47	41.84	11.78
SR	21	5%	30.18	26.86	-7.16	-7.83	75.13	74.09	9.94	12.03
SR	21	15%	98.17	88.66	-0.11	-0.74	75.65	75.13	94.95	10.54
SOR	5	5%	148.04	139.23	2.46	2.07	82.90	81.87	57.49	6.33
SOR	5	15%	69.34	60.26	-2.24	-3.09	75.65	74.61	25.79	14.51
SOR	21	5%	48.63	44.45	-4.37	-4.90	75.65	74.61	17.06	9.30
SOR	21	15%	52.55	44.43	-3.91	-4.91	84.46	83.42	41.64	17.37

Tabella 4.11: Riepilogo delle statistiche relative al titolo Intesa Sanpaolo S.p.A. (input: indicatori AT).

Per quanto riguarda la seconda applicazione dell'algoritmo, tutte le statistiche relative ottenute relativamente al titolo Intesa Sanpaolo S.p.A. sono contenute nella Tabella 4.11. Anche in questo caso, le performance del sistema di trading non sono buone. La redditività è quasi sempre negativa, per quanto riguarda invece la preservazione del capitale la statistica % presenta dei valori buoni per alcune configurazioni mentre per altre risulta essere insufficiente. Non si rileva una particolare tendenza ad operare meglio con una funzione di *reward* rispetto all'altra; con lo Sharpe ratio il sistema di trading ottiene il peggior risultato tra tutti – si veda nella Tabella 4.11 la configurazione $N = 3, L = 21, \epsilon = 5\%, SR$ –, con il Sortino ratio ottiene invece una performance accettabile dal punto di vista della redditività e della preservazione del capitale – precisamente la configurazione $N = 3, L = 5, \epsilon = 5\%, SOR$ – in contrasto con gli altri risultati in perdita. Questi risultati contrastanti non permettono di dare un giudizio su quale delle due funzioni di ricompensa sia stata più utile.

Passando ad un confronto delle performance generali delle due applicazioni del sistema di trading, non ci sono particolari differenze tra le due da rilevare. In termini di performance i due algoritmi in questo caso sono equiparabili e nemmeno l'utilizzo di input

più sofisticati quali gli indicatori AT ha permesso di migliorare la redditività e l'operatività del sistema di trading nell'ambito del presente titolo. Per completezza, si riporta il confronto tra le due performance contenuto nella Tabella 4.12. Come si può notare, in alcuni casi la redditività migliora leggermente come nel caso della configurazione $L = 5, \varepsilon = 5\%$, SR ma in altri peggiora, come per il caso $L = 21, \varepsilon = 5\%$, SR . Per quanto riguarda invece il confronto dei valori della statistica % valgono le medesime considerazioni fatte per il rendimento netto medio annuo, dato che ci sono situazioni in cui questa si attesta oltre la soglia del 50% e altri in cui peggiora diminuendo drasticamente rispetto al valore ottenuto con la prima applicazione. Al termine dell'analisi si può dire che Intesa Sanpaolo sia il titolo più complesso da comprendere date le caratteristiche della serie storica, e osservandone l'andamento una volta terminato il periodo di investimento si arriva a comprendere una performance generale insufficiente come questa.

			Rendimento netto annuo			Statistica %			Statistica #		
Reward	L	ε	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione
SR	5	5%	-2.15	-1.32	0.18	10.70	50.46	83.07	4.46	6.83	5.21
SR	5	15%	-2.48	-3.89	-1.73	28.62	10.15	41.84	11.66	13.40	11.78
SR	21	5%	-6.51	-4.08	-7.83	28.72	22.09	9.94	7.56	8.94	12.03
SR	21	15%	-5.03	-4.89	-0.74	37.65	11.67	94.95	13.15	12.91	10.54
SOR	5	5%	-4.51	2.19	2.07	10.51	71.85	57.49	8.43	5.46	6.33
SOR	5	15%	-2.82	-0.39	-3.09	14.84	69.61	25.79	10.42	7.32	14.51
SOR	21	5%	-1.36	-4.48	-4.90	42.67	23.39	17.06	7.44	6.21	9.30
SOR	21	15%	0.03	-0.78	-4.91	65.99	7.26	41.64	6.82	4.10	17.37

Tabella 4.12: Confronto tra le applicazioni dell'algorithmo al titolo Intesa Sanpaolo S.p.A.

Date le performance negative sia che si utilizzino i rendimenti logaritmici come input che gli indicatori AT, può essere utile analizzare nel dettaglio qualche configurazione di entrambe le applicazioni per trarne qualche osservazione aggiuntiva.

Per quanto riguarda la prima applicazione del sistema di trading, si riportano nella Figura 4.14 i grafici relativi alla configurazione $N = 5, L = 5, \varepsilon = 5\%$, SOR . Si tratta del settaggio che ha prodotto il miglior risultato in termini di redditività e permanenza del capitale al di sopra del livello del capitale iniziale. Questa particolare performance può essere suddivisa in diversi momenti: ci sono periodi in cui l'algorithmo identifica un trend in corso e adatta le proprie scelte di conseguenza, altri in cui la tendenza dei prezzi non risulta così facilmente identificabile dove il sistema di trading preferisce rimanere fuori dal mercato. Sfruttando i momenti in cui il trend è più marcato, il sistema effettua qualche transazione con un buon

tempismo, come ad esempio attorno all'osservazione $t = 2200$. Ci sono però alcuni momenti in cui l'algoritmo riesce a cogliere l'andamento del prezzo e ad agire conseguentemente ma perdendo immediatamente quando guadagnato a causa di un crollo verticale del prezzo. Si osservi ad esempio il picco della *equity line* all'altezza dell'istante $t = 1000$: qui l'algoritmo intercetta un trend ribassista mette a segno alcune operazioni in grado di spingere in alto il livello del capitale investito; l'algoritmo però non risulta pronto al successivo crollo del prezzo, causando quindi la perdita di tutto quanto guadagnato con le precedenti transazioni. Una performance come questa è da definirsi accettabile, con un rendimento netto medio annuo del 2.19% ed una statistica $\% = 71.85$. L'operatività media annua è pari a $\# = 5.46$, in questo caso un numero di operazioni come questo risulta essere adeguato visto il risultato prodotto; ciò non toglie che l'algoritmo avrebbe potuto eseguire qualche transazione in più o in meno e ottenere comunque un risultato soddisfacente, pertanto non si può dare una valutazione oggettiva di tale valore.

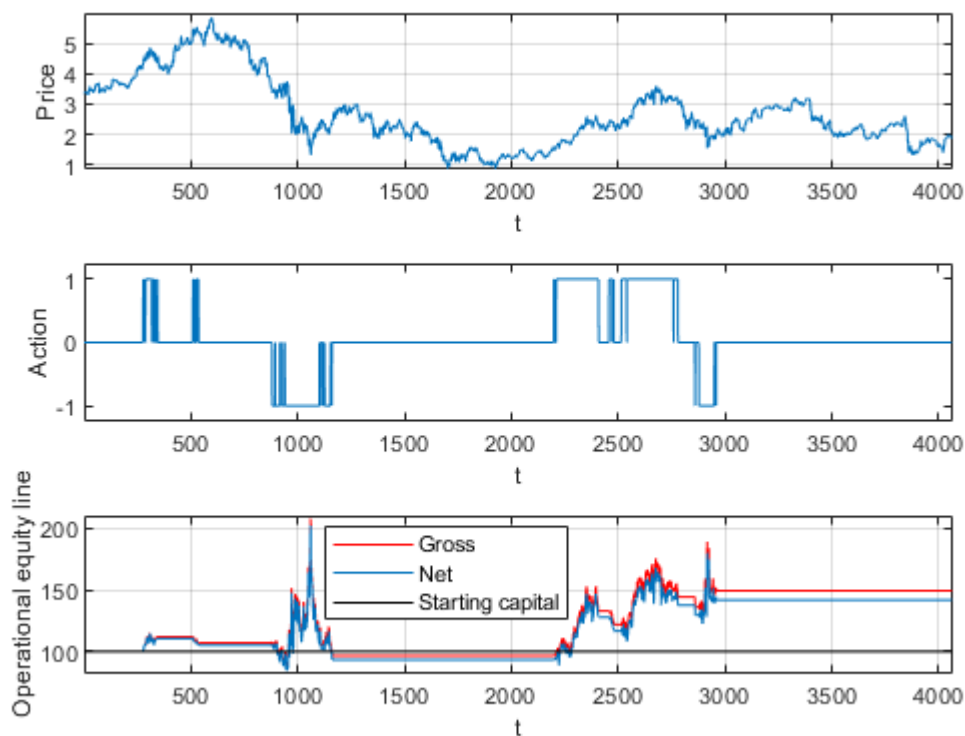


Figura 4.14: Intesa Sanpaolo S.p.A. $N = 5, L = 5, \varepsilon = 5\%$, *SOR* (input: rendimenti logaritmici).

Per quanto riguarda la seconda applicazione dell'algoritmo, c'è una particolare configurazione che attira l'attenzione nella Tabella 4.11: si tratta del settaggio $N = 3, L = 21, \varepsilon = 15\%$, *SR*, il quale riporta un rendimento netto negativo con una permanenza del capitale al di sopra del

livello del capitale iniziale pari al 94.95%. Ciò significa che il sistema è stato in grado di mantenere un buon livello del capitale per quasi tutto l'orizzonte temporale osservato e che poi verso la fine sia collassato, terminando la corsa con un capitale finale netto inferiore al capitale iniziale. Per capire meglio cosa è successo vengono presentate nella Figura 4.15 le rappresentazioni grafiche della presente configurazione.

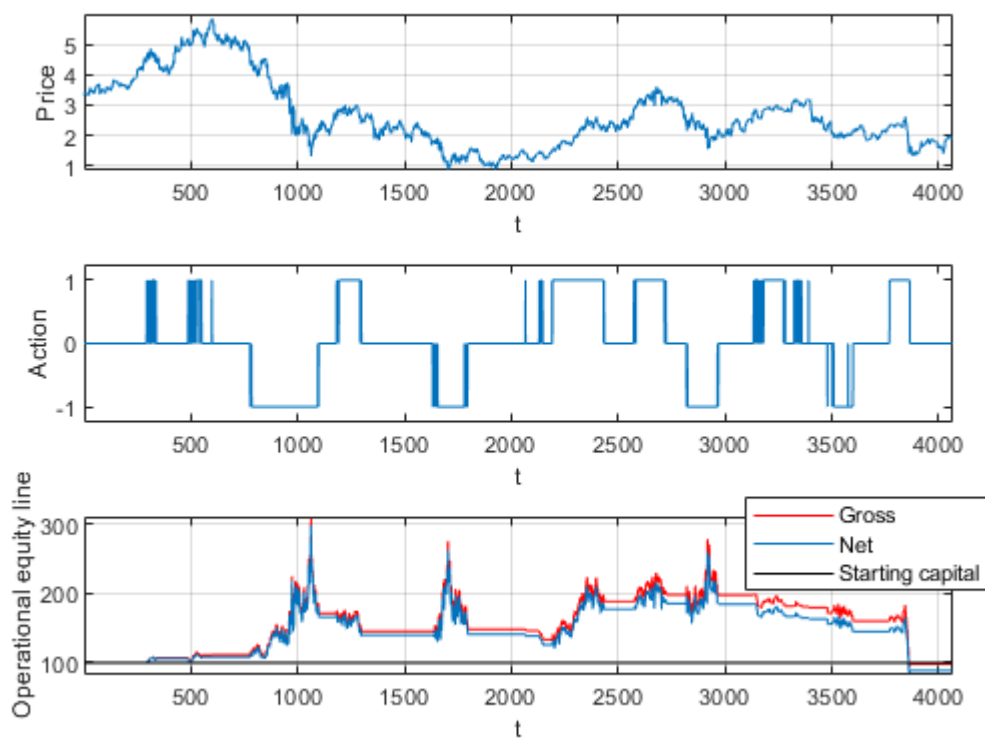


Figura 4.15: Intesa Sanpaolo S.p.A. $N = 3, L = 21, \varepsilon = 15\%, SR$ (input: indicatori AT).

Il terzo pannello della figura conferma quanto ipotizzato poco sopra: il sistema di trading produce effettivamente un livello del capitale che permane al di sopra del capitale iniziale per tutta la durata dell'orizzonte temporale. Appena poco prima del termine delle osservazioni il prezzo subisce un crollo verticale, il che porta l'*equity line* a scendere al di sotto del livello del capitale iniziale. Nonostante il capitale finale ottenuto sia inferiore al livello iniziale investito, questa rimane una performance accettabile e interessante, visto e considerato l'andamento dell'*equity line* durante tutto il periodo osservato.

4.6 Telecom Italia S.p.A.

N=2			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	97.55	90.98	-0.15	-0.58	72.02	72.02	40.65	7.19
SR	5	15%	46.38	40.11	-4.65	-5.51	62.69	61.14	28.49	15.01
SR	21	5%	38.05	33.58	-5.82	-6.54	69.95	69.95	26.33	12.90
SR	21	15%	47.66	43.60	-4.49	-5.02	77.20	76.68	47.02	9.18
SOR	5	5%	132.08	123.19	1.74	1.30	75.65	75.13	86.54	7.19
SOR	5	15%	67.89	58.71	-2.37	-3.25	76.17	75.13	41.83	15.01
SOR	21	5%	72.63	66.77	-1.96	-2.47	83.42	82.90	10.36	8.68
SOR	21	15%	87.76	86.60	-0.81	-0.89	96.89	96.89	20.08	1.36

N=5			Capitale Finale		Rendimento medio annuo		Incremento del capitale			
Reward	L	ϵ	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	81.02	74.56	-1.30	-1.80	75.65	75.13	35.85	8.56
SR	5	15%	59.81	52.72	-3.14	-3.90	66.84	65.80	28.86	13.03
SR	21	5%	27.58	24.43	-7.68	-8.37	68.91	67.36	31.15	12.53
SR	21	15%	56.97	51.14	-3.43	-4.08	76.68	75.65	34.23	11.17
SOR	5	5%	94.41	86.48	-0.36	-0.90	71.50	71.50	42.06	9.06
SOR	5	15%	43.78	36.92	-5.00	-6.00	76.68	75.65	1.53	17.62
SOR	21	5%	62.16	56.94	-2.91	-3.43	86.01	85.49	7.68	9.06
SOR	21	15%	84.19	82.58	-1.06	-1.18	96.37	95.85	5.66	1.99

Tabella 4.13: Riepilogo delle statistiche relative al titolo Telecom Italia S.p.A. (input: rendimenti logaritmici)

L'ultimo titolo del *pool* da analizzare è Telecom Italia S.p.A., di cui si trovano nella Tabella 4.13 le statistiche relative ai risultati ottenuti dalla prima applicazione dell'algoritmo.

Accanto al titolo Intesa Sanpaolo S.p.A., Telecom Italia S.p.A. si presta ad essere uno dei titoli con le performance peggiori tra i cinque analizzati. Come si può notare dall'ultima colonna della tabella, la quale riporta l'operatività media annua, l'algoritmo tende ad eseguire un numero di transazioni particolarmente elevato. Va precisato che un numero di transazioni elevato è collegato anche un valore di ϵ elevato: se la percentuale di selezione dell'azione esplorativa ad ogni step temporale è alta esiste la possibilità che il numero di transazioni risulti a sua volta elevato. Se guardiamo le configurazioni con il numero medio annuo di transazioni eseguite si vede che queste sono configurazioni che utilizzano, per l'appunto, una percentuale di selezione dell'azione esplorativa pari a $\epsilon = 15\%$. Tuttavia, non è sempre vero che ad un valore di ϵ elevato corrisponde un'operatività elevata, come dimostra la statistica $\#$ relativa alla configurazione $N = 5, L = 21, \epsilon = 15\%, SOR$. La serie dei prezzi di Telecom Italia S.p.A. – si veda il primo pannello della Figura 4.16 – presenta una peculiarità, ossia una

volatilità che si riduce dopo il primo trend al ribasso individuabile. Il fatto che i prezzi oscillino in maniera meno evidente incide sull'efficacia delle due funzioni di *reward* utilizzate; se la volatilità diminuisce, le osservazioni dei rendimenti di discostano meno dalla media, ed essendo sia lo Sharpe ratio che il Sortino ratio due funzioni che mappano gli scostamenti dalla media (sia positivi che negativi nel primo caso, solo negativi nel secondo), esse possono rivelarsi poco utili se questi si verificano in maniera meno marcata. In casi come questo sarebbe più opportuno utilizzare funzioni di ricompensa basate, ad esempio, sul massimo *drawdown* al posto di funzioni basate sulla deviazione standard. Utilizzando un una funzione basata sul massimo declino della serie da un punto di massimo al successivo punto di minimo in un intervallo temporale di ampiezza L , l'algoritmo potrebbe essere in grado di valutare meglio la volatilità dei prezzi, e di conseguenza dei rendimenti, in un momento in cui questi seguono un andamento piuttosto piatto.

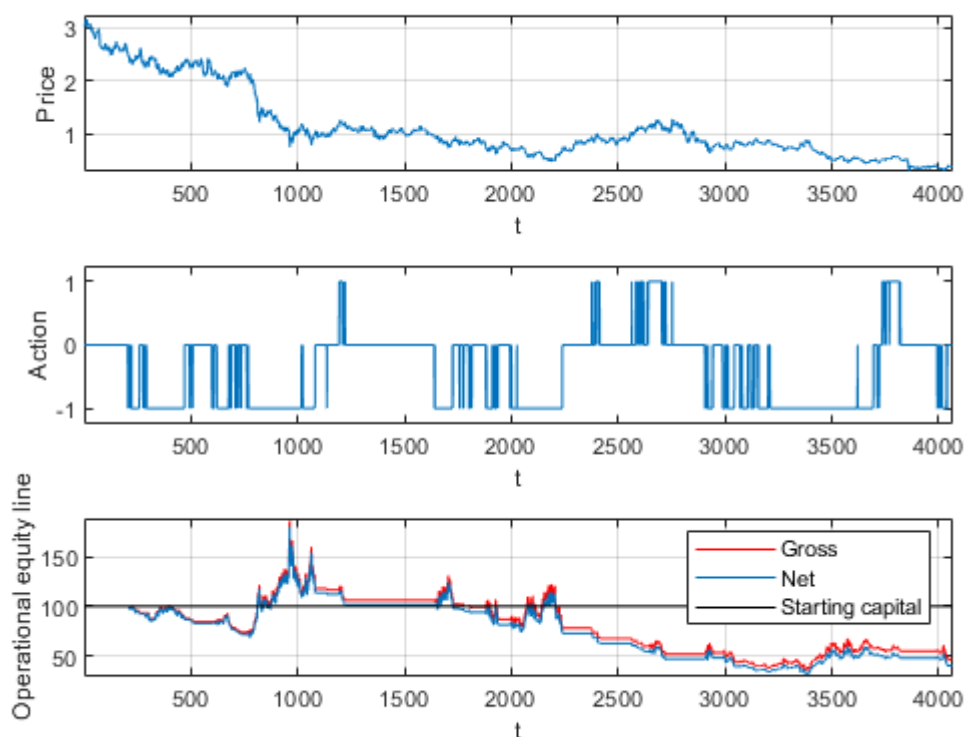


Figura 4.16: Telecom Italia S.p.A. $N = 2, L = 5, \varepsilon = 15\%, SR$ (input: rendimenti logaritmici).

La tendenza che comunque si può notare confrontando i risultati ottenuti con le due diverse funzioni di ricompensa è una preferenza per le azioni *greedy* nei casi in cui viene utilizzato un parametro $L = 5$, di fatto le performance peggiorano con un parametro $\varepsilon = 15\%$ a parità di L . Al contrario, utilizzando un parametro $L = 21$ la situazione si ribalta e l'algoritmo

ottiene risultati migliori con una incidenza dell'azione esplorativa più alta. L'inversione di tendenza è probabilmente dovuta all'utilizzo di un periodo più ampio per il calcolo delle funzioni di *reward*, in questo modo l'algoritmo è in grado di valutare in modo più completo l'andamento della serie dato che la volatilità – che abbiamo visto essere ridotta per gran parte dell'orizzonte temporale – risalta di più. A dimostrazione delle performance negative ottenute dall'algoritmo applicato al presente titolo, si possono osservare nella Figura 4.17 le $K = 500$ *equity lines* prodotte: queste si presentano molto disperse e con una tendenza al ribasso, comportamento già osservato con la precedente analisi del titolo Intesa Sanpaolo S.p.A.

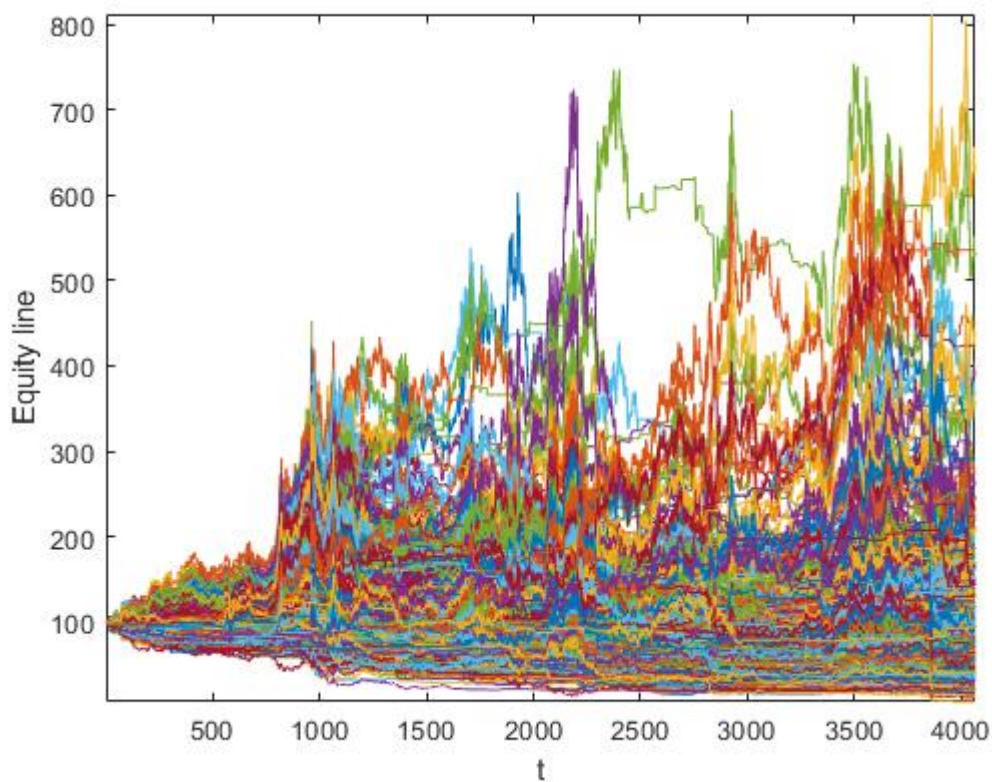


Figura 4.17: *Equity lines* totali del titolo Telecom Italia S.p.A. $N = 2, L = 5, \varepsilon = 5\%, SOR$ (input: rendimenti logaritmici).

Per quanto riguarda la seconda applicazione dell'algoritmo, di cui si ritrova un riepilogo nella Tabella 4.14, le performance non sono buone. Anche in questo caso si rileva una preferenza per le azioni di tipo *greedy* e una redditività migliore utilizzando lo Sharpe ratio come funzione di ricompensa. Le percentuali di permanenza della *equity line* al di sopra del capitale iniziale non sono buone tranne in un caso, che tra l'altro è quello che ha ottenuto il miglior risultato tra tutti. L'operatività media tende ad essere leggermente elevata ma come

si può osservare è in corrispondenza delle configurazioni con $\varepsilon = 15\%$ che l'operatività media è più alta, un comportamento normale per i motivi già elencati precedentemente. In questo caso, le configurazioni che utilizzano un parametro ε più elevato sono anche quelle che ottengono i risultati peggiori in termini sia di redditività che di preservazione del capitale. In generale, la statistica % non è buona tranne in un caso, il che non è sufficiente a caratterizzare la performance generale come sufficiente. Anche in questo caso le $K = 500$ equity lines si presentano disperse e tendenti al ribasso, esattamente come nel caso della prima applicazione.

N=3			Capitale Finale		Rendimento annuo		Incremento del capitale			
Reward	L	ε	Lordo	Netto	Lordo%	Netto%	Lordo%	Netto%	%	#
SR	5	5%	96.75	89.39	-0.20	-0.69	82.38	81.87	47.43	8.19
SR	5	15%	51.90	45.74	-3.99	-4.74	68.39	66.32	39.50	13.02
SR	21	5%	114.46	111.75	0.84	0.69	83.42	82.90	72.73	2.48
SR	21	15%	56.49	49.92	-3.48	-4.22	73.58	73.58	49.67	12.78
SOR	5	5%	92.48	84.12	-0.48	-1.07	74.61	74.61	36.99	9.80
SOR	5	15%	51.00	44.96	-4.09	-4.84	71.50	70.98	25.28	13.02
SOR	21	5%	76.26	72.93	-1.67	-1.94	79.79	79.79	49.47	4.59
SOR	21	15%	59.13	53.14	-3.21	-3.85	85.49	85.49	38.15	11.04

Tabella 4.14: Riepilogo delle statistiche del titolo Telecom Italia S.p.A. (input: indicatori AT)

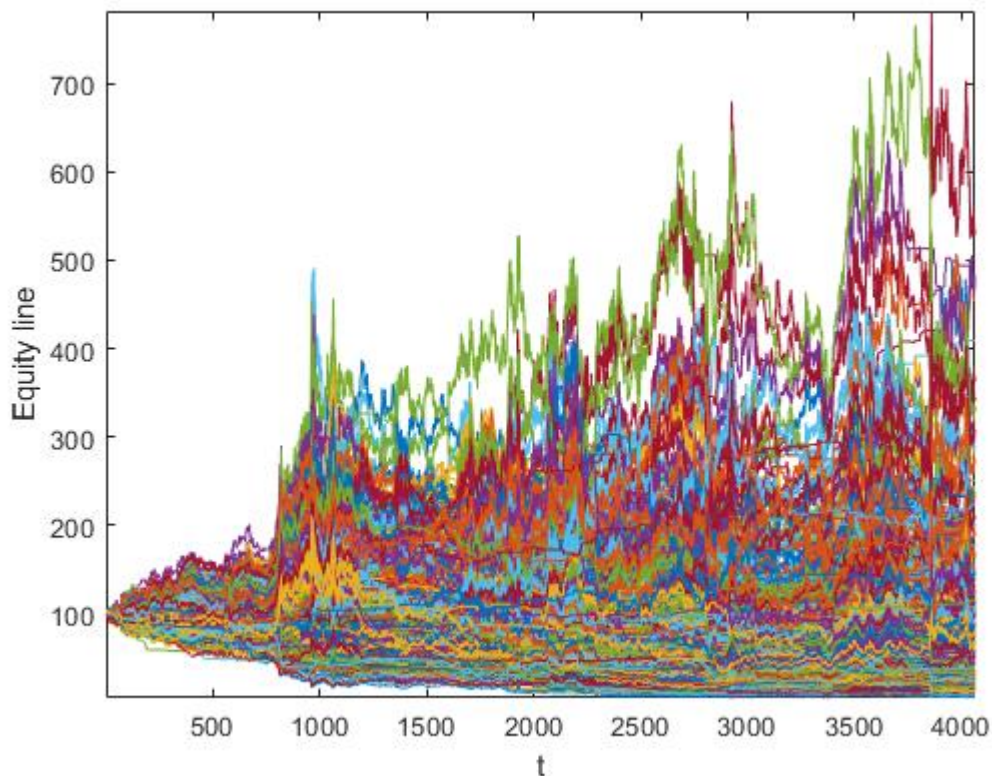


Figura 4.18: Equity lines totali del titolo Telecom Italia S.p.A. $N = 3, L = 21, \varepsilon = 5\%, SR$ (input: indicatori AT)

			Rendimento netto annuo			Statistica %			Statistica #		
Reward	L	ε	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione	Prima applicazione		Seconda applicazione
SR	5	5%	-0.58	-1.80	-0.69	40.65	35.85	47.43	7.19	8.56	8.19
SR	5	15%	-5.51	-3.90	-4.74	28.49	28.86	39.50	15.01	13.03	13.02
SR	21	5%	-6.54	-8.37	0.69	26.33	31.15	72.73	12.90	12.53	2.48
SR	21	15%	-5.02	-4.08	-4.22	47.02	34.23	49.67	9.18	11.17	12.78
SOR	5	5%	1.30	-0.90	-1.07	86.54	42.06	36.99	7.19	9.06	9.80
SOR	5	15%	-3.25	-6.00	-4.84	41.83	1.53	25.28	15.01	17.62	13.02
SOR	21	5%	-2.47	-3.43	-1.94	10.36	7.68	49.47	8.68	9.06	4.59
SOR	21	15%	-0.89	-1.18	-3.85	20.08	5.66	38.15	1.36	1.99	11.04

Tabella 4.15: Confronto tra le applicazioni dell'algorithmo al titolo Telecom Italia S.p.A.

Per quanto riguarda le performance dei due diversi algoritmi messe a confronto – si veda la Tabella 4.15 –, non è possibile rilevare un miglioramento degno di nota utilizzando il secondo. Un aspetto che sembra aver giovato leggermente dell'utilizzo di input diversi dai rendimenti logaritmici è l'operatività media: la prima applicazione dell'algorithmo ha prodotto risultati in cui l'operatività media risultava essere alle volte troppo elevata e altri in cui era estremamente ridotta. In questa seconda parte applicativa, invece, il numero medio di azioni è meglio distribuito, pur registrando ancora qualche caso di elevata operatività – come, ad esempio, per la configurazione $N = 3, L = 21, \varepsilon = 15\%, SR$ –. Quantità di transazioni come queste alle volte possono anche rivelarsi utili per ottenere dei profitti, in questo caso però eseguire tante operazioni non sembra essere stato utile per ottenere un rendimento positivo, pertanto l'operatività media elevata si traduce esclusivamente in maggiori costi di transazione applicati. Per il resto, non sembrano esserci notevoli differenze tra le performance dei due algoritmi, pertanto anche in questo caso l'utilizzo di indicatori AT come input non ha migliorato la redditività e l'operatività del sistema di trading.

Infine, per completare l'analisi vengono presentati alcuni risultati relativi alla prima applicazione e alla seconda, al fine di dare una valutazione più completa sull'operatività del sistema di trading automatico nel contesto del titolo Telecom Italia S.p.A. per quanto riguarda la prima applicazione, viene presentata nella Figura 4.19 la configurazione $N = 2, L = 5, \varepsilon = 5\%, SOR$. Fino a circa metà del periodo temporale osservato le performance sono buone, nonostante sia presente qualche periodo in cui il capitale cresce vertiginosamente per poi crollare subito dopo. Successivamente la performance non è altrettanto buona, di fatto le *equity lines* nel terzo pannello scendono notevolmente. In ultima battuta comunque c'è una ripresa che permette al sistema di trading di ottenere un capitale finale netto superiore al capitale iniziale pari a 123.19 €, con un rendimento netto medio

annuo del 1.30%. La statistica $\% = 86.54$ si qualifica come un buon risultato, mentre l'indice di preservazione del capitale con un valore pari al 75.13% – ricordando di non tenere conto dei periodi di prolungata inattività – non è propriamente un risultato brillante. L'operatività in questo caso risulta essere adeguata con un numero medio annuo di transazioni pari a $\# = 7.19$. In un contesto caratterizzato da un'operatività molto spesso esagerata – con alcune configurazioni che hanno un'operatività media annua pari al doppio di quella in esame – una statistica come questa dimostra come in questo caso il sistema sia stato in grado di “scremare” i segnali operativi e di prendere una posizione solo quando ve ne era la convenienza.

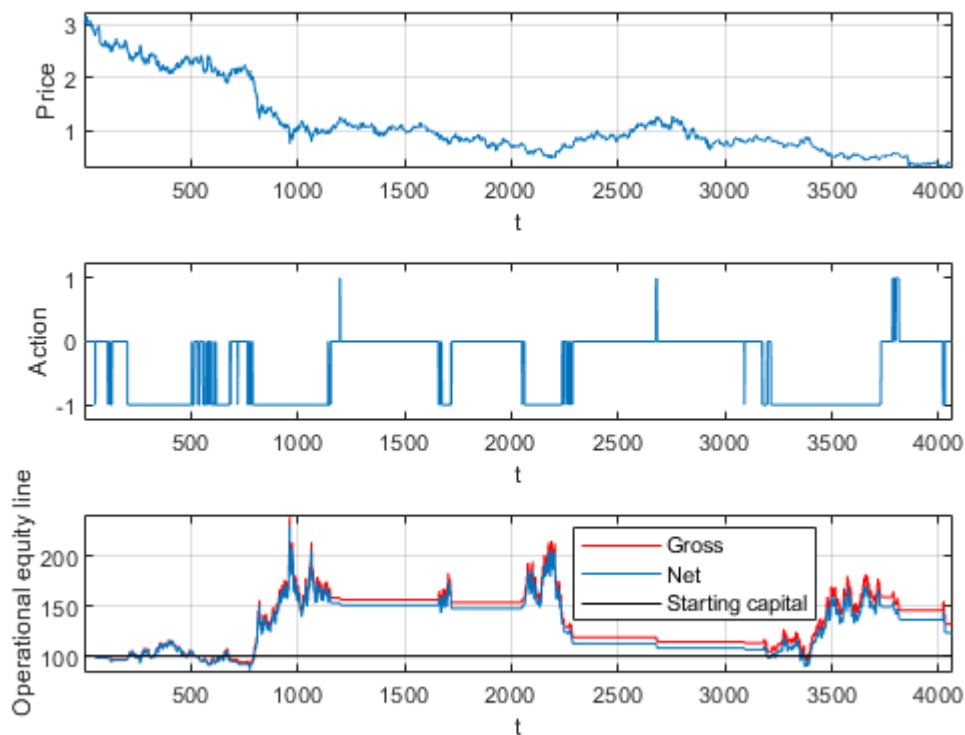


Figura 4.19: Telecom Italia S.p.A. $N = 2, L = 5, \varepsilon = 5\%, SOR$ (input: rendimenti logaritmici).

Per quanto riguarda la seconda applicazione del sistema di trading, si presenta nella Figura 4.20 la configurazione $N = 3, L = 21, \varepsilon = 5\%, SR$. Concentrando l'azione sul secondo pannello della figura, si nota che l'operatività è caratterizzata da posizioni aperte e mantenute anche per lunghi periodi. Un andamento dell'*equity line* come quello rilevabile nel terzo pannello normalmente non è indice di una buona performance, dato che il capitale nel corso di tutto il periodo considerato subisce notevoli oscillazioni. Inoltre, sono presenti periodi di inattività piuttosto lunghi in cui l'algoritmo avrebbe potuto sfruttare eventuali opportunità di profitto; si può notare comunque come questi periodi coincidano con un

andamento del prezzo piuttosto piatto (dettaglio rilevabile osservando il primo pannello). *Ex-post*, è facile vedere come il prezzo del titolo sia declinato lungo tutto l'orizzonte temporale; il sistema di trading, tuttavia, non conosce la serie storica nella sua interezza, il che giustifica un comportamento come questo che ha portato ad una performance appena sufficiente.

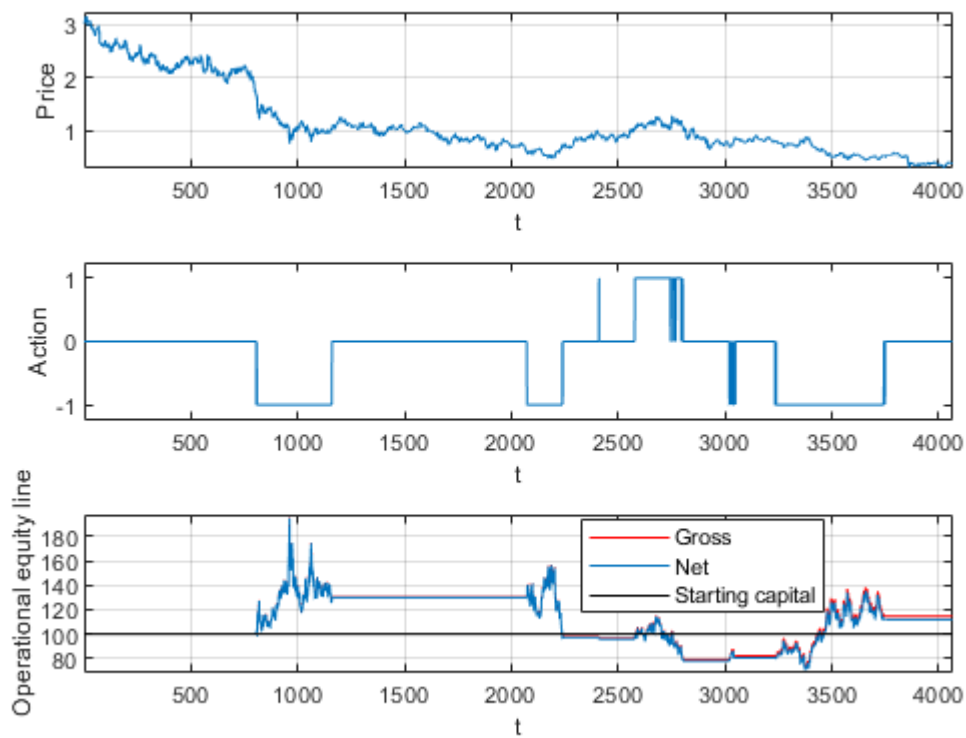


Figura 4.20: Telecom Italia S.p.A. $N = 3, L = 21, \varepsilon = 5\%, SR$ (input: indicatori AT)

4.7 Considerazioni finali

Al fine di dare una valutazione più completa sull'efficacia di entrambi i sistemi di trading automatico qui implementati, i risultati ottenuti da entrambe le applicazioni sono stati raggruppati in una tabella che mostra le performance ottenute a seconda del parametro considerato. Per dare una valutazione sintetica si è deciso di calcolare quante volte il sistema di trading automatico produce un rendimento positivo e un indice di preservazione del capitale pari ad almeno il 50%. Si è deciso di raggruppare i risultati a seconda del parametro considerato, come si può osservare nella Tabella 4.16. Viene fatta una distinzione tra i risultati della prima applicazione e quelli della seconda applicazione proprio per capire se l'utilizzo degli indicatori di Analisi Tecnica nel vettore di stato abbia portato ad un miglioramento in termini di redditività e consistenza delle performance.

Parametri		Prima applicazione		Seconda applicazione	
		Input: Rendimenti Logaritmici		Input: Indicatori analisi tecnica	
		Rendimento medio annuo > 0	Statistica % > 50%	Rendimento medio annuo > 0	Statistica % > 50%
L	5	30.00%	20.00%	40.00%	35.00%
	21	22.50%	17.50%	40.00%	40.00%
ε	5%	20.00%	12.50%	50.00%	40.00%
	15%	32.50%	25.00%	30.00%	35.00%
Reward	Sharpe	30.00%	15.00%	50.00%	45.00%
	Sortino	22.50%	22.50%	30.00%	30.00%

Tabella 4.16: Confronto delle performance raggruppate in base al parametro utilizzato.

La Tabella 4.16 evidenzia molto chiaramente come l'utilizzo di input diversi dai rendimenti logaritmici abbia permesso di migliorare notevolmente le performance del sistema di trading. In un unico caso il sistema di trading "originale" ha una percentuale di volte in cui il rendimento è positivo che è di poco maggiore rispetto alla sua controparte; più precisamente si tratta del caso relativo al parametro $\varepsilon = 15\%$, il quale nel 32.50% dei casi ha portato ad un rendimento positivo nella prima applicazione dell'algorithm mentre per la seconda applicazione si arriva ad una percentuale del 30%. Si tratta comunque di una differenza talmente ridotta da passare praticamente inosservata; le performance del secondo sistema di trading testato sono senza ombra di dubbio migliori.

Un aspetto interessante rilevabile dai dati riportati nella Tabella 4.16 riguarda l'efficacia delle due funzioni di *reward* utilizzate per le applicazioni dell'algorithm: con la prima applicazione, lo Sharpe ratio ottiene ottime performance in termini di redditività ma la persistenza

dell'*equity line* al di sopra del capitale investito lascia un po' a desiderare; al contrario, il Sortino ratio, pur non producendo risultati brillanti, riesce a garantire una persistenza maggiore delle *equity lines* portando di fatto a risultati piuttosto bilanciati dal punto di vista della redditività e della consistenza. La situazione, però, cambia con la seconda applicazione dell'algoritmo: le configurazioni che hanno utilizzato lo Sharpe ratio hanno avuto un aumento delle performance con rendimento positivo del 20% e la permanenza dell'*equity line* al di sopra del capitale iniziale per almeno metà del periodo osservato è passata dal 15% al 45%, un salto di qualità davvero interessante. Questo fatto induce a pensare che l'utilizzo di input leggermente più raffinati rispetto ai semplici rendimenti logaritmici abbia permesso di migliorare la qualità dei risultati ottenuti. Per quanto riguarda le configurazioni che hanno utilizzato il Sortino ratio, invece, si può registrare un miglioramento sia della redditività che della persistenza dell'*equity line* anche se non ai livelli dei risultati raggiunti con lo Sharpe ratio e mantenendo comunque un rapporto bilanciato tra i due indicatori.

Altro aspetto interessante è il confronto tra performance con $L = 5$ e $L = 21$: con la prima applicazione si rileva una difficoltà per il sistema di trading ad ottenere risultati quantomeno accettabili con un parametro $L = 21$; di fatto, la percentuale di volte in cui la permanenza dell'*equity line* sopra il capitale iniziale è maggiore del 50% è davvero bassa. La seconda applicazione dell'algoritmo porta ad uno stravolgimento di questo risultato, le configurazioni con $L = 21$ arrivano ad essere equiparabili – se non migliori – alle configurazioni con $L = 5$, ottenendo un aumento del 32.50% per quanto riguarda la permanenza delle *equity lines*. Anche l'aumento della redditività è notevole, passando da una percentuale di volte in cui il rendimento è positivo del 22.50% ad una percentuale pari al 40%.

In ultima battuta, si rileva un miglioramento notevole anche delle performance relativamente alle configurazioni con un'incidenza dell'azione esplorativa più bassa: la prima applicazione del sistema di trading ottiene risultati migliori quando viene utilizzato un parametro $\varepsilon = 15\%$, la situazione però si ribalta con la seconda applicazione dell'algoritmo portando ad una redditività positiva nel 50% dei casi e ad una statistica % maggiore del 50% nel 40% dei casi con un parametro $\varepsilon = 5\%$. Come già anticipato, la seconda applicazione dell'algoritmo porta ad un leggero peggioramento della redditività per le configurazioni che utilizzano un parametro $\varepsilon = 15\%$ ma una variazione negativa del 2.50% non influisce sul giudizio finale che viene attribuito ai due diversi sistemi di trading.

Dopo questa analisi, l'idea che l'impiego di strumenti dell'Analisi Tecnica nella costruzione di sistemi di trading automatici basati sul Reinforcement Learning possa portare a performance migliori sembra essere confermata. Si evidenzia comunque una certa difficoltà

dell'algoritmo ad ottenere risultati soddisfacenti nei periodi in cui l'andamento dei prezzi non segue una tendenza ben definita e individuabile oppure quando i prezzi oscillano con una velocità notevole. A titolo di esempio, si riporta la Figura 4.21, che rappresenta il risultato della configurazione $N = 3, L = 21, \varepsilon = 5\%, SR$ relativamente al titolo Intesa Sanpaolo S.p.A. Come si può rilevare dal terzo pannello della figura, questa non è una performance brillante. Come già accennato più volte in precedenza, la serie storica dei prezzi del titolo Intesa Sanpaolo S.p.A. è caratterizzata da un andamento estremamente volatile, il che ha inevitabilmente causato qualche problema al sistema di trading in fase di simulazione operativa. In più fasi l'algoritmo non è in grado di capire quale direzione prenderanno i prezzi, finendo per selezionare azioni non propriamente adatte alla situazione. Si nota, inoltre, una tendenza ad aprire e chiudere numerose posizioni in brevi lassi temporali, comportamento che denota una certa difficoltà ad apprendere l'ambiente finanziario.

Un ulteriore esempio dei punti di debolezza dell'algoritmo che utilizza gli indicatori AT come input si può ritrovare nella Figura 4.22, la quale riporta la rappresentazione grafica dei risultati ottenuti con la configurazione $N = 3, L = 5, \varepsilon = 15\%, SR$ applicata al titolo Telecom Italia S.p.A. La serie storica dei prezzi di Telecom Italia S.p.A. è caratterizzata da un marcato trend decrescente nella primissima fase del periodo osservato, dopodiché i prezzi scendono inesorabilmente ma in modo molto meno evidente. Un andamento in alcuni punti quasi piatto, causa problemi al sistema di trading in quanto non è in grado di individuare una tendenza definita. L'algoritmo, pertanto, non essendo in grado di individuare la direzione dei prezzi reagisce aprendo numerose posizioni. L'elevata operatività in questo caso specifico è anche da attribuire al livello di ε elevato, tuttavia si può notare come nei periodi caratterizzati da prezzi che diminuiscono in modo poco evidente ci sia un'elevata operatività da parte del sistema di trading. Un'idea che potrebbe essere utile a risolvere in parte queste difficoltà è l'utilizzo di periodi diversi per il calcolo degli indicatori di Analisi Tecnica. Come si ricorderà, sono state usate due medie mobili all'interno del vettore che descrive lo stato: una media mobile corta calcolata a 5 giorni ($MA(5)$) e una più lunga calcolata a 21 giorni ($MA(21)$). Il rischio con la media mobile $MA(5)$ è che vengano prodotti falsi segnali che potrebbero indurre il sistema di trading ad aprire o chiudere una posizione quando non ve ne è la necessità. Utilizzando però una media mobile più lunga come $MA(21)$ molti di questi falsi segnali vengono filtrati limitando l'operatività ai casi necessari. Per le serie in cui il *trend* dei prezzi non è molto evidente, però, può essere utile diminuire l'ampiezza delle finestre temporali su cui queste medie mobili vengono calcolate al fine di renderle più sensibili ai cambiamenti dei prezzi. Utilizzando quindi una media mobile corta a 5 giorni combinata con

una media mobile non più a 21 giorni ma, ad esempio, a 10 giorni l'algoritmo potrebbe essere in grado di cogliere segnali diversi e di individuare con più facilità la tendenza dei prezzi¹⁷. Al contrario, per serie dei prezzi in cui l'andamento è piuttosto imprevedibile potrebbe essere utile ampliare le finestre di calcolo delle medie mobili, in modo da renderle meno sensibili ai repentini cambi di prezzo e da permettere un *focus* del sistema di trading sulla tendenza generale. Utilizzando, ad esempio, una media mobile a 14 giorni combinata con una media mobile più lunga a 50 giorni, l'algoritmo avrebbe modo di osservare un lasso di tempo più ampio e visualizzare meglio l'eventuale tendenza nei prezzi ignorando i numerosi segnali prodotti dalle continue oscillazioni di questi.

Modifiche come quelle appena proposte potrebbero rappresentare il punto di inizio per lavori futuri in questo ambito, nella speranza che possano portare ad un miglioramento di quanto evidenziato in questo elaborato.

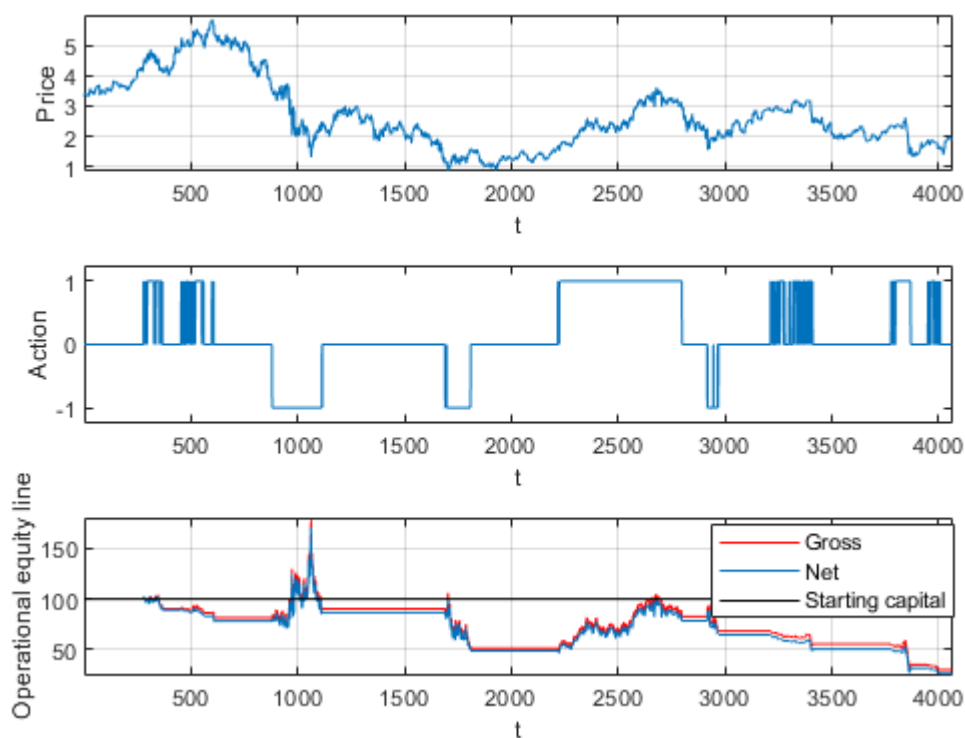


Figura 4.21: Intesa Sanpaolo S.p.A. $N = 3, L = 21, \varepsilon = 5\%, SR$.

¹⁷ A tale scopo, è stato effettuato un piccolo esperimento utilizzando come input una media mobile a 10 giorni al posto di quella a 21 giorni utilizzata precedentemente. Tale esperimento è stato condotto sul titolo Telecom Italia S.p.A., il quale presentava una serie storica dei prezzi caratterizzata da una bassa volatilità. La configurazione utilizzata per l'esperimento è $N = 3, L = 21, \varepsilon = 5\%, SR$, la quale aveva ottenuto un rendimento medio netto annuo pari a -5.82% ed una statistica $\% = 26.33$. Il rendimento medio netto annuo ottenuto cambiando l'intervallo della media mobile è pari a 0.93%, con una permanenza del capitale al di sopra del livello del capitale iniziale pari a 83.98%.

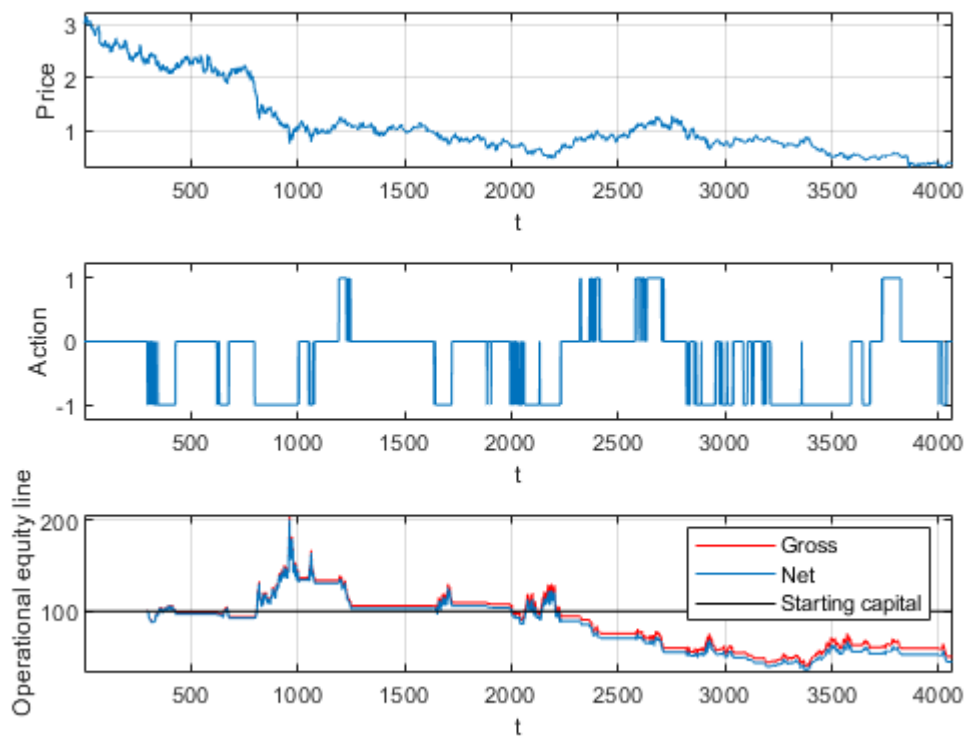


Figura 4.22: Telecom Italia S.p.A. $N = 3, L = 5, \varepsilon = 15\%, SR$.

Conclusioni

Nel presente elaborato è stato implementato un sistema di trading automatico basato su un algoritmo appartenente alla categoria dei metodi di apprendimento per differenze temporali, ossia il *Q-Learning*. L'algoritmo in questione appartiene ad una più ampia classe di metodi di apprendimento, ossia il *Reinforcement Learning*, in grado di interagire in un ambiente dinamico come i mercati finanziari. Tale sistema di trading trova la propria giustificazione grazie ad una nuova teoria che concilia i paradigmi del mercato efficiente con concetti evuzionistici quali la competizione, l'adattamento e la selezione naturale. Questa è la Teoria del Mercato Adattivo, la quale descrive uno scenario in cui i soggetti sono intelligenti ma possono comunque fallire nel prendere delle decisioni.

L'algoritmo *Q-Learning* qui implementato ha assunto il ruolo di agente artificiale con lo scopo di fornire le migliori indicazioni possibili riguardo le operazioni di trading da eseguire nell'ambito di uno specifico titolo. Il sistema è stato infatti applicato a cinque diverse serie storiche dei prezzi di titoli quotati sul mercato italiano, al fine di testarne le performance in base ai diversi andamenti dei prezzi.

Diverse configurazioni di questo algoritmo sono state applicate, variando i valori attribuiti a specifici parametri, in modo da indagare nella maniera più completa possibile quali combinazioni hanno portato ai migliori risultati sul lungo periodo. Un aspetto particolare che si è voluto approfondire è l'utilizzo di diversi input all'interno del vettore che descrive lo stato. È possibile, infatti, distinguere l'applicazione dell'algoritmo in due parti: la prima, in cui vengono utilizzati i rendimenti logaritmici come input, e la seconda, in cui si utilizzano due medie mobili e il *Relative Strength Index* come input. L'idea alla base è che gli indicatori AT possano agevolare il sistema di trading nell'individuazione di tendenze nei prezzi, permettendogli di agire conseguentemente. È stata eseguita un'analisi dei risultati prodotti da entrambe le applicazioni per poi convergere in un confronto delle due performance generali. A parità di tutti gli altri parametri, l'algoritmo che utilizza gli indicatori di Analisi Tecnica

come input ha prodotto risultati migliori sia in termini di redditività che di operatività, confermando di fatto quanto ipotizzato nella fase iniziale del presente elaborato.

Nonostante gli ottimi risultati evidenziati, permane una difficoltà dell'algoritmo ad individuare il modello di comportamento più adatto da seguire in presenza di serie storiche il cui andamento è o estremamente imprevedibile o caratterizzato da periodi senza una tendenza ben precisa. Al fine di superare tale ostacolo è stata proposta una modifica degli intervalli temporali utilizzati per il calcolo delle medie mobili per rendere l'algoritmo più o meno sensibile alle oscillazioni dei prezzi.

In generale, si possono ritenere i risultati qui ottenuti come più che soddisfacenti, rendendoli di fatto un punto di partenza per eventuali futuri sviluppi.

Bibliografia

- [1] J. E. Stiglitz, «The allocation role of the stock market: Pareto optimality and competition,» *The Journal of Finance*, vol. 36, n. 2, pp. 235-251, 1981.
- [2] E. F. Fama, «Random Walks in Stock Market Prices,» *Financial Analyst Journal*, vol. 51, n. 1, pp. 55-59, 1965.
- [3] P. A. Samuelson, «Proof that Properly Anticipated Prices Fluctuate Randomly,» *Management Review*, vol. 6, n. 2, pp. 41-49, 1965.
- [4] E. F. Fama, L. Fisher, M. C. Jensen e R. Roll, «The Adjustment of Stock Prices to New Information,» *International Economic Review*, vol. 10, n. 1, pp. 1-21, 1969.
- [5] A. W. Lo e A. C. MacKinlay, «Stock market prices do not follow random walks: Evidence from a simple specification test,» *The Review of Financial Studies*, pp. 41-66, 1988.
- [6] S. J. Grossman e J. E. Stiglitz, «On the Impossibility of Informationally Efficient Markets,» *The American Economics Review*, vol. 70, n. 3, pp. 393-408, 1980.
- [7] A. W. Lo, «Adaptive Markets and the New World Order,» *Financial Analyst Journal*, vol. 68, n. 2, pp. 18-29, 2012.
- [8] P. J. Kaufman, *Trading Systems and Methods*, Wiley, 2013.
- [9] J. J. Murphy, *Technical Analysis of the financial markets: A comprehensive guide to trading methods and applications*, Penguin, 1999.
- [10] M. Corazza e F. Bertoluzzo, «Reinforcement Learning for automatic financial trading: Introduction and some applications,» *Procedia Economics and Finance*, vol. 3, n. C, pp. 68-77, 2012.

- [11] M. Corazza e A. Sangalli, «Q-Learning and SARSA: a Comparison Between Two Intelligent Stochastic Control Approaches for Financial Trading,» *Università Ca'Foscari di Venezia, Dipartimento di Economia - Research Paper Series*, vol. 15, 2015.
- [12] A. Gosavi, *Simulation-Based Optimization. Parametric Optimization Techniques and Reinforcement Learning*, Springer, 2015.
- [13] A. Gosavi, «Reinforcement Learning: A Tutorial Survey and Recent Advances,» *INFORMS Journal on Computing*, vol. 21, n. 2, pp. 178-192, 2009.
- [14] D. P. Bertsekas e J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.
- [15] R. S. Sutton e A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 2018.
- [16] E. J. Elton, M. J. Gruber, S. J. Brown e W. N. Goetzmann, *Modern portfolio theory and investment analysis*, Wiley, 2014.
- [17] E. Barucci e C. Fontana, *Financial Markets Theory*, Springer, 2017.
- [18] E. F. Fama, «The Behavior of Stock-Market Prices,» *The Journal of Finance*, vol. 38, n. 1, pp. 34-105, 1965.
- [19] B. G. Malkiel e E. F. Fama, «Efficient Capital Markets: A Review of Theory and Empirical Work,» *The Journal of Finance*, vol. 25, n. 2, pp. 383-417, 1970.
- [20] A. W. Lo, «Reconciling Efficient Markets with Behavioral Finance: the Adaptive Markets Hypothesis,» *Journal of Investment Consulting*, vol. 7, n. 2, pp. 21-44, 2005.
- [21] A. W. Lo, «The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective,» *The Journal of Portfolio Management*, vol. 30, n. 5, pp. 15-29, 2004.
- [22] A. W. Lo, *Adaptive Markets: Financial Evolution at the Speed of Thought*, Princeton University Press, 2017.

- [23] M. Corazza, G. Fasano, R. Gusso e R. Presenti, «A comparison among Reinforcement Learning algorithms in financial trading systems,» *Università Ca'Foscari di Venezia, Dipartimento di Economia - Research Paper Series*, vol. 33, 2019.
- [24] M. Corazza e F. Bertoluzzo, «Q-Learning-Based Financial Trading Systems with Applications,» *SSRN Electronic Journal*, 2014.
- [25] M. Wiering e M. v. Otterlo, *Reinforcement Learning: State-of-the-Art*, Springer Science & Business Media, 2012.