



Università
Ca' Foscari
Venezia

Corso di Laurea Magistrale in Economia e Finanza

(ordinamento ex D.M. 270/2004)

Tesi di Laurea

Q-Learning: un algoritmo ad apprendimento per rinforzo applicato al trading finanziario

Relatore

Ch. Prof. Marco Corazza

Laureando

Enrico Cesaro
Matricola 837791

Anno Accademico

2016 / 2017

Sommario

INTRODUZIONE	5
1. IL TRADING FINANZIARIO: STORIA RECENTE ED <i>EXPERTISE</i> ATTUALE	9
1.1 Il trading online come fenomeno diffuso: dalle origini ad oggi.....	10
1.2 La conoscenza dei mercati e degli investimenti: la situazione in Italia	18
2.1 Il concetto di “mercato efficiente”	22
2.2 Gli articoli di Samuelson e di Fama del 1965.....	24
2.3 La teorizzazione di Fama delle EMH nel 1969	28
2.3.1 I modelli submartingale e random walk come implicazioni del <i>fair game</i>	31
2.3.2 Condizioni del mercato consistenti con le EMH	33
2.4 Test empirici sulle EMH	33
2.5 Critiche e sviluppi empirici delle EMH	36
2.6 Le Adaptive Market Hypothesis di Andrew Lo	40
2.6.1 Un nuovo ordine mondiale.....	40
2.6.3 Applicazioni concrete delle AMH.....	46
2.6.4 <i>Sopravvivenza</i> e nuove tecnologie: sfide rinnovate per il <i>portfolio management</i> ...	51
3. REINFORCEMENT LEARNING	53
3.1 La storia del <i>reinforcement learning</i> : cenni	55
3.2 Elementi base del <i>reinforcement learning</i>	57
3.2.1 La funzione di ricompensa.....	60
3.2.2 Il processo decisionale <i>à la</i> Markov	61
3.2.3 La funzione di valore: computazione ed ottimizzazione	62
3.2.4 Iterazione e valutazione del criterio π	66
3.3 Apprendimento per differenze temporali: i metodi TD	69
3.4 Q-Learning: algoritmo <i>off-policy</i>	72
3.4.1 Approssimazione della funzione di valore.....	73
3.4.2 Metodo della <i>discesa del gradiente</i>	75
3.4.3 Forma lineare della funzione $Q(s, a, \theta)$	78
4. APPLICAZIONE PRATICA DEL QLa: STRUTTURA	81
4.1 Elementi base dell’algoritmo	82
4.1.1 Descrittori dello stato s : rendimenti logaritmici.....	82
4.1.2 La funzione di <i>reward</i> r_t	83
4.1.3 Costi di transazione δ	85

4.1.4 Struttura di selezione di a_t	86
4.1.5 Funzione di <i>squashing</i>	88
4.2 Serie storiche dei prezzi.....	89
4.3 Fase di simulazione, fase operativa e rappresentazione dei risultati.....	95
5. APPLICAZIONE PRATICA DEL QLa: RISULTATI	101
5.1 Risultati: singoli titoli, settaggi $N = 1, N = 5$	105
5.2 Considerazioni finali	128
CONCLUSIONI.....	137
Bibliografia.....	139

INTRODUZIONE

La teoria dei mercati efficienti, nota come *Efficient Market Hypothesis*, è quasi certamente una delle teorie dei mercati finanziari più dibattute ed argomentate dell'ultimo secolo. In Lo (2004) è riportata una divertente storiella, ovviamente di natura metaforica, riguardo un economista a passeggio in compagnia di un collega. Questi trova casualmente una banconota da 100 dollari per la strada e, non appena il tale si china per raccoglierla, ecco che l'economista esclama: «Non affaticarti! Se fossero davvero 100 dollari autentici, qualcuno li avrebbe già raccolti». Questa logica economica vagamente contorta è in realtà un'accurata riproposizione del principale assioma della teoria dei mercati efficienti, secondo il quale i prezzi di mercato riflettono istantaneamente tutta l'informazione disponibile e non vi è alcuna possibilità di sfruttare qualsivoglia vantaggio economico per ottenere dei profitti. L'economista della storia è assolutamente certo, dunque, che se quella banconota fosse realmente autentica qualcuno l'avrebbe già raccolta poiché conscio della sua autenticità, e d'altro canto sarebbe un comportamento economicamente irrazionale non sfruttare un'opportunità gratuita di guadagno.

Le basi ed i costrutti teorici sui quali si fonda la teoria dei mercati efficienti sono relativamente semplici, tuttavia l'aderenza tra teoria e reale è di difficile attuazione, fattore evidenziato soprattutto nel corso delle ultime crisi economiche globali che hanno imperversato sui mercati, quando razionalità ed efficienza informativa hanno lasciato il posto a crolli finanziari, volatilità, incertezza. Nonostante ciò, tale teoria è ancora, a circa sessant'anni dalla sua formulazione, *resiliente* a qualsiasi tipo di confutazione. Tra gli economisti e gli accademici non vi è tuttora un consenso unanime sulla validità di tale teoria, introdotta da Paul Samuelson ed Eugene Fama alla fine degli anni '60, e questa viene ancora oggi studiata, approfondita, verificata, confutata. In un mercato pienamente efficiente, gli agenti si comportano razionalmente in base alle

proprie aspettative sui prezzi, in cerca di un maggior profitto: tale comportamento permette l'incorporazione istantanea delle informazioni a disposizione degli agenti di mercato nei livelli dei prezzi, in questo modo le opportunità di profitto vengono repentinamente eliminate ed i prezzi variano randomicamente. In un contesto come quello descritto, quale vantaggio vi sarebbe ad operare nei mercati? Gli investitori sono spinti ad effettuare operazioni di *trading* dai potenziali livelli di extra-profitto futuri ed agiscono in modo tale da sfruttare le proprie conoscenze per ottenere un vantaggio competitivo che permetta loro di accrescere il proprio patrimonio. In un *framework* come quello previsto dalla teoria dei mercati efficienti non esiste la possibilità di ottenere tali profitti, poiché questi sono istantaneamente incorporati nei livelli dei prezzi. È dunque verosimile immaginare un mercato descritto dalle ipotesi di Eugene Fama, oppure è necessario spingersi oltre?

A partire dagli anni '90, cominciarono a sorgere correnti di pensiero alternative all'impostazione classica, in particolar modo attorno alla natura del comportamento degli individui ed alle decisioni da questi intraprese nel partecipare attivamente ai mercati finanziari. Queste scelte, molto spesso, sono guidate da processi decisionali distorti, specialmente quando gli individui proiettano sé stessi nel futuro e sono dunque chiamati a prevedere scenari incerti. Non è raro il verificarsi di un'eccessiva confidenza e fiducia nelle proprie decisioni, oppure di possedere un livello di attitudine finanziaria spesso sovrastimato ed una conoscenza dei mercati altrettanto erratamente ottimistica. Un tale scenario sembrerebbe essere più verosimile: immaginare i mercati come non pienamente efficienti darebbe spazio a concrete possibilità di *trading* e profitti, poiché le informazioni non verrebbero istantaneamente incorporate nei prezzi e poiché errori comportamentali, se implicano una perdita di capitale per taluni individui, potrebbero configurarsi come opportunità di profitto per altri. Supporre la razionalità come cardine del comportamento umano è un'ipotesi *verosimile* ma non *reale*, valida in alcune situazioni ma semplicistica in altre. Come conciliare dunque la teoria dei mercati efficienti con l'irrazionalità insita nel comportamento umano? La risposta a siffatto quesito viene fornita da un nuovo approccio *evoluzionistico* applicato alle interazioni economiche, secondo il quale gli individui agiscono in un contesto estremamente dinamico e mutevole e, nel tentativo di adattarsi ad esso, incappano necessariamente

in errori di comportamento che tuttavia permettono loro di migliorare il proprio processo di apprendimento e di assumere comportamenti sempre più efficienti ed ottimi. Sono le leggi di selezione naturale a determinare l'evoluzione del mercato, a selezionare la strategia di *trading* vincente, a condurre all'*estinzione* (metafora di uscita dal mercato) quegli individui non in grado di intraprenderle. Questi sono solo alcuni dei capisaldi del nuovo possibile modello descrittivo dei mercati, che prende il nome di *Adaptive Market Hypothesis* e del quale Andrew Lo, professore universitario di finanza alla MIT Sloan School of Management di Cambridge (Massachusetts) nonché direttore del MIT Laboratory for Financial Engineering, è uno dei maggiori sostenitori, tanto più padre fondatore. Imperniata su concetti di psicologia, biologia evuzionistica, neuroscienza ed altre idee interdisciplinari, la teoria dei mercati adattivi dimostra come le ipotesi di Fama non siano propriamente inesatte, quanto meramente incomplete: in presenza di mercati instabili, gli investitori agiscono istintivamente ed irrazionalmente, creando inefficienze sfruttabili da altri; perciò, il grado di efficienza di un mercato dipende da fattori plurimi quali il numero dei competitors, la loro capacità di adattamento, la possibilità di profitti.

La *Adaptive Market Hypothesis* può dunque essere vista come un'estensione della *Efficient Market Hypothesis*, e secondo questo *framework* innovativo gli individui agiscono ed interagiscono con l'ambiente economico circostante, perseguendo il proprio interesse personale, commettendo spesso errori i quali permettono loro di migliorare il proprio processo di apprendimento ed adattare il proprio comportamento. I mercati finanziari sono un'ambiente estremamente complesso con il quale interagire, dinamici ed estremamente veloci nel modificare la propria configurazione; per questo motivo, potrebbe essere profittevole *creare* un agente artificiale ed intelligente in grado di apprendere più celermente, rispetto ad un individuo umano, l'ambiente circostante, così da sviluppare una strategia di *trading* efficace ed adattiva. Uno dei metodi principalmente usati per la costruzione di tale agente viene definito *Reinforcement Learning*, ovvero *apprendimento per rinforzo*: più che di metodo, è corretto parlare di una branca di metodi computazionali concernenti lo sviluppo di agenti artificiali in grado di apprendere un determinato comportamento, adatto ad una situazione specifica, interagendo direttamente con l'ambiente nel quale sono immersi. Nell'ambito di un

ambiente finanziario, l'obiettivo è costruire un metodo capace di scegliere quale strategia adottare (*buy, stay out of the market, sell*) dati un numero definito di titoli ed un set di informazioni basilari, e verificare se tale metodo risulta soddisfacente in termini di profitto e di strategia.

Questo elaborato di tesi si propone di esporre i punti di forza e di debolezza della teoria dei mercati efficienti, di presentare la nuova teoria dei mercati adattivi proposta da Andrew Lo e dimostrare come questa si dimostra essere più aderente a rappresentare la realtà dei mercati finanziari. Verranno dunque districate le principali caratteristiche dei metodi ad apprendimento per rinforzo, entrando poi nello specifico del metodo *Q-Learning*, un algoritmo a differenze temporali largamente utilizzato nella letteratura empirica di genere. Verrà infine sviluppata una realizzazione pratica ai risvolti teorici presentati, applicando tale algoritmo ad uno stock di titoli reali del mercato italiano. Il primo capitolo di questa tesi tratta i principali traguardi storici che il fenomeno del *trading online* ha perseguito nel corso della storia recente, per poi rappresentare lo spaccato degli investitori italiani *retail* e dimostrare come sia impensabile immaginare tali investitori *razionali*. Il secondo capitolo presenta la teoria dei mercati efficienti in tutte le sue principali sfaccettature, partendo dagli articoli di Samuelson e Fama e dalle controversie accademiche che questi hanno generato; si passa poi ad un'analisi della teoria dei mercati adattivi di Andrew Lo e delle novità che essa comporta rispetto alla teoria dei mercati efficienti. Il terzo capitolo presenta i metodi computazionali ad apprendimento per rinforzo e la struttura teorica dell'algoritmo *Q-Learning*, il cui sviluppo pratico viene implementato in ambiente *Matlab®* ed esposto nel quarto capitolo. Il quinto ed ultimo capitolo raccoglie invece risultati e considerazioni finali. La costruzione dell'algoritmo *Q-Learning* è ispirata ad un modello generale, presentato in alcune delle fonti bibliografiche analizzate¹, al quale sono state apportate delle modifiche raramente implementate in letteratura, al fine di verificarne l'efficacia e le *performance* da esse derivanti.

¹ Corazza e Bertoluzzo (2012, 2015), Corazza e Sangalli (2015).

1. IL TRADING FINANZIARIO: STORIA RECENTE ED *EXPERTISE* ATTUALE

Al giorno d'oggi, qualora si volesse interrogare un campione di clientela di una qualsiasi banca operante sul territorio italiano riguardo al significato della parola *trading online*, sicuramente la maggioranza degli interlocutori risponderebbe con certezza che significa operare compravendita di titoli su piattaforme on-line il cui accesso può essere effettuato tramite il portale della propria banca o tramite piattaforme specializzate facilmente accessibili. Inoltre, una buona parte di questi risponderebbero che loro stessi eseguono operazioni giornaliere di investimento, senza avvalersi di consulenti specializzati o esperti del settore. Il fenomeno del *trading online*, perché è di fenomeno di cui sembra corretto scrivere, ha assunto proporzioni sempre maggiori, di pari passo con l'introduzione di innovazioni tecnologiche, per ultima lo smartphone: qualsiasi banca offre ai propri clienti la possibilità di operare digitalmente tramite app, effettuando operazioni in *multicurrency*, acquistando e vendendo CFD², titoli ed obbligazioni. Un'indagine effettuata recentemente da KPMG Advisory³, nota agenzia di consulenza internazionale, riporta che il 66% della popolazione italiana fruisce di connessione internet, mentre il 47% dispone di servizi internet da *mobile*, dato in costante crescita ed aggiornamento; volendo invece indagare sul numero di italiani che

² CDF è l'acronimo dell'inglese "Contract for difference", ovvero *contratto per differenza*. È uno strumento derivato tramite il quale due parti si accordano in tal modo: l'acquirente si impegna al pagamento di un tasso di interesse a fronte della corresponsione, da parte del venditore, del rendimento dell'attività finanziaria regolata dal CDF. Le due parti si accordano per scambiarsi i flussi finanziari derivanti dal *differenziale* tra il prezzo del sottostante al momento dell'apertura del contratto ed al momento della chiusura. (Fonte: Borsa Italiana)

³ Tale indagine è stata condotta nel mese di luglio 2016 su un campione di 1.081 individui che risulta essere rappresentativo della "popolazione italiana *bancarizzata*" (KPMG, 2017).

utilizzano il *mobile banking*, la percentuale scende attorno al 27% e, nonostante si attesti ad un livello modesto, conferma comunque il ruolo sempre più crescente dello smartphone come mezzo tecnologico principale di fruizione di servizi bancari, in sostituzione di un *internet banking* ancora in auge, ma in flessione rispetto agli anni precedenti. Il sempre più elevato utilizzo dei servizi bancari digitalizzati, congiuntamente ad una crescente diversificazione degli stessi, sembra avvicinare nuova e vecchia clientela ad interessarsi a quei servizi complementari rispetto ai servizi classici, quali appunto i servizi di investimento. Sorge dunque spontanea una domanda: essendo oramai il *trading online* accessibile tramite smartphone o pc alla maggioranza dei clienti bancari, il livello di *expertise* e di conoscenza dei mercati è tale da permettere a tali clienti di operare in maniera oculata e conscia dei rischi connessi alle suddette operazioni? Prima di approfondire tali aspetti, è opportuno comprendere la portata storica del fenomeno, soffermandosi brevemente sulle origini del *trading online* e su come questo si sia sviluppato nel corso della storia recente. In questo capitolo vengono ripercorsi gli episodi più significativi della storia del trading online, soffermandosi in particolare sullo sviluppo del trading in America per analizzare poi brevemente anche a come il fenomeno sia arrivato in Italia e come qui si sia sviluppato.

1.1 Il trading online come fenomeno diffuso: dalle origini ad oggi

Non è corretto, innanzitutto, parlare di *invenzione* del *trading online* poiché si tratta di un processo sviluppatosi nel tempo, formato dall'interrelazione di tre principali macro aree: la *normativa finanziaria*, la *tecnologia* e i *sistemi hardware e software*. Le continue interconnessioni tra queste macro aree rendono il fenomeno del *trading online* in costante e continua evoluzione, di cui però è possibile individuare alcuni storici momenti chiave. Andrea Fiorini, editore dell'Annuario del Trading Online Italiano, nel libro "Storia del Trading Online" individua negli ingenti investimenti in favore degli sviluppi tecnologici che vennero promossi dal Governo degli Stati Uniti a cavallo tra le due Guerre Mondiali - ma soprattutto nel corso della Seconda Guerra Mondiale (dagli anni '40 in poi) - il processo scatenante che ha portato alla nascita delle prime reti Internet e dei primi elaboratori di calcolo utilizzati a scopo non militare. Fu proprio l'intervento

degli Stati Uniti, superpotenza mondiale che solo negli ultimi anni del conflitto poteva contare su un budget a disposizione della Difesa pari a 108 miliardi di dollari⁴, a dare una svolta definitiva al conflitto bellico. Queste enormi somme a disposizione vennero investite in sviluppi ed innovazioni tecnologiche finalizzate a finanziare l'esercito prima e, anche se in misura minore, a sviluppare tecnologie all'avanguardia una volta terminato il conflitto. Un ulteriore impulso ad investire in tecnologia militare lo diede sicuramente lo scontro USA-URSS a partire dagli anni '50, il quale non si tradusse mai in una guerra diretta ma, com'è noto dai libri di storia, in una *Guerra Fredda* la cui finalità principale tendeva alla prevaricazione sull'avversario soprattutto per quanto concerneva la tecnologia bellica: a partire dal 1956, ad esempio, il Pentagono intraprese la realizzazione di una rete di difesa aerea all'avanguardia denominata S.A.G.E. (*Semi Automatic Ground Environment*), a protezione di tutto il territorio Nord Americano e finalizzata all'avvistamento di bombardieri sovietici. Questi potenti mezzi militari vennero successivamente utilizzati a supporto delle prime reti Internet che progressivamente vennero aperte anche all'uso civile, secondo un fenomeno che viene definito di *reverse engineering*, la cui traduzione – ingegneria inversa – indica proprio l'utilizzo ad uso civile di tecnologie inizialmente sviluppate per scopi militari (Fiorini, 2015). Continui ed ingenti investimenti che andavano di pari passo, negli Stati Uniti come in Europa, con un numero sempre maggiore di banche che sfruttavano queste nuove tecnologie, in particolar modo i moderni e potenti elaboratori di calcolo digitali per lo svolgimento di attività interne; già verso la fine degli anni '70, questi elaboratori cominciarono a trovare diffusione presso le grandi aziende del settore, per poi diffondersi in un secondo momento anche presso le piccole e medie imprese e presso gli utenti privati, assieme ad altre innovazioni automatizzate⁵.

A tale processo di rinnovamento tecnologico cominciarono a conformarsi, già dalla fine degli anni '50, alcune Borse – tra cui la *New York Stock Exchange* o N.Y.S.E – con la finalità principe di rendere più efficienti le operazioni di negoziazione dei titoli azionari: basti

⁴ Fiorini, Andrea. *Storia del trading online*. Milano, Mediosfera, 2015, pag. 8.

⁵ Una delle più famose novità tecnologiche di quegli anni fu l'introduzione, tra il 1960 ed il 1970, delle ATM, acronimo per *Automated Teller Machines*, che in Italia prendono il nome di Bancomat.

pensare, infatti, che verso la fine del decennio citato il mercato USA vantava un volume di scambio di azioni pari a 350.000 unità annue dalle quali operazioni di compravendita, a causa di errori dovuti ai sistemi di ordine (fax, reti telefoniche, cartacei), andavano sprecati all'incirca 100 milioni di dollari l'anno⁶. Il primo punto di svolta si ebbe nel 1961 quando la SEC statunitense⁷ intraprese un'indagine sui volumi di scambio nei mercati OTC⁸ americani il cui responso, giunto due anni dopo, individuò la necessità di introdurre l'automazione dei processi di borsa per regolamentare il mercato. Intervenne dunque il NASD⁹ il quale, qualche anno dopo (1968) e su richiesta del SEC, affidò alla Bunker Ramo, una società specializzata nella costruzione di tecnologia elettronica militare, l'implementazione di un sistema che permettesse la contrattazione telematica dei titoli. Tre anni dopo, nel 1971, viene ultimato quella che fu una vera e propria rivoluzione, ovvero la prima piattaforma telematica per la negoziazione di titoli OTC disponibile esclusivamente per gli associati del NASD. L'acronimo utilizzato per suddetta piattaforma fu *National Association of Security Dealers for Automatic Quotation*, in breve NASDAQ. Questo fu solo il primo e storico passo verso la digitalizzazione dei servizi di *trading*: con NASDAQ infatti si avviò un processo di telematizzazione delle negoziazioni, che portò gradualmente anche le istituzioni a guardare con attenzione sempre maggiore al fenomeno e ad adeguare le normative di settore per regolamentarizzare il mercato. Tra i primi emendamenti più importanti, si ricordano senza dubbio l'*Electronic Fund Transfer Act* (1978) e il *Tax Equity and Fiscal Responsibility*

⁶ Indagine della Borsa di New York effettuata nei primi anni '60 (citata in Fiorini, Andrea. *Storia del trading online*. Milano, Mediosfera, 2015, pag. 8).

⁷ SEC è l'acronimo di *Security and Exchange Commission* ovvero l'autorità federale preposta alla vigilanza della borsa valori statunitense, l'analogo della Consob italiana.

⁸ Un mercato *Over the Counter* (OTC) è un mercato non soggetto alla regolamentazione ufficiale le cui modalità di contrattazione non sono standardizzate. Secondo la maggioranza delle fonti, questa locuzione nacque proprio in America per descrivere tutte quelle contrattazioni di azioni societarie che avvenivano al di fuori della borsa di Wall Street, molto spesso sui banconi dei bar adiacenti la Borsa, da cui il termine inglese "over the counter" ovvero "sopra il bancone" (Fonte: Borsa Italiana).

⁹ Il NASD, per esteso *National Association of Security Dealers*, venne fondato nel 1939 e divenne l'ente preposto alla Vigilanza ed all'autoregolamentazione dei mercati "over the counter". Nel 2007 viene inglobato nel FINRA (*Financial Industrial Regulatory Authority*), organo di Vigilanza non governativo – ma autorizzato dal Congresso degli Stati Uniti – per la tutela degli investitori (Fonte: www.finra.org).

Act (TE-FRA, 1982), normative che miravano entrambe a migliorare la trasparenza informativa degli intermediari e ad obbligare gli stessi a fornire agli enti federali più informazioni riguardanti l'attività svolta, la cui mole era gestibile solo attraverso l'utilizzo ed il supporto dei mezzi informatici. Se nel nuovo continente viene ricordato il NASDAQ, in Europa in quegli anni veniva introdotto (nel 1977) lo SWIFT, un sistema di pagamento elettronico finalizzato al mercato interbancario sviluppato a Bruxelles ma che già dai primi mesi del suo avvio poteva contare su una rete di 293 banche di 15 paesi in tutto il mondo. Tutte tecnologie che gettano importantissime basi per la nascita dei sistemi bancari digitalizzati così come li conosciamo ai giorni nostri.

Il processo di telematizzazione della Borsa, a fine anni '80, portò negli USA alla nascita dei primi servizi di negoziazione di titoli finanziari mediante l'utilizzo di tastiere telefoniche, innovazione pensata proprio per utenti privati: il telefono era ancora il primo mezzo di comunicazione globale poiché il computer ed internet, all'epoca, non avevano una diffusione così capillare da considerarli come strumenti principe per la contrattazione e la compravendita di titoli. Ancora una volta, l'Europa si muove in parallelo agli Stati Uniti: nel medesimo periodo, nel Regno Unito e successivamente nel resto del continente, viene introdotto Videotex, il primo servizio telematico per la diffusione di dati e di messaggistiche che oggi viene definito uno dei predecessori di internet. Era infatti un sistema di reti centralizzato e gestito da società di telecomunicazioni, attraverso il quale era possibile visualizzare, sullo schermo televisivo, una serie di informazioni e trasmettere, come output di risposta, le proprie decisioni tramite apparecchi specificatamente realizzati, connessi al televisore ed al telefono. Questa tecnologia vide luce anche negli USA: al primo servizio di tal specie, Channel 2000, introdotto nel 1980 e subito sostituito, a causa del suo scarso successo, fece seguito un altro sistema rinnovato che prese il nome di Applause. In Italia questa tecnologia arriverà qualche anno dopo (1985) sotto il nome di Videotel. Nasce dunque il primo rudimentale servizio di *home banking*, anche se ebbe vita e successo molto brevi. Erano, come precedentemente descritto, servizi fruibili tramite reti telefoniche di gestione monopolistica, i cui costi di fruizione risultavano ancora troppo alti per utenti non professionisti. Inoltre, tali servizi nella maggioranza dei casi venivano trasmessi in reti chiuse e private, per cui non vi era la possibilità di interconnettere la pluralità di reti

a livello globale. Per questa e per altre ragioni, i Videotex vennero sostituiti nel 1994, anno che segna la nascita del Web, vera e propria invenzione rivoluzionaria per tutta la tecnologia del nuovo millennio. Prima di arrivare a tale innovazione, è bene però tornare indietro di qualche anno: se da un lato infatti iniziarono a moltiplicarsi le tecnologie per un utilizzo più semplificato dei servizi di negoziazione, anche le agenzie di brokeraggio cominciarono ad adattarsi offrendo servizi agli utenti sempre più telematizzati; per citarne alcuni, si ricordano E*Trade e NAICO-Net, due delle più famose agenzie di brokeraggio americane dell'era pre-internet. Si viene a creare dunque un duplice sistema di concorrenza: *verticale* tra agenzie broker ed *orizzontale* tra società finanziarie e banche per disponibilità finanziarie e di prodotti offerti (Fiorini, 2015).

Il fermento di innovazioni che a fine anni '80 popolava il mondo del trading online vide però una brusca ed improvvisa frenata il 19 ottobre 1987, giorno del celeberrimo *lunedì nero*, ovvero il crollo dei mercati più grave di tutta la storia economica mondiale: le Borse di tutto il mondo, a partire da quella di Hong Kong, videro un crollo in una seduta in alcuni casi pari anche al 60%; l'indice Dow Jones americano perse 508 punti (pari a circa il 23%), l'indice della Borsa di Milano (all'epoca ancora non era stata istituita la Borsa Italiana) perse l'11% il lunedì ed il 12% il martedì. Secondo uno studio della società *Amen & Associates*, i più grandi gestori di fondi statunitensi, rispetto ad i dati di gennaio 1987, persero circa 500 miliardi di dollari pari al 40% dei propri capitali in gestione¹⁰. Tale improvvisa *caduta*, secondo alcuni analisti di settore tra cui anche Andrea Fiorini, fu in parte causata dai cosiddetti *program trading*, una tipologia di negoziazione di gruppi di titoli che, grazie alla possibilità dei mezzi informatici di trattare moli sempre più elevate di ordini, permise la vendita incontrollata di titoli causando il *crash* del 19 ottobre. Ad ogni modo, la causa principale individuata dalla Commissione Brady, incaricata dal Parlamento degli Stati Uniti di scovare le ragioni principali della crisi, fu l'errata strategia di credito attuata da alcuni grandi investitori istituzionali che portò al crollo del 45% dell'indice della Borsa di Hong Kong, crisi che si estese poi a macchia d'olio anche al mercato statunitense ed al mercato europeo. La diretta conseguenza di tale crollo fu un arresto del processo di telematizzazione del *trading online*: broker e società

¹⁰ Fiorini, Andrea. *Storia del trading online*. Milano, Mediosfera, 2015, pag. 169.

finanziarie impiegheranno qualche anno per riprendere spazio e fiducia nei mercati, anche a fronte di un'opinione pubblica improvvisamente ostile o quanto meno diffidente dopo il crollo registrato il *lunedì nero*. Lo sviluppo dei servizi di *trading online* vedrà una ripresa già nei primi anni '90, ma è soprattutto con lo sviluppo del World Wide Web che tale fenomeno raggiunge livelli e volumi importanti; volendo fornire qualche indicazione numerica, nel 1995 il 20% della popolazione USA possiede almeno uno strumento finanziario, nel 1999 Forrester Reserach (nota compagnia americana per le ricerche ed indagini di mercato) stima l'esistenza di circa 8,4 milioni di conti di *trading online* aperti ed operativi sul territorio USA¹¹. La negoziazione telematica giunge finalmente anche in Europa: è dal 1994 infatti che si può affermare l'esistenza della compravendita di titoli online, poiché i servizi quali Videotex precedentemente citati erano troppo limitati in termini di volumi ed utenti da considerarli tali. La compravendita di titoli online è certamente favorita anche da un progressivo abbassamento delle commissioni richieste dai broker, fattore importante che rende più accessibile, in termini di costi, il servizio e contribuisce ad ampliare la portata del fenomeno.

Il 1994 segna dunque la nascita di un nuovo ciclo economico, definito *New Economy*, il quale in pochi anni vede moltiplicarsi il numero di aziende operanti nel settore informatico e di Internet, denominate "dot.com Companies". Tali compagnie, sospinte dal fermento tecnologico e dalla totale euforia verso le recenti innovazioni di internet e del Web, vennero progressivamente sovra quotate nei mercati borsistici a causa di una eccessiva fiducia nelle aspettative future di continui e crescenti aumenti di valore di tali aziende, a prescindere dalle indicazioni che potevano derivare da indicatori tradizionali di analisi fondamentale (quali ad esempio indice di indebitamento, disponibilità liquide, previsioni di crescita nel breve e nel lungo periodo). Questa eccessiva fiducia in tali aziende portò di fatto le aspettative ad autorealizzarsi, montando nel corso del tempo una vera e propria *bolla speculativa* destinata, come la storia insegna, ad esaurirsi nel breve o medio periodo. Lo scoppio di una bolla speculativa può derivare, nella maggior parte dei casi, dalle seguenti ragioni:

¹¹ Fiorini, Andrea. *Storia del trading online*. Milano, Mediosfera, 2015, pag. 185.

- una situazione di saturazione del mercato, ossia da un mercato *ipervenduto* dove non vi sono più investitori disposti ad acquistare ai livelli di prezzo elevati raggiunti dal titolo;
- l'entrata nel mercato da parte di una componente speculativa orientata soltanto a cavalcare l'onda della profittabilità senza invece analizzare le cause iniziali dei primi aumenti;
- una revisione prospettica negativa delle potenzialità di crescita del titolo;
- la nascita di nuove e più promettenti potenzialità di crescita di altri titoli dello stesso settore o di altri settori.

Nel caso della bolla speculativa delle *dot.com Companies*, lo scoppio avvenne nel marzo 2000 quando, a seguito della pubblicazione dei bilanci da parte delle diverse aziende del settore, fu evidente che l'investimento nelle società del comparto poteva rivelarsi non così performante come sperato, vista la scarsa profittabilità delle stesse. Le quotazioni cominciarono a decrescere per effetto delle vendite da parte degli speculatori che non volevano perdere ulteriori quote di capitali investiti. Il NASDAQ, l'indice azionario di riferimento per queste società, perse in tre giorni quasi il 9% ed arrivò a fine anno a perdere il 45% del suo valore¹². Nel corso del 2001 molte *dot.com Companies* fallirono o furono oggetto di operazioni di acquisizione e fusione, e tre anni dopo solo la metà delle aziende presenti agli albori del nuovo millennio era ancora attiva: le più solide che riuscirono a resistere, oggi dominano il mercato mondiale del settore (Amazon, Ebay, Apple). Tale crollo dei mercati ebbe chiaramente conseguenze anche nelle attività di trading online, caratterizzate in questo periodo da un comportamento che si può definire tipico in presenza di bolle speculative: l'*herding behaviour*, la cui traduzione – che letteralmente significa “il comportamento del gregge” – indica l'attitudine degli investitori ad assumere comportamenti imitativi, spinti dalla generale euforia per un mercato in continua crescita ed espansione. Questo comportamento si traduce, al momento dello scoppio della bolla, in un rapido e tempestivo disinvestimento delle quote possedute per paura di perdere i capitali investiti, definito molto chiaramente *panic selling*. Gli studi sulle attitudini e comportamenti degli investitori rientrano nella

¹² Fonte: rielaborazione dati disponibili sulla piattaforma Yahoo Finanza.

più ampia branca degli studi economici definita *finanza comportamentale*, la quale applica allo studio dei mercati finanziari i principi di psicologia e di scelta legati ai comportamenti dell'uomo in un contesto sia individuale che sociale.

Se la disanima del fenomeno del *trading online* è stata principalmente focalizzata sul mercato statunitense, vero e proprio pioniere delle innovazioni e sempre all'avanguardia sulle tecnologie informatiche applicate alla Borsa, è sicuramente interessante guardare più da vicino anche al mercato italiano, la cui storia è molto più breve poiché di origini più recenti e meno articolate. Sostanzialmente, il *trading online* arriva in Italia nel 1993, quando Borsa Italiana decise di accettare le connessioni delle SIM non più solo dai classici terminali ma anche da altri elaboratori, per permettere la trascrizione automatica degli ordini eseguiti dalle stesse. Le SIM, acronimo di Società di Intermediazione Mobiliare, sono società che svolgono l'esercizio professionale dei servizi di investimento presso il pubblico, articolato in diverse attività: negoziazione di titoli per conto terzi, inclusa la raccolta e la gestione degli ordini di acquisto e vendita (brokeraggio), collocamento di titoli con o senza garanzie e gestione dei portafogli finanziari. Istituite nel 1991, videro la prima grande riforma nel 1996 con l'emanazione del "decreto Eurosim" (D.Lgs. n° 415 del 23/7/1996 in recepimento della Direttiva Europea 93/22/CEE del 10 maggio 1993), decreto fondamentale che recepisce la normativa europea introducendo sia una riforma dei *servizi di investimento* dei valori mobiliari, sia nuovi vincoli patrimoniali per le *imprese di investimento* e per gli *enti creditizi*. Nello stesso anno nacque Directa SIM, il primo operatore ad offrire una piattaforma di *trading online* in Italia e vero e proprio pioniere del settore. Dal 1996, dunque, il *trading online* vide il boom anche in Italia, fenomeno agevolato anche dalla diffusione di Internet e da un sistema di interconnessioni di rete sempre più articolato e veloce tanto che, l'anno seguente (1997), la Borsa Italiana diviene completamente telematizzata. Un'altra tappa fondamentale per la diffusione del *trading* avviene nel 1998, anno in cui la Consob, l'organismo che vigila sulla trasparenza del mercato finanziario italiano, emana il Testo Unico della Finanza, in breve TUF (D.Lgs. 24 febbraio 1998). Il TUF è di importanza fondamentale per la storia del mercato borsistico italiano in quanto tende a razionalizzare tutta la normativa finanziaria italiana dell'ultimo trentennio, disciplinando gli intermediari finanziari in modo finalmente completo e

conforme alla normativa comunitaria. La crescita e diffusione del *trading online* è stata agevolata anche dalla progressiva espansione dei servizi di *home banking*: dalla fine anni '90, in Italia gli istituti di credito cominciarono ad affiancare, all'offerta del conto corrente, anche la gestione dei conti direttamente online. Dal 2003, invece, vengono implementati i primi servizi di *trading online* per telefoni mobili, il cui utilizzo ha seguito un crescendo progressivo per arrivare alle percentuali esposte all'inizio del capitolo.

Il progressivo sviluppo tecnologico ha portato all'espansione del *trading online* da attività di nicchia a fenomeno di massa: del resto, più aumenta tale portata più è difficoltoso tener traccia dei nuovi servizi e delle nuove società sorte in questi anni. Dalla storia si passa infatti ai fatti di cronaca: lo sviluppo delle reti web e di Internet fornisce, come si è ripetuto spesso, un supporto fondamentale verso la creazione del mercato globalizzato così com'è conosciuto oggi.

1.2 La conoscenza dei mercati e degli investimenti: la situazione in Italia

Il *trading online*, come visto in precedenza, ha vissuto negli anni un'espansione progressiva e notevole. Il supporto tecnologico rende tale attività accessibile a chiunque e sembrerebbe corretto affermare che oggi "*tutti possono investire*". In realtà, la situazione in Italia è ancora molto arretrata per quanto riguarda la cultura e la conoscenza dei mercati e dei basilari principi di investimento e di rischio. A supporto di tale affermazione, vengono di seguito riportati alcuni significativi risultati del report della Consob incentrato sugli investimenti finanziari delle famiglie, basato sul sondaggio di Multifinanziaria sul mercato *retail* italiano (la numerosità del campione è di 2.500 intervistati) e sullo studio dell'Osservatorio Italiano riguardante l'approccio alla finanza ed agli investimenti delle famiglie italiane (numerosità in questo caso pari a 1.000 intervistati). Il report, pubblicato nel 2017 elaborando i risultati raccolti nel 2016, presenta uno spaccato sulle conoscenze e competenze finanziarie, in materia di investimenti e mercati, mediamente negativo: solo una media del 43% conosce nozioni economiche basilari quali diversificazione di portafoglio e relazione rischio-rendimento, percentuale che si abbassa notevolmente, attestandosi intorno al 15%, quando si passa

alle conoscenze sui profili di rischio di un prodotto finanziario, riportati nelle appendici descrittive a favore dei risparmiatori. Altro dato rilevante è che il 59% degli intervistati che afferma di preferire una composizione di portafoglio azionaria asserisce che la principale motivazione della scelta risiede nella convinzione che le azioni siano meno rischiose delle obbligazioni. La *tabella 1* riporta tutti i risultati dell'indagine, relativi alle conoscenze finanziarie.

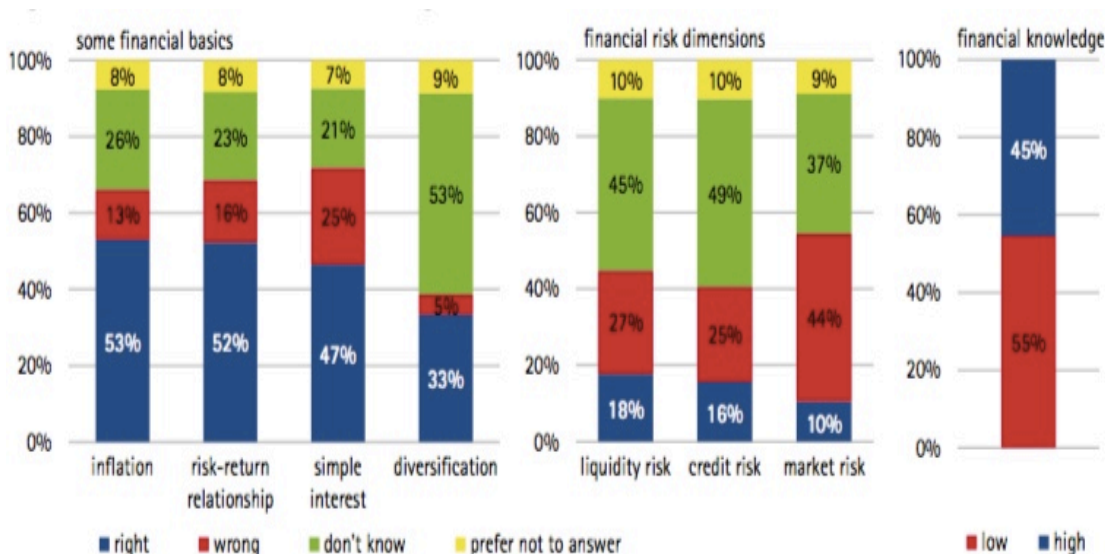


Tabella 1: risultati relativi all'indagine sulle conoscenze finanziarie degli investitori italiani.

Come si evince dalla tabella, le nozioni più conosciute risultano essere quelle di *inflazione* e *relazione rischio-rendimento* (53% e 52% di risposte corrette rispetto al campione analizzato), mentre la percentuale scende 20 punti quando la domanda riguarda il concetto di *diversificazione* (33% di risposte esatte), nozione alla quale il 62% del campione ha risposto "non so" oppure "non rispondo" (sezioni verde e gialla delle barre nel grafico). Le percentuali di risposte giuste crollano vertiginosamente nel campo della conoscenza delle dimensioni di rischio: nozioni quali rischio di liquidità, rischio di credito e rischio di mercato vedono, nello scenario più "ottimistico", una percentuale di risposte giuste pari al 18% per il rischio di liquidità, mentre la nozione di rischio di mercato è conosciuta soltanto dal 10% del campione. Nonostante i risultati esposti, il 40% degli intervistati ritiene che il proprio livello di conoscenza sia medio-alto, convinzione che si rivela essere errata evidenziando una *sovrastima* delle proprie competenze finanziarie rispetto all'effettivo livello posseduto, fattore che può portare

ad un'eccessiva *self confidence* che conduce a prendere decisioni di investimento errate. Il livello di "ansia finanziaria" è un altro fattore determinante che può condizionare scelte e decisioni di investimento: il 50% del campione (specialmente la porzione meno preparata in materia finanziaria) afferma di essere fortemente condizionata da emotività ed ansia nei processi di scelta, questo a conferma di quanto questa esposizione mira ad evidenziare, ovvero che scarsa conoscenza dei mercati e delle dinamiche che li regolano, unita ad un'eccessiva confidenza nelle proprie conoscenze e capacità, spinge gli investitori che non si avvalgono di consulenti specializzati ad intraprendere scelte di collocamento dei propri capitali errate, emotive i cui profitti molto spesso si rivelano randomici e fortunosi. Rilevanti ed esplicativi, inoltre, sono le percentuali risultanti dalle indagini su scelte ed abitudini di investimento: del 45% degli intervistati detenente uno o più strumenti finanziari, più del 50% elabora le proprie scelte servendosi del supporto di famigliari, conoscenti e colleghi (questo fenomeno viene definito *consulenza informale*), il 25% di essi opera in totale autonomia e solo il restante 25% gestisce i propri risparmi personali servendosi di consulenza specializzata o affidando questi ultimi ad intermediari finanziari. Si evince infine come la partecipazione ai mercati finanziari sia correlata positivamente con l'attitudine alla sopravvalutazione delle proprie competenze e conoscenze nonché all'ottimismo ed alla fiducia di poter operare in maniera profittevole (Liciano, Gentile e Soccorso, 2017).

Il quadro appena descritto, seppur rappresentativo di un campione limitato, sembra cogliere in modo soddisfacente lo spaccato degli investitori *retail* italiani. Tale indagine comporta una correzione necessaria a quanto affermato precedentemente ("tutti possono investire"), poiché è evidente che *non tutti possono investire*. La conoscenza anche solo delle nozioni base della terminologia finanziaria sembra non appartenere a buona parte degli investitori, nonostante i mezzi di comunicazione siano ora molteplici ed offrano la possibilità di fruire di formazione di base sulle tecniche di investimento e sulle teorie economiche dei mercati. L'opinione di chi scrive rimane comunque quella che, senza una solida preparazione in materia, usufruire dei mezzi che intermediari e piattaforme di trading mettono a disposizione non è consigliato e risulterebbe oltremodo deleterio per le finanze personali.

2. LA TEORIA DEI MERCATI: DISAMINA ED APPLICABILITÀ

La teoria dei mercati finanziari è stato l'argomento centrale delle pubblicazioni e discussioni economiche di tutta la seconda metà del secolo scorso, e riveste tutt'ora materia di altissimo interesse. Come si è avuto modo di apprendere leggendo il capitolo precedente, il fenomeno del *trading online* va storicamente di pari passo con l'ampliamento progressivo, sia in termini di volumi che in termini partecipativi, dei mercati finanziari e perciò studiosi ed esperti hanno cercato, nel corso del tempo, di teorizzarne l'andamento e le leggi che li regolano. Il presente capitolo si pone dunque l'obiettivo di presentare la celeberrima *teoria dei mercati efficienti* (dall'inglese, *Efficient Markets Hypothesis*, d'ora in avanti indicata con l'acronimo EMH), sviluppata inizialmente in maniera indipendente da Paul Samuelson e da Eugene Fama negli anni '60: tale teoria ha permeato ed influenzato studi e pubblicazioni economiche per tutto il trentennio successivo alla sua introduzione, e di seguito verranno esposti i suoi principali argomenti e trattazioni. Tale teoria, come si avrà modo di appurare, presenta delle assunzioni che si sono rivelate di difficile verifica empirica: in particolare, la maggior critica a tale teoria proviene da studiosi di finanza comportamentale sulla natura del comportamento umano, che viene assunto *razionale* nella teoria dei mercati efficienti. Per questo la letteratura economica, negli ultimi anni, ha cercato di percorrere strade alternative che evolvessero la teoria proposta da Fama e ne affrontassero i limiti in modo tale da conciliare l'efficienza dei mercati con le peculiarità comportamentali degli agenti di mercato. In questo elaborato verrà considerata e trattata la *teoria dei mercati adattivi* (traduzione dell'inglese *Adaptive Markets Hypothesis*, AMH), introdotta dallo studioso Andrew J. Lo, la quale risulta essere più adeguata a descrivere il comportamento dei prezzi e degli agenti che operano nel mercato.

Il punto di partenza di tale disamina è necessariamente il concetto di “mercato efficiente” ed il suo significato nozionistico. Una volta districato il concetto, si passerà ad argomentare i principali risultati esposti negli articoli più rinomati che Fama e Samuelson pubblicarono negli anni tra il 1965 ed il 1970, le successive critiche alle EMH ed il fondamentale contributo di Andrew Lo.

2.1 Il concetto di “mercato efficiente”

Il termine *efficiente*, per uno studioso di economia o pratico della materia, ha un significato molto semplice: secondo la definizione classica, l’allocazione delle risorse di un mercato si definisce *efficiente* (Pareto ottima) se non esiste un’allocazione alternativa tale da rendere migliore la condizione di alcuni individui senza peggiorarne la condizione di alcuni altri. Il criterio di *efficienza Pareto* è uno dei principi fondanti del liberismo economico, sul quale questo lavoro non mira però a soffermarsi, rimandando alla lettura di altri testi eventuali approfondimenti in materia¹³. Stiglitz (1981) distingue tre livelli di analisi riguardanti l’efficienza del mercato, ovvero:

- *Exchange efficiency* (efficienza degli scambi), raggiunta laddove, dati come noti l’insieme di attività e le opinioni ed informazioni dei diversi partecipanti al mercato, tali attività vengono scambiate in modo tale che non sia possibile alcuna riallocazione degli stessi che incrementi l’utilità attesa di un individuo senza alcun decremento dell’utilità attesa di un altro¹⁴.
- *Production efficiency* (efficienza produttiva), concernente la determinazione della disponibilità degli *assets* presenti sul mercato, date come note le risorse

¹³ Si veda, ad esempio, Barucci, E. *Teoria dei mercati finanziari: equilibrio, efficienza, informazione*, pagg. 17-21 e pagg. 52-57.

¹⁴ La funzione di utilità attesa indirizza le scelte di un individuo, ordinate in base alle proprie preferenze. Si consideri, ad esempio, una generica lotteria X i cui possibili esiti siano x_1 e x_2 con rispettive probabilità di realizzarsi pari a p ed a $1 - p$, dunque ogni individuo associa una propria funzione di utilità a tutti i possibili esiti del mondo – nel caso dell’esempio, $u(x_1)$ ed $u(x_2)$ – tali che la funzione di utilità attesa dell’individuo j -esimo è data dalla seguente funzione di Vonon Neumann-Morgensten:

$$U_j(X) = p \times u_j(x_1) + (1 - p)u_j(x_2) = E(u_j(X)).$$

tecnologiche ed informative. L'efficienza produttiva in questo caso viene raggiunta se le imprese massimizzano il proprio valore di mercato organizzando in modo ottimale le risorse a loro disposizione, se la volontà unanime degli *shareholders* è quella che le imprese massimizzino il loro valore di mercato e se esiste un qualche meccanismo di controllo che assicuri che il management delle imprese persegua le politiche aziendali nell'interesse degli *shareholders*.

- *Information efficiency* (efficienza informativa), conseguita se il mercato fornisce in maniera appropriata e consistente tutta l'informazione disponibile, se i prezzi di mercato riflettono integralmente ed immediatamente tali informazioni nel loro andamento e se le imprese sono in grado di trasmettere in maniera efficiente tutte le nuove informazioni, concernenti le proprie potenzialità di crescita future, ad investitori e ad eventuali nuovi *shareholders*.

Il terzo livello di efficienza definito da Stiglitz, ovvero l'efficienza informativa, è una delle colonne portanti della teoria dei mercati efficienti proposta da Fama, e da ora in avanti ci si riferirà ad *efficienza di mercato* relativamente a questo livello di trattazione. Secondo Lo (2007), una delle prime definizioni formali del termine "mercato efficiente" applicata al contesto dei mercati finanziari può essere ricondotta a Fama (1965), ed è la seguente:

«an "efficient" market is defined as a market where there are large numbers of rational, profit-maximizers actively competing, [...] and where important current information is almost freely available to all participants».

Ball (1989) riporta invece la definizione di Fama, Fisher, Jensen e Roll di qualche anno seguente come testimonianza di prima formale definizione di mercato efficiente, ovvero:

« [...] independence of successive price changes is consistent with an "efficient" market, i.e., a market that adjust rapidly to new information».¹⁵

¹⁵ Fama, E.F, L. Fisher, M.C. Jensen, e R. Roll. «The adjustment of stock prices to new information.» *International Economic Review* (Wiley) 10, n. 1 (1969): pag. 1.

Argomentazioni sulla teoria del mercato efficiente, in realtà, risalgono a circa un secolo prima dei lavori dei sopracitati Fama *et al.* (1969), in particolare riconducibili ad un broker della Borsa di Parigi il quale si diletta di teoremi matematici, Jules Regnault: nella sua pubblicazione “Calcul des chances et philosophie de la bourse” del 1863 egli scrisse che il prezzo di un titolo, ad un certo istante di tempo t , riflette pienamente “la saggezza della folla”¹⁶. In altri termini, Regnault arrivò ad affermare che l’unica possibilità di guadagnare un extra profitto è basare le proprie scelte di *trading* su informazioni private, se non esclusive, possedute da una cerchia ristretta di individui. Un altro contributo che è doveroso citare è la prima verifica empirica per dimostrare che i prezzi seguono un modello *random walk* che Bachelier, statistico e matematico francese vissuto a cavallo tra l’800 ed il ’900, effettuò nel 1900. Il suo “teorema fondamentale” riguardo al comportamento dei prezzi, tuttavia, fu che la previsione degli stessi sia un *fair game*¹⁷ e che dunque i profitti attesi al tempo $t+1$ siano zero (Fama, 1969). Ad ogni modo, la letteratura accademica riconosce il fiorire di trattazioni riguardanti l’efficienza dei mercati nel periodo situato tra gli anni ’60 e gli anni ’70 del secolo scorso, in ragione di discussioni corpose ed approfondite che svilupparono le teorie in parte trattate dai due studiosi francesi. Due dei più celebri contributi al dibattito furono indubbiamente gli articoli di Fama e Samuelson del 1965: i due accademici, in maniera del tutto indipendente, svilupparono l’ipotesi di processi *random walk* per i prezzi da due programmi di ricerca piuttosto diversi.

2.2 Gli articoli di Samuelson e di Fama del 1965

Il contributo di Samuelson alle EMH può essere facilmente riassunto dal titolo del suo articolo del 1965, “Proof that Properly Anticipated Prices Fluctuate Randomly”. Premio Nobel nel 1970, Samuelson porta all’attenzione accademica la questione che le EMH

¹⁶ Regnault (1863) in Elton, et al. (2014).

¹⁷ L’espressione *fair game*, ovvero gioco equo, implica che non è in alcun modo possibile ottenere un rendimento sopra le aspettative utilizzando l’informazione disponibile al tempo t . Nelle pagine seguenti verranno esposti i fondamentali costrutti teorici di tale teorema.

implicano solamente che il prezzo di mercato «contains in itself all that can be known about the future and in that sense has discounted future contingencies as much as humanly possible» (Samuelson, 1965). L'economista, nel suo articolo, presenta la composita dimostrazione a tale affermazione, i cui passaggi chiave vengono riportati di seguito. Si consideri, innanzitutto, la sequenza temporale dei prezzi di un titolo come $\{\dots, X_{t-1}, X_t, X_{t+1}, \dots, X_{t+\tau}, \dots\}$; conoscendo solamente la sequenza di prezzi passati ed il prezzo corrente, ovvero $[X_t, X_{t-1}, X_{t-2}, \dots]$, e dunque non conoscendo nessun prezzo futuro $X_{t+\tau}$, la sola assunzione possibile è che esista una distribuzione di probabilità per ogni prezzo futuro, la cui forma è indipendente da quanti periodi τ vengono scelti per prevedere il prezzo $X_{t+\tau}$, ovvero la seguente:

$$P\{X_{t+\tau} \leq X | X_t = x_0, X_{t-1} = x_1, \dots\} = P(X, x_0, x_1 \dots; \tau), \quad (1)$$

la quale implica di fatto l'assunzione di un processo stazionario. Asserito ciò, è possibile relazionare tutte le probabilità, per ogni ampiezza τ futura, alla probabilità base $P(X, x_0, x_1 \dots; \tau = 1)$, ovvero:

$$P(X, x_0, x_1 \dots; \tau) = \int_{-1}^1 P(X, y, x_0, x_1, \dots; \tau - 1) dP(y, x_0, x_1, \dots; \tau = 1), \quad (2)$$

implicando dunque che qualsiasi previsione futura è fondamentalmente deducibile dalla probabilità $P(X, x_0, x_1, \dots; \tau = 1)$. Si definisca inoltre, dati X_t il prezzo al tempo t ed $X_{t+\tau}$ il prezzo che avrà il bene al tempo $t + \tau$, il prezzo *futuro* $Y(\tau, t)$, ovvero la quotazione al tempo t del prezzo del bene $X_{t+\tau}$. Dopo un periodo, la sequenza dei prezzi conosciuta sarà $(X_{t+1}, X_t, X_{t-1}, \dots)$ e non solamente (X_t, X_{t-1}, \dots) perciò il prezzo *futuro* sarà $Y(\tau - 1, t + 1)$. Costruendo ricorsivamente le due sequenze $\{X_{t+\tau}\}$ ed $\{Y(\tau - n, t + n)\}$, e considerando che al tempo $t + \tau$ vale la condizione di arbitraggio (al netto di eventuali commisioni) tale per cui

$$Y(0, t + \tau) = X_{t+\tau}, \quad (3)$$

Samuelson definisce l'assioma del valore atteso del prezzo¹⁸ ("axiom of Mathematically expected price formation") come segue:

$$Y(\tau, t + \tau) = E[X_{t+\tau} | X_t, X_{t-1}, \dots] \text{ con } \tau = 1, 2, \dots \quad (4)$$

Applicando questi costrutti teorici, l'economista afferma che i prezzi dei futures siano basati sull'informazione disponibile nell'esatto momento in cui vengono stabiliti e non vi è possibilità alcuna di speculazione. Samuelson infatti esplicita il *Theorem of Fair-Game Futures Pricing*, la cui riformulazione può essere espressa come segue:

Se i prezzi $\{X_t\}$ sono soggetti alle leggi di probabilità (1) e (2), e la sequenza dei prezzi futuri $\{Y(\tau, t + \tau), Y(\tau - 1, t + \tau - 1), \dots, Y(0, t + \tau)\}$ è soggetta agli assiomi del valore atteso dei prezzi (3) e (4), allora tale sequenza è un *gioco equo* nel senso di presentare cambiamenti di prezzi non distorti (*unbiased*), ovvero

$$E[Y(\tau - 1, t + \tau - 1) | X_t, X_{t-1}, \dots] = Y(\tau, t) \quad (5)$$

oppure, allo stesso modo, considerando $Y(\tau - 1, t + \tau - 1) - Y(\tau, t) = \Delta Y(\tau, t)$,

$$E[\Delta^n Y(\tau, t)] = 0 \text{ dove } n = 1, 2, \dots, \tau. \quad (6)$$

In questo modo, non vi è alcuna possibilità di ottenere un extra profitto dall'analisi dei cambiamenti di prezzo passati, «by chart or any other esoteric devices of magic or mathematics» (Samuelson, 1965). Un'ulteriore e diretta implicazione di questa teoria è che i prezzi futuri, rispetto al set informativo a disposizione degli investitori al tempo t ,

¹⁸ La condizione di partenza è la trascurabilità dei tassi di interesse e dei profili di rischio dei singoli investitori. Samuelson arriva agli stessi risultati anche dopo aver riformulato l'assioma includendo un tasso d'interesse *free risk* positivo ed un certo livello di *risk aversion* degli investitori. Ad ogni modo, l'intento di questa esposizione è dimostrare come Samuelson giunga ad una prima teorizzazione di *fair game*, rimandando alla lettura del sopra citato articolo le dimostrazioni di questa seconda asserzione e del modello generale.

definito con Φ_t , seguono un modello *martingale* (si veda pag. 31 per la formalizzazione di tale modello).

Eugene Fama, quasi simultaneamente alla divulgazione di Samuelson, pubblica l'articolo "Random Walks in Stock Market Prices", di stampo volutamente non tecnicistico ma di importanza fondamentale, nel quale teorizza che i prezzi dei titoli seguono, generalmente, un modello *random walk*. L'analisi di Fama parte da alcune considerazioni di partenza distanti da quelle esaminate da Samuelson, ovvero l'annoso dibattito tra gli analisti *tecnici* e gli analisti sostenitori della teoria del *fondamentale*, entrambi orientati alla previsione del comportamento futuro dei prezzi da due diversi punti di vista: l'analisi tecnica, basata sulle asserzioni di Charles Dow¹⁹, presuppone che l'evoluzione dei prezzi ed il loro comportamento passato possano essere utilizzati per prevederne l'andamento futuro (teoria che Fama definisce impermeata di un *certo grado di misticismo e analoga all'astrologia*); l'analisi fondamentale trae invece spunto dal metodo del *valore intrinseco*, la cui assunzione principale presuppone che ogni prezzo è equivalente al valore attuale di tutti i possibili dividendi futuri e la conoscenza di tale valore intrinseco permette un confronto tra quest'ultimo ed il prezzo odierno di un titolo, al fine di valutare se la quotazione sia *sottostimata* o *sovrastimata*²⁰. La teoria di un mercato *random walk*, asserisce Fama, è invece la più coerente con l'ipotesi di un mercato *efficiente* secondo la definizione dello stesso Fama vista in precedenza: tale comportamento dei prezzi implica che i vari cambiamenti nel corso del tempo non conservano memoria e, di conseguenza, i movimenti passati non hanno alcuna potenza informativa per poter prevedere significativamente i movimenti futuri di prezzo, questo in totale contrasto con quanto teorizzato dai fondamenti dell'analisi tecnica. Inoltre, eventuali discrepanze di prezzo, continua Fama, tra il valore intrinseco ottenuto con

¹⁹ Charles Dow (1851 – 1902), giornalista statunitense è stato il fondatore del Wall Street Journal, uno dei più rinomati quotidiani finanziari a livello mondiale. Il suo più grande contributo fu quello di istituire l'indice Dow Jones Industrial Average in collaborazione con Edward Jones, famoso matematico statunitense, che sicuramente oggi è il più celebre indice della borsa NYSE di New York. Dow negli anni sviluppò teorie sull'evoluzione dei prezzi nei mercati finanziari che presentò in una serie di articoli pubblicati sul Wall Street Journal e che divennero la base dei principi di analisi tecnica.

²⁰ Tale modello è definito *dividend-discount model* ed è il modello base per valutare il valore intrinseco di un'azione.

l'analisi fondamentale ed il valore del prezzo al tempo di valutazione, sotto l'ipotesi di mercato efficiente, vengono *istantaneamente* livellate dal comportamento degli agenti di mercato, i quali operano seguendo le indicazioni derivanti da tale analisi spingendo così il prezzo verso il suo valore intrinseco (chiaramente tale ipotesi è assumibile considerando un mercato ideale, senza costi di transizione e senza asimmetrie informative). L'analisi fondamentale può essere di una certa utilità solo laddove gli analisti siano in possesso di nuove informazioni non ancora correttamente e pienamente considerate nella costruzione dei prezzi correnti di mercato. Fama, nel suo articolo del 1965, non espone empiricamente assunzioni e test sull'effettiva indipendenza temporale dei prezzi, al contrario ne riporta le maggiori conclusioni e di come le principali risultanze supportino la sua tesi.

Fama e Samuelson, dunque, giungono alle stesse evidenze empiriche: più un mercato è *efficiente*, più la sequenza dei cambiamenti di prezzo generata da tale mercato è randomica. Di conseguenza, un mercato *pienamente efficiente* è un mercato nel quale i cambiamenti di prezzo sono completamente imprevedibili e casuali.

2.3 La teorizzazione di Fama delle EMH nel 1969

Il contributo più completo alle EMH si raggiunge con la pubblicazione di Fama del 1969, "Efficient Capital Markets: A Review of Theory and Empirical Work", il massimo sunto della teoria dei mercati efficienti da un punto di vista sia teorico che empirico. Fama ripercorre i principali risultati della letteratura economica sul tema, sintetizzando la definizione di mercato efficiente come segue:

«in an efficient market prices "fully reflect" all available information».

Assunto Φ_t l'insieme informativo disponibile al tempo t , è possibile interpretare tale insieme in tre differenti casistiche, le quali costituiscono la base di tutte le ricerche empiriche effettuate ed individuano tre diverse forme di efficienza di mercato:

- I. *Efficienza in forma debole*, dove Φ_t rappresenta il set di informazioni limitatamente all'andamento della serie storica dei prezzi, senza incorporare altre informazioni (storicamente, questa fu la prima forma di efficienza ipotizzata ed empiricamente testata). La diretta implicazione di tale ipotesi è che vi possono essere agenti che battono sistematicamente il mercato utilizzando strategie che sfruttino informazioni non sintetizzate ed incorporate nel comportamento storico delle serie dei prezzi, delineando una situazione di *asimmetria informativa*.
- II. *Efficienza in forma semi-forte*, dove Φ_t rappresenta il set di informazioni pubblicamente disponibili (ad esempio split azionari, fusioni, acquisizioni, report annuali ecc). In questa seconda forma di efficienza, non è possibile ottenere degli extra profitti investendo nel mercato, a meno di non poter sfruttare informazioni private che permettano di assumere una posizione *monopolistica* rispetto ad esso, poiché in possesso di notizie che nessun altro investitore è in grado di recepire.
- III. *Efficienza in forma forte*, dove Φ_t rappresenta l'insieme informativo totale, includendo tutte le informazioni disponibili, siano esse pubbliche o private. In questo ultimo caso, nessun trader è in grado di guadagnare extra profitti in misura maggiore rispetto al mercato poiché tutte le informazioni disponibili sono già scontate nei prezzi, il cui andamento non risulta essere prevedibile.

Alla luce di questa distinzione, appare dunque chiaro che con l'espressione *all available information* Fama si riferisca a tutto l'insieme informativo disponibile²¹. L'equilibrio di mercato, nel quale i prezzi *riflettono pienamente* tutta l'informazione disponibile, può

²¹ Ball (1989) riporta la critica di Le Roy alla definizione di Fama, una delle tante pubblicazioni avvenute in risposta al lavoro dell'americano, al quale innegabilmente va dato il merito di aver spostato l'attenzione accademica sul tema dell'efficienza dei mercati, innescando un dibattito estremamente acceso. Tornando alla critica di Le Roy, egli afferma che la definizione di Fama è una tautologia in quanto permette ad ogni tipologia di distribuzione dei rendimenti di essere consistente con l'ipotesi di efficienza. Corretta o meno, la critica di Le Roy dimostra come la definizione di Fama sia alquanto rigida. Successivamente, Fama presenterà una definizione rivisitata di efficienza, nella quale afferma che "the market correctly uses all available information" e dunque la distribuzione congiunta dei prezzi *futures* stabiliti dal mercato è equivalente alla distribuzione corretta ottenuta implicitamente utilizzando il set informativo Φ_t .

essere formalizzato anche in termini dei rendimenti attesi secondo la seguente relazione:

$$E(\tilde{p}_{j,t+1}|\Phi_t) = [1 + E(\tilde{r}_{j,t+1}|\Phi_t)]p_{j,t} \text{ dove } n = 1,2, \dots, T \quad (7)$$

dove $p_{j,t}$ è il valore del prezzo j al tempo t , $E(\tilde{p}_{j,t+1}|\Phi_t)$ è il valore atteso della stima del prezzo $\tilde{p}_{j,t+1}$ al tempo t del titolo j -esimo, ed $\tilde{r}_{j,t+1}$ è la stima del rendimento del periodo considerato²². Il valore atteso, chiarisce Fama, è solo una delle misure sintetiche di una distribuzione dei rendimenti e l'efficienza dei mercati non implica formalmente l'utilizzo del *momento primo* della distribuzione come metodo esplicativo per teorizzarne le principali implicazioni. Ad ogni modo, questo sembra essere il viatico necessario per fornire alla disamina delle EMH un certo contenuto empirico (Fama, 1969). Dalla relazione (7) si deduce inoltre che, qualsiasi sia il modello applicato per il calcolo dei rendimenti attesi, il set informativo Φ_t è pienamente utilizzato per determinare i rendimenti attesi di equilibrio.

Tale ultima affermazione ha un'ulteriore implicazione empirica: se le informazioni in Φ_t sono pienamente utilizzate per determinare i rendimenti di equilibrio, allora è esclusa la possibilità per qualsiasi *trading system* basato sul set Φ_t di ottenere extra profitti rispetto a quelli di mercato. Si consideri infatti l'*excess market value* del titolo j -esimo così definito:

$$\tilde{x}_{j,t+1} = \tilde{p}_{j,t+1} - E(\tilde{p}_{j,t+1}|\Phi_t). \quad (8)$$

Il suo valore atteso, condizionatamente ad Φ_t , sarà:

$$E(\tilde{x}_{j,t+1}|\Phi_t) = 0. \quad (9)$$

Analogamente, in termini di rendimenti attesi, si possono esprimere le seguenti proprietà:

²² Si ricorda che il rendimento uniperiodale è pari a $(p_{j,t+1} - p_{j,t})/p_{j,t}$.

$$\tilde{z}_{j,t+1} = \tilde{r}_{j,t+1} - E(\tilde{r}_{j,t+1}|\Phi_t) \quad (10)$$

$$E(\tilde{z}_{j,t+1}|\Phi_t) = 0. \quad (11)$$

La sequenza $\{z_{j,t}\}$, secondo le equazioni (10) ed (11), risulta essere un *gioco equo* rispettivamente alla sequenza informativa $\{\Phi_t\}$. Tale risultato è una diretta implicazione delle assunzioni esposte precedentemente, ovvero che le condizioni di equilibrio di mercato possano essere descritte in termini di rendimenti attesi e che il set informativo Φ_t sia completamente utilizzato per la determinazione dei rendimenti attesi di equilibrio. Fama dunque raggiunge le stesse conclusioni di Samuelson, per altro riconoscendogli il merito di essere stato il primo ad individuare il ruolo che il modello a *gioco equo* possiede nella formalizzazione della teoria dei mercati efficienti ed a studiarne rigorosamente le implicazioni²³.

2.3.1 I modelli submartingale e random walk come implicazioni del *fair game*

In base alle proprietà appena esposte, possono essere assunte due ulteriori specificazioni del modello dei prezzi, ovvero dei rendimenti di equilibrio: il modello *submartingale* ed il modello *random walk*, entrambi aventi un ruolo fondamentale in tutta la letteratura empirica (Fama, 1969). Tali esiti vennero già argomentati, come descritto in precedenza, da Samuelson e Fama nel 1965, ad ogni modo il contributo che l'economista fornisce nel 1969 è riepilogativo e di una certa utilità, poiché racchiude in un unico articolo la teoria dei mercati efficienti nelle sue asserzioni teoriche e dimostrazioni empiriche.

La sequenza dei prezzi $\{p_{j,t}\}$, per essere considerate un submartingale, deve soddisfare la seguente proprietà:

$$E(\tilde{p}_{j,t+1}|\Phi_t) \geq p_{j,t} \text{ o equivalentemente } E(\tilde{r}_{j,t+1}|\Phi_t) \geq 0, \quad (12)$$

²³ Fama cita anche Benoit Mandelbrot e il suo "Forecasts of Future Prices, Ubiased Markets and Martingale Models" del 1966 come altro contributo fondamentale allo studio dei modelli a *gioco equo*.

assunta valida, nell'espressione (7), per ogni t e per ogni Φ_t . Qualora la disuguaglianza in (12) si trasformi in un'uguaglianza, allora il modello sarà un martingale. Tale proprietà ha una diretta implicazione empirica, ovvero che non è possibile sviluppare alcuna regola di trading capace di sovraperformare, in termini di profitto, una strategia basata semplicemente sul *buy and hold* che permanga durante tutto l'*holding period* considerato, proprio per la proprietà (12) che dimostra come rendimenti attesi condizionali a Φ_t sono sempre maggiori o uguali a zero.

Del modello *random walk* si è già discusso in precedenza, ad ogni modo è bene ricordare come alla base di tale ipotesi vi siano due condizioni necessarie:

- I. cambiamenti di prezzo consecutivi o, più precisamente, rendimenti attesi uniperiodali, indipendenti;
- II. cambiamenti di prezzo, o rendimenti attesi, identicamente distribuiti.

A livello formale, il modello *random walk* implica la seguente proprietà:

$$P(r_{j,t+1} | \Phi_t) = P(r_{j,t+1}) \quad (13)$$

$$E(\tilde{r}_{j,t+1} | \Phi_t) = E(\tilde{r}_{j,t+1}) \quad (14)$$

dove l'equazione (13) indica che per tale modello la distribuzione di probabilità condizionale coincide con la distribuzione di probabilità marginale, mentre l'equazione (14) indica che il rendimento atteso $\tilde{r}_{j,t+1}$ è indipendente da Φ_t . Questo non significa, come asserisce Fama, che la distribuzione dei rendimenti passati non abbia alcuna utilità per stimare la distribuzione dei rendimenti futuri, supposto che tale distribuzione sia stazionaria e dunque sia la miglior fonte a disposizione. Il modello *random walk* asserisce che è la sequenza, o meglio, l'ordine dei rendimenti passati ad essere insignificante per la stima dei rendimenti futuri (Fama, 1969). Tale modello, conclude Fama, risulta essere un'estensione del modello a *gioco equo* dei mercati efficienti: infatti il modello a *gioco equo* presuppone semplicemente le due condizioni viste precedentemente riguardo ai rendimenti attesi ed al set informativo Φ_t , il modello *random walk* invece presuppone delle ipotesi più forti.

2.3.2 Condizioni del mercato consistenti con le EMH

Per la sussistenza delle EMH vengono individuate, da Fama, delle condizioni *sufficienti* ma non necessarie, concernenti il mercato dei capitali:

- I. assenza di costi di transazione nelle operazioni di compravendita di titoli;
- II. accessibilità a tutte le informazioni (pubbliche e private) senza alcun tipo di costo per i partecipanti al mercato, quindi perfetta *simmetria* informativa;
- III. omogeneità nel considerare l'impatto che Φ_t possiede sui prezzi correnti e sulla distribuzione dei prezzi futuri di ogni titolo disponibile sul mercato.

Se queste tre condizioni vengono rispettate, il mercato è efficiente ed i prezzi *riflettono pienamente* tutte le informazioni disponibili al tempo t . Tali condizioni sono sufficienti ma non necessarie, ovvero la loro sussistenza o meno non implica direttamente l'efficienza o la non efficienza dei mercati. Un mercato può essere efficiente anche solamente se «sufficient numbers of investors have ready access to available information» (Fama, 1969). Chiaramente, tali ipotesi sono delle forzature poiché nel mercato reale esistono costi di transazione, alcuni *traders* possiedono un set informativo migliore o sfruttano meglio le informazioni a loro disposizione (quindi l'informazione ha un costo, non necessariamente monetario) e non sempre vi è omogeneità di opinioni riguardo la distribuzione futura dei prezzi. Fama persevera nella sua linea a favore delle EMH definendo situazioni di tal genere come fonti potenziali di inefficienza e non inefficienza di per sé, portando a supporto delle sue affermazioni i test empirici effettuati, considerando le tre diverse forme di efficienza (debole, semi-forte, forte).

2.4 Test empirici sulle EMH

Non vi è alcun dubbio che la letteratura empirica riguardo le EMH abbia avuto una notevole proliferazione, in particolar modo dopo la formalizzazione di Fama appena analizzata. Si può, nemmeno troppo forzatamente, distinguere tale letteratura in due diversi blocchi, aventi come spartiacque l'articolo dell'economista statunitense nel

1969, pertanto si può disquisire sui test empirici *pre Fama* e *post Fama*. Questo lavoro non ha come obiettivo l'esposizione sistematica e dettagliata dei test empirici riguardanti le EMH, tuttavia non può prescindere da considerarne i principali risultati, argomentando in tal merito alcune riflessioni. Innanzitutto, è bene ricordare che le prime ricerche e diatribe sull'efficienza del mercato, a partire dal già citato Bachelier, erano finalizzate a ricercare evidenze che provassero la sussistenza di un modello *random walk* per i prezzi di equilibrio. Lo (2007) ripercorre i principali lavori di ricerca sulle ipotesi di mercati *random walk*: il primo contributo proviene da Cowels e Jones i quali, nel 1937, compararono la frequenza di *sequenze* ed *inversioni* nelle serie storiche dei rendimenti del mercato statunitense, dove le prime sono coppie di rendimenti consecutivi con il medesimo segno (+ + o - -) mentre con le seconde si intendono coppie di rendimenti consecutivi con segno opposto (+ - o - +). I due studiosi arrivarono a validare l'ipotesi *random walk*, così come fece Kendall nel 1953: lo studioso, esaminando i comportamenti delle variazioni settimanali di 19 indici del settore industriale britannico, dopo diverse analisi riguardo la presenza di correlazioni seriali nelle osservazioni, affermò – in termini scenografici – che

«the series looks like a wandering one, almost as if once a week the Demon of Chance drew a random number from a symmetrical population of fixed dispersion and added it to the current price to determinate the next week's price». ²⁴

Gli appena citati autori, assieme a molti altri, basarono i propri studi empirici su osservazioni storiche senza però fornire dimostrazioni convincenti a supporto dei fondamenti logici usati per la presentazione delle ipotesi. È soltanto con Samuelson (1965) e Mandelbrot (1966) che i test empirici assumono consistenza teorica ed il ruolo del modello per i rendimenti attesi a *gioco equo* comincia a prendere piede, così come si comincia a considerare e studiare in maniera rigorosa le relazioni tra i modelli a *gioco equo* ed i modelli *random walk*, espresse in modo molto chiaro e semplice da Fama.

²⁴ Kendall (1953), in Fama (1969).

Fama presenta la mole consistente di lavoro empirico suddividendolo in tre diverse categorie, definite in base alle ipotesi riguardanti la natura del set informativo Φ_t . Delle tre macro classi di test presentati (efficienza debole, semi-forte e forte), quella relativa all'efficienza debole è certamente la più voluminosa e corposa e, asserisce Fama, supporta fortemente l'ipotesi di mercati efficienti: nonostante venga individuata da Fama una certa correlazione seriale positiva tra le osservazioni giornaliere di prezzi²⁵, queste evidenze secondo Fama non sono sufficientemente consistenti per dimostrare l'inefficienza dei mercati. Inoltre, lo studioso attesta che, anche qualora vi fosse una certa dipendenza infra giornaliera tra i prezzi delle azioni ed esistesse un certo numero di *traders* in grado di sfruttarla, nella quasi totalità dei casi il numero di transazioni necessarie per ottenere un profitto rilevante sarebbe tale che i costi di transazione dovuti per eseguire queste operazioni pareggerebbero il profitto ottenuto. Tuttalpiù, i test empirici sull'efficienza in forma debole provano, in alcuni casi, l'insussistenza dell'ipotesi *random walk* per la serie dei prezzi, ma non vi sono evidenze sufficienti a rifiutare la più rilevante ipotesi del modello a *gioco equo*. I test concernenti l'efficienza in forma semi-forte confermano, secondo il lavoro di Fama, le EMH: in particolare, viene dimostrato²⁶ come l'informazione contenuta negli split azionari riguardante i possibili dividendi futuri corrisposti dalle imprese è, in media, pienamente riflessa nel prezzo dell'azione al momento esatto dello split. Infine, i test riguardanti l'efficienza in forma forte vengono considerati da Fama come una sorta di *benchmark* di riferimento per valutare eventuali deviazioni dalle EMH, vista la componente forzosamente teorica e poco realizzabile della forma di efficienza. Chiaramente, tale ipotesi non viene empiricamente validata in quanto si evincono due principali deviazioni dalla condizione di efficienza, ovvero l'accesso monopolistico ad alcune informazioni che gli specialisti di borsa e i *corporate insiders* posseggono riguardanti le movimentazioni dei principali titoli, in particolare la loro possibilità di conoscere i limiti di ordini di vendita e acquisto ancora non eseguiti e conseguentemente la possibilità di generare profitti sfruttando queste informazioni, rendendoli *market makers* di fatto. Fama stoicamente persegue la sua linea a favore anche di questa ultima forma di efficienza, delimitando a solo queste

²⁵ Si veda Fama (1969), pag. 393 e seguenti.

²⁶ Fama, E.F, L. Fisher, M.C. Jensen, e R. Roll. «The adjustment of stock prices to new information.» *International Economic Review* (Wiley) 10, n. 1 (1969) in Fama (1969).

due categorie l'accesso monopolistico ad informazioni riservate ed affermando di conseguenza che «for the purposes of most investors the efficient markets model seems a good first (and second) approximation to reality»²⁷.

Le varie casistiche riguardanti le EMH vengono esposte linearmente da Fama, il quale sintetizza il lavoro empirico precedente a lui, plasmandolo e strutturandolo in maniera organizzata. L'estrema sintesi del suo operato si può racchiudere nel seguente concetto: più un mercato è efficiente, più la sequenza dei cambiamenti dei prezzi generata da quel mercato è casuale ed un mercato pienamente efficiente è un mercato nel quale le variazioni di prezzo non sono prevedibili in alcun modo; se dunque i prezzi riflettono pienamente tutta l'informazione disponibile e scontano istantaneamente le nuove informazioni prodotte dal mercato, allora i prezzi seguono il modello martingale. Vi furono, nel decennio successivo ai lavori di Fama e Samuelson, svariate pubblicazioni che espansero e svilupparono le EMH, non distanziandosi tuttavia più di tanto dai pilastri della teoria di Fama: gli individui agiscono e producono aspettative in maniera razionale, il mercato aggrega e recepisce le informazioni in modo efficiente, i prezzi di equilibrio scontano tutte le informazioni disponibili istantaneamente (Lo, 2007).

2.5 Critiche e sviluppi empirici delle EMH

Come più volte affermato, il merito di Fama non fu solamente quello di presentare la teoria dei mercati efficienti ma anche l'input che diede, grazie al suo lavoro, alle trattazioni economiche successive, influenzandone fortemente contenuti e pubblicazioni. È solo recentemente infatti che si è cercato di conciliare le ipotesi di efficienza dei mercati con l'economia comportamentale, formalizzando in tal senso nuove ipotesi di efficienza più adatte ad un contesto reale e concreto: fino agli anni '90 del secolo scorso, infatti, l'attenzione era rivolta a dimostrare validità o insussistenza delle EMH da vari punti di vista, e non ad usurparla definitivamente ricercando una valida e più completa alternativa. Vi è pertanto una mole piuttosto considerevole di

²⁷ Fama (1969), pag. 416.

trattazioni e prove empiriche pro ed a sfavore delle EMH, di cui è bene considerare quanto meno le più importanti, spartendole in differenti tipologie di argomenti²⁸.

Per quanto riguarda l'ipotesi *random walk* per descrivere l'andamento dei prezzi di mercato (o equivalentemente dei rendimenti), uno dei lavori più interessanti è quello di Lo e MacKinlay del 1998: i due studiosi, attraverso la costruzione di un test utilizzando come ratio la varianza delle osservazioni settimanali dei rendimenti di alcuni indici del mercato statunitense, individuano una correlazione seriale positiva nelle osservazioni, poiché le varianze analizzate presentavano una crescita più rapida rispetto ad una crescita lineare, man mano che si incrementa l'*holding period*. Tale risultato (correlazione seriale positiva) rigetta l'ipotesi *random walk* per i rendimenti. Lo stesso Lo, nel 1991, presenta l'implementazione di un test per verificare la presenza di memoria a lungo termine nelle serie di prezzi: il risultato conferma tale ipotesi, individuando dunque un certo grado di persistenza a lungo termine nelle osservazioni, rigettando anche in questo caso l'ipotesi *random walk*. Un altro set rilevante di analisi empiriche sulle EMH parte dalla semplice generalizzazione del *dividend discount model*, considerando uno scenario in cui è presente incertezza: il prezzo di mercato al tempo t , in questo caso, è equivalente al valore atteso, condizionale al set informativo disponibile Φ_t , del valore attualizzato di tutti i dividendi futuri, scontati ad un certo tasso *risk-adjusted*²⁹. In base a questo modello, l'errore di previsione, ovvero la differenza tra il valore effettivo del prezzo realizzato in t ed il suo valore atteso condizionale, dovrebbe essere per costruzione incorrelato con il valore atteso condizionale del prezzo. Ciò implica che la varianza del valore attuale del prezzo al tempo t *ex post*, cioè dopo la sua effettiva realizzazione, corrisponde alla somma tra la varianza dell'errore di previsione e la varianza del valore atteso condizionale del prezzo. Da questa decomposizione, data la non negatività della volatilità, deriva la condizione che la varianza dei prezzi di

²⁸ Vengono qui esposti i principali test riportati in Lo (2007) che analizzano le implicazioni delle EMH, presentandone brevemente i risultati che essi comportano. L'intento non è certo quello di fornire una disamina dettagliata delle innumerevoli pubblicazioni *contra EMH*, ma di portare all'attenzione come l'accademia abbia sviluppato prove e test che hanno messo in luce le debolezze delle EMH, evidenziando la necessità di ricercare nuove e più adatte ipotesi di efficienza.

²⁹ Questa generalizzazione, come riporta Lo, è stata sviluppata nel 1981 da Grossman e Shiller.

mercato delle azioni *non può eccedere* la varianza dei valori attuali dei prezzi *ex post*. Come riporta Lo (2007), Shiller nel 1981 effettua dei test empirici utilizzando dati annuali di vari prezzi azionari del mercato USA, trovando che tale limite viene ripetutamente violato, finendo per affermare che i prezzi di mercato presentano un livello tale di volatilità da inficiare le basi delle EMH. Tale risultato suscitò reazioni controverse nel mondo accademico, e lavori successivi a quello di Shiller provarono invece che la violazione empirica di tale limite non necessariamente comporta un rifiuto delle EMH *in toto*, tutt'al più indica un certo grado di avversione al rischio degli investitori o un fenomeno di *smoothing* dei dividendi, ed in alcuni casi, come dimostrano Marsh e Merton nel 1986, qualora i rendimenti seguano un modello *random walk* geometrico, le violazioni dei limiti imposti alla varianza possono essere interpretati come un supporto dell'efficienza dei mercati³⁰. Un ulteriore ambito di trattazione sulla validità empirica delle EMH riguarda le reazioni degli investitori a nuove informazioni disponibili sul mercato: in particolare, spesso vengono effettuate operazioni impulsive di vendita di titoli che nel recente passato hanno registrato performance negative oppure di acquisto di titoli che in precedenza hanno performato positivamente. Questo tipo di reazioni spingono il livello dei prezzi ad oscillare dal suo valore *razionale* di mercato, fin quando investitori razionali non compiono operazioni di mercato opposte che riconducono i prezzi al loro livello di equilibrio. Tali comportamenti provocano *anomalie* nel mercato, molte delle quali seguono uno schema definito e conservano dunque un certo grado di predicibilità: ad esempio, il cosiddetto *January effect*, ovvero il fenomeno secondo il quale titoli aventi volumi di capitalizzazione molto ridotti sovraperformano titoli ad alta capitalizzazione, anche di un margine considerevolmente ampio, nel periodo di inizio anno. Tale anomalia è solo una delle tante analizzate ed individuate da vari studi empirici³⁰, e risulta essere una conferma evidente contro le EMH (che presuppongono agenti razionali di mercato e non prevedibilità dell'andamento dei prezzi).

Infine, è doveroso riportare il risultato del lavoro di Grossman e Striglitz nel 1981 sull'impossibilità di mercati efficienti: i due economisti affermano come, nel caso di mercato perfettamente efficienti, non vi sarebbe alcun impulso per gli investitori a

³⁰ Si veda, per i riferimenti completi, Lo (2007), pagg. 5-9.

spendere risorse per raccogliere informazioni che portino ad un profitto e, proprio per questa ragione, vi sarebbero scarsi incentivi allo scambio di titoli ed i mercati andrebbero inesorabilmente verso il crollo. Dunque, un equilibrio di mercato *non degenerato* si ottiene solamente qualora vi siano opportunità di profitto che compensino gli investitori per i costi di trading e di raccolta delle informazioni, ipotesi che equivalgono ad inefficienza per le EMH e dunque conducono alla loro insussistenza. Argomentazioni a supporto delle EMH hanno portato a controbattere che, sebbene errori comportamentali sintomo di inefficienza siano talvolta ricorrenti, vi è un limite alla loro prevalenza ed impatto sui mercati poiché investitori razionali potranno trarre vantaggio da queste inefficienze ed i loro comportamenti spingeranno nuovamente i livelli dei prezzi di mercato verso l'equilibrio efficiente³¹. Questa ultima affermazione, tuttavia, trae necessariamente supporto dalla convinzione che gli agenti di mercato siano sufficientemente potenti da incorporare, se non istantaneamente, in tempi brevi le aspettative razionali nei livelli di prezzo, o in alternativa che comportamenti inefficienti ed errati non abbiano forza sufficiente per condizionare nel lungo periodo l'equilibrio di mercato. Tale assunzione ha necessariamente bisogno di riscontri empirici basati su inferenza statistica, non è sufficiente darle forma teorica per asserirne la sua validità; inoltre, è possibile che tali comportamenti errati ed irrazionali possano sopraffare agenti razionali anche per un periodo di tempo prolungato (mesi ed in alcuni casi anche anni), discostando di molto i mercati dalla condizione di efficienza declamata dalle EMH.

Risulterà chiaro ora come, ancora oggi, non vi sia un consenso unanime tra gli economisti riguardo le EMH, e queste rivestano ancora importanza a livello accademico. La ragione di ciò è insita nel fatto che le EMH, oltre ad essere una idealizzazione dell'ambiente economico irrealizzabile nel concreto, sono anche un utilissimo *benchmark* per misurare il grado di efficienza di un mercato, o di confrontare il livello di efficienza di due mercati diversi. Inoltre, risultano sorprendentemente resistenti a prove empiriche che ne determinino il rifiuto, proprio per la natura di ipotesi non così ben

³¹ Un esempio a supporto di tale affermazione è il cosiddetto "Dutch book", nel quale si illustra come informazioni irrazionali creino profitti garantiti per investitori che agiscono razionalmente. Si veda Lo (2007), pag. 11.

definite nella loro interezza (Lo, 2004) (Lo, 2007). Tuttavia, il desiderio di ricercare ipotesi teoriche basate su assunzioni più realistiche è stato l'input che ha spinto la letteratura accademica a sviluppare nuove interpretazioni alle EMH in maniera tale da riconciliarle con il comportamento degli individui e degli agenti che popolano il mercato. In questa direzione si è mosso e si muove tutt'ora il lavoro di Andrew W. Lo il quale, utilizzando le sue stesse parole, propone «a new framework that reconciles market efficiency with behavioral alternatives by applying the principles of evolution – competition, adaptation, and natural selection – to financial interaction.» (Lo, 2004).

2.6 Le Adaptive Market Hypothesis di Andrew Lo

2.6.1 Un nuovo ordine mondiale

Prima di inoltrarsi nella disamina della teoria evoluzionistica di Lo, è opportuno avere chiaro il seguente concetto: non è corretto etichettare le EMH come ipotesi coerenti o inconsistenti, ma meramente come ipotesi lacunose, utilizzate come semplificazione di una realtà ben più complessa. Tale incompletezza ha assunto contorni definiti soprattutto a seguito delle recenti crisi economiche: sotto condizioni stabili e di stazionarietà, i mercati presentano un andamento lineare e le EMH sono uno strumento efficace per descrivere la realtà; nel momento in cui si presentano condizioni di instabilità, dinamismo e volatilità le EMH perdono invece validità. Il paradigma fondamentale di investimento, sotto le EMH, è il seguente (Lo, 2012):

- I. La relazione tra rischio e rendimento è lineare.
- II. Tale relazione è statica sia in relazione al tempo che in relazione alle circostanze di mercato.
- III. I parametri che descrivono tale relazione possono essere accuratamente stimati (utilizzando, ad esempio, il CAPM).
- IV. Gli investitori hanno aspettative razionali.
- V. I rendimenti degli assets sono stazionari.
- VI. I mercati sono efficienti.

Durante il periodo storico compreso tra il 1940 e i primi anni 2000, quando si era in presenza di mercati finanziari stabili, tale paradigma poteva essere una discreta approssimazione della realtà. Lo definisce tale periodo storico “The Great Modulation”: la *figura 1* rappresenta la volatilità annualizzata dei rendimenti giornalieri dell’indice CRSP americano (Center for Research in Security Prices) e si può verificare come il periodo indicato presenti una volatilità decisamente attenuata. In seguito, viene registrato uno dei più alti picchi di volatilità dell’ultimo secolo, in coincidenza con la bancarotta di Lehman Brothers e con l’inizio della crisi che dal 2008 ha imperversato sui mercati. I recenti avvenimenti economici hanno dimostrato chiaramente come i mercati dei capitali abbiano subito una delle trasformazioni più radicali della storia moderna, crescendo in volumi, interazioni, partecipanti, questi ultimi cresciuti non solo in termini di numerosità, ma anche per tipologia di preferenze ed attitudini comportamentali.

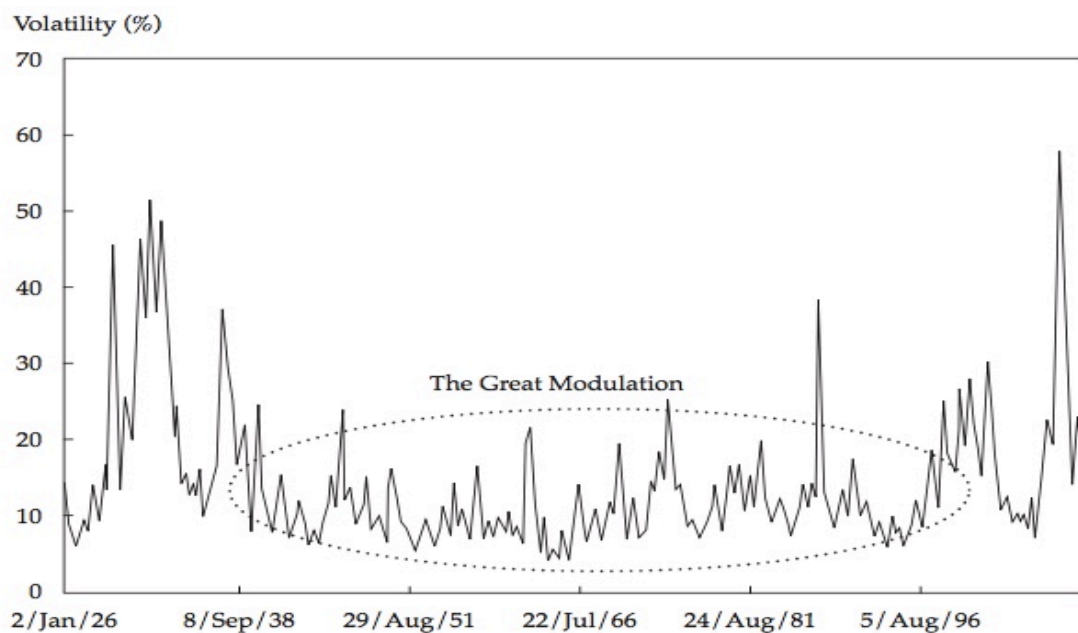


Figura 1: volatilità mobile annualizzata dei rendimenti giornalieri dell'indice CRSP value-weighted, dal 2/1/1926 al 31/12/2010 (Fonte: Lo, 2012).

Tale evoluzione, nella direzione di quello che Lo definisce «a new world order», crea un contesto nel quale le ipotesi di mercato adattivo risultano essere un *framework* migliore per descrivere il comportamento dei mercati finanziari e degli agenti che in essi operano.

2.6.2 L'approccio evoluzionistico ed il comportamento umano

Le critiche più aspre e durevoli alle EMH evolvono proprio attorno al paradigma di razionalità comportamentale che, secondo questa teoria, permea gli individui partecipanti al mercato e il modo in cui essi costruiscono le proprie preferenze: gli investitori agiscono razionalmente ottimizzando la propria utilità attesa dati una serie di parametri noti, come ad esempio il grado personale di avversione al rischio. Come visto nei paragrafi precedenti, l'individuo non sempre assume comportamenti razionali, ma è forzato ad assumere decisioni in contesti molto spesso incerti, condizione che lo conduce a commettere errori comportamentali: numerosi studi hanno infatti dimostrato come il processo decisionale umano non sembra conformarsi ai concetti di razionalità e di efficienza dei mercati, ma esibisce certi errori di comportamento, chiaramente controproducenti se visti da una prospettiva finanziaria, quali ad esempio eccessivo ottimismo nelle proprie capacità di investimento, eccessiva reazione agli eventi del mercato, emulazione, pentimento nelle scelte (Lo, 2005). Tutti comportamenti, questi, impregnati di irrazionalità, guidati da sentimenti impulsivi che sono insiti nella natura umana.

Lo, nei suoi articoli, riporta spesso una versione ristrutturata di un esperimento condotto da Kahneman e Tversky riguardo gli errori comportamentali, il cui sunto può essere il seguente: si supponga di avere due alternative, A e B, la prima prevede un profitto certo di 240.000\$ mentre la seconda è un biglietto di una lotteria la cui probabilità di vincere 1.000.000 \$ è pari al 25%, mentre il restante 75% rappresenta la probabilità di non ottenere nulla. Il valore atteso dell'alternativa B è 250.000\$, più alto rispetto al valore atteso di A, tuttavia l'alternativa scelta dalla maggioranza è sempre la A, dato il comune comportamento di *risk aversion* degli individui in una situazione profittevole. Si considerino ora due ulteriori alternative, C e D, la prima rappresenta una perdita certa di 750.000 \$, la seconda una perdita di 1.000.000\$ con probabilità del 75% ed una perdita nulla con probabilità del 25%. In questo caso, l'individuo è portato a scegliere per l'alternativa più rischiosa D poiché, in una situazione di perdita, assume generalmente un comportamento *risk seeking*. Dunque, gli individui scelgono in base alle proprie preferenze ed avversione al rischio dettata dalle condizioni di mercato

esistenti, ed in questo caso non vi sono scelte apparentemente giuste o sbagliate. Se però cambiano le prospettive di scelta, ecco che vi sono scelte economicamente razionali e corrette, ed in questo caso evidenze empiriche e sperimentali hanno evidenziato che la maggior parte degli individui seleziona l'alternativa errata. Si considerino infatti la combinazione di A e D e la combinazione di B e C, la prima equivalente ad un biglietto della lotteria fruttante 240.000\$ con probabilità 25% e una perdita di 760.000\$ con probabilità 75%, la seconda invece prevede un guadagno di 250.000\$ con probabilità 25% ed una perdita di 750.000\$ con probabilità 75%. Nonostante le alternative B e C combinate prevedano un guadagno maggiore di 10.000\$ e una perdita inferiore di 10.000\$ rispetto alla combinazione delle alternative A e D, gli individui sono portati a scegliere queste ultime per il comune comportamento – errato – di essere *risk averse* in situazioni profittevoli e *risk seeking* in situazioni di perdita. Tali comportamenti irrazionali, dei quali l'esperimento appena esposto ne è un esempio, hanno imperversato nei mercati dei capitali innumerevoli volte: si varia da aneddoti storici molto curiosi quali la “mania dei tulipani”, che tra il 1634 ed il 1636 in Olanda spinse il prezzo dei fiori a livelli vertiginosi, per poi crollare rovinosamente creando enormi dislocazioni di capitali e panico economico, alle ben note e più recenti crisi finanziarie del 1929 e del 1987 nel mercato statunitense, o alla bolla tecnologica dei primi anni 2000.

Come conciliare tali errori comportamentali con le ipotesi di mercato efficiente? La nuova sintesi di Lo, le AMH, trae sostentamento ed ispirazione dal concetto di *psicologia evolutivista* introdotto da E.O. Wilson nel 1975, il quale applica i principi di *competizione, riproduzione e selezione naturale* alle interazioni sociali giungendo a spiegare, mediante questo nuovo paradigma, determinati comportamenti umani. Wilson coniò il neologismo “sociobiologia” per etichettare tali idee. Gli stessi principi ideati da Wilson possono essere applicati anche in contesti economici e finanziari, conciliando così le EMH con la pluralità di alternative comportamentali percorribili dagli individui che popolano tali contesti. Questo nuovo approccio evolutivista *à la Darwin* non è certo nuovo, ed è stato più volte applicato in contesti finanziari. Per citare due dei più importanti contributi riportati in Lo (2004), Niederhoffer nel 1997 descrive i mercati finanziari come un *ecosistema* (intitola infatti il suo capitolo “The Ecology of Markets”)

nel quale gli speculatori sono i “carnivori”, i venditori sono gli “erbivori” ed i *floor traders* e gli investitori stressati sono gli organismi in decomposizione; Bernstein invece argomenta che il concetto di equilibrio, centrale per le EMH, si raggiunga difficilmente in un contesto reale e che le dinamiche di mercato vengono meglio descritte da processi evolucionistici. Tali processi, applicati ai sistemi economici, ambienti fondati e formati da interazioni sociali, prevedono un ruolo centrale per il comportamento umano, il quale si evolve e sviluppa seguendo *euristiche*³², è un comportamento adattivo e non sempre risulta completamente prevedibile. Sotto questa nuova luce, la prospettiva evolucionistica si propone di teorizzare il comportamento umano come condotta finalizzata alla sopravvivenza del proprio materiale genetico, comportamento che *evolve* sulla spinta della selezione naturale e che dipende dall’ambiente nel quale agisce ed è immerso.

Al fine di dare valenza e contestualizzazione economica a tali teorie, è imprescindibile innanzitutto considerare il concetto di “razionalità limitata” coniato da Herbert Simon nel 1955: contrariamente al postulato neoclassico sul comportamento degli individui (massimizzazione dell’utilità attesa in base ad aspettative razionali), Simon afferma che difficilmente l’individuo persegue la massimizzazione della propria utilità, risultato estremamente costoso e difficilmente raggiungibile, ma intraprende scelte e comportamenti che lo portino a raggiungere uno stato di *satisficing* (letteralmente, “soddisfazione”), non necessariamente ottima. Ad un primo impatto, sembrerebbe che il problema di ottimizzazione abbia solo cambiato forma e non sostanza, poiché Simon non chiarisce a quale punto esatto un individuo smette di ricercare l’ottimo e si “accontenta” di una soluzione soddisfacente. Tale limite viene risolto considerando la prospettiva evolucionistica, nella quale il punto di arrivo non viene determinato analiticamente, ma attraverso la selezione naturale ed il metodo *trial and error* (basato su continui tentativi di prova di diverse strategie finché si raggiunge il successo o si smette di tentare): gli individui sviluppano *euristiche* basate sulle proprie esperienze

³² L’euristica (dal termine greco *eurísko*, «scopro» o «trovo») è quel determinato procedimento non formale che prevede la formulazione di ipotesi riguardanti un dato fenomeno o comportamento, cercando poi applicazioni pratiche delle formulazioni ipotetiche compiute.

passate e sulle proprie preferenze, evolvendole a seconda dei *rinforzi*, positivi o negativi, che ricevono come output dall'ambiente circostante. Quelle euristiche che conducono ad una soluzione positiva saranno *ottime* ed avranno applicabilità futura fin tanto che non muteranno nuovamente le condizioni dell'ambiente di riferimento: in questo caso, applicando le medesime euristiche, si registreranno errori comportamentali dettati non da comportamenti irrazionali ma "maladattivi". Ad esempio, il salto di un pesce dall'acqua verso la terraferma può essere considerato come un comportamento irrazionale e controproducente, ma sott'acqua può essere individuato come una strategia vincente per sfuggire ai propri predatori (Lo 2005).

Le Adaptive Markets Hypothesis si presentano come una versione evoluta delle EMH, anche poiché riformulano il paradigma principale di queste ultime ("prices fully reflect all available information"): i prezzi degli strumenti finanziari riflettono tante informazioni quante sono quelle derivanti dall'interazione tra le condizioni ambientali e le specie che popolano l'economia/ecologia, dove per specie si intendono le pluralità di agenti partecipanti al mercato (investitori, *market makers*, istituzioni, *portfolio managers* ed altri). In questo contesto, il grado di efficienza dipende da:

- I. Il numero di competitors che popolano il mercato.
- II. La forza e la consistenza delle opportunità di profitto.
- III. L'adattabilità dei partecipanti al mercato.

Ad esempio, se un numero elevato di "specie" facilmente adattive compete in un mercato con scarse opportunità di profitto (come il mercato dei Treasury Notes, i titoli di Stato del governo federale USA), tale mercato risulta essere quasi perfettamente efficiente, al contrario se un numero ristretto di "specie" compete in un mercato con forti opportunità di profitto e margini di crescita, tale mercato sarà lontano dalla condizione di efficienza, come ad esempio può essere il mercato dei quadri rinascimentali italiani (Lo 2004). Numero di competitors, opportunità di profitto ed adattabilità dei partecipanti non sono costanti nel tempo ma si manifestano a cicli, e la selezione naturale è un deterrente importante poiché seleziona quelle specie che sviluppano euristiche sempre efficaci e che permettono loro di adattarsi continuamente ai cicli di mercato e di *sopravvivere*. Lo sviluppo di strategie di investimento produrrà

dunque profitti o perdite a seconda del cambiamento delle condizioni di business. Tali cicli, in alcuni casi, convergono in situazioni cosiddette *di angolo* e che comportano l'estinzione di quelle specie che non hanno assunto un comportamento sufficientemente adattivo.

In un contesto in cui comportamento ed errori comportamentali sono parte integrante nel processo di *decision making* che porta a scelte ed alle strategie di investimento, l'*emozione* gioca un ruolo fondamentale in tale scacchiera: non è corretto definire un comportamento *emotivo* ed un comportamento *razionale* come due concetti diametralmente opposti, al contrario risultano essere complementare poiché l'emozione è un input fondamentale per sviluppare ed accelerare quei comportamenti adattivi che migliorano il processo di ottimizzazione delle scelte, nonché un potente mezzo per migliorare il processo attraverso il quale gli individui apprendono dall'ambiente circostante e dalle scelte passate. I recenti sviluppi nella neuroscienza cognitiva affermano proprio questo: le emozioni sono la base di un sistema *reward-and-punishment* che facilita la selezione di comportamenti vantaggiosi, fornendo parametri di scelta per una corretta analisi dei costi e dei benefici derivanti dalle azioni compiute. La prospettiva della neuroscienza cognitiva suggerisce dunque un'alternativa rispetto a quella paventata dalle EMH, in cui le forze di mercato e gli individui e le loro preferenze (che non sono stabili nel tempo, ma mutano e si evolvono adattandosi alle condizioni di mercato) interagiscono in un sistema economico molto più dinamico, guidato da selezione naturale, competizione e dalla pluralità di comportamenti assunti da individui ed istituzioni. Questa è la fondamentale essenza delle AMH, i cui risvolti pratici vengono esposti nel prossimo paragrafo.

2.6.3 Applicazioni concrete delle AMH

Il nuovo *framework* di Andrew Lo, dato l'innovativo contesto appena presentato, non può che essere retto dai seguenti pilastri (Lo, 2005):

- I. Gli individui agiscono nel proprio interesse personale (unico punto di contatto con le EMH).

- II. Gli individui commettono errori (non vi è più razionalità comportamentale).
- III. Attraverso il proprio comportamento, imparano dagli errori e si adattano.
- IV. La competizione tra individui è un input fondamentale per adattamento ed innovazione.
- V. La selezione naturale modella l'ecologia di mercato.
- VI. L'evoluzione determina le dinamiche di mercato.

Il carattere pervadente ed influente attribuito al comportamento umano consente un'immediata applicazione pratica: individui ed istituzioni agiscono e si muovono nel mercato sviluppando *preferenze*, in particolare relative alla propensione al rischio, le quali non sono costanti ma si evolvono nel corso del tempo. Una delle sfide delle AMH è quella di riuscire a modellizzare tali *risk preferences* mediante un approccio non più unidimensionale (funzione di utilità attesa) ma mediante un'aggregazione complessa di una pluralità di fattori. L'approccio neuro scientifico suggerisce che, seguendo le analogie con la struttura del cervello umano³³, si può giungere ad una corretta misura e gestione delle preferenze, grazie alla combinazione degli approcci quantitativi di natura psicologica (le cui indagini sono finalizzate a catturare quante più caratteristiche comportamentali possibili) ed economica (i cui scopi sono lo studio delle preferenze dei consumatori e dei loro profili di rischio).

Un'ulteriore applicazione concreta, concernente in maniera più diretta il campo dell'asset management, riguarda la relazione tra *rischio* e *rendimento*, non più costante nel tempo ma variabile e funzione delle preferenze dei partecipanti al mercato, dell'ambiente di riferimento e degli aspetti istituzionali che regolano tale ambiente. A corollario di ciò, lo stesso *equity risk premium* non è un parametro universale ma variabile in base ai fattori appena esposti. Questa condizione rappresenta sicuramente una rottura con la visione economica classica della relazione rischio-rendimento (ad alto rischio corrisponde un alto rendimento, e viceversa) poiché situazioni e comportamenti degli agenti di mercato possono portare a plurime violazioni. In *figura 2* vengono

³³ Si veda Lo (2005), pagg. 27-29 per i riferimenti completi di tali studi. In questa sede vengono riportati i principali spunti e congetture sulla base delle quali si sono sviluppate le AMH.

riportati rendimenti e deviazioni standard del CRSP *Value-Weighted Index Returns* e si evidenzia come, in alcuni periodi storici, la relazione tra rischio e rendimento sia negativa (la correlazione tra le due serie è di -59,9%).

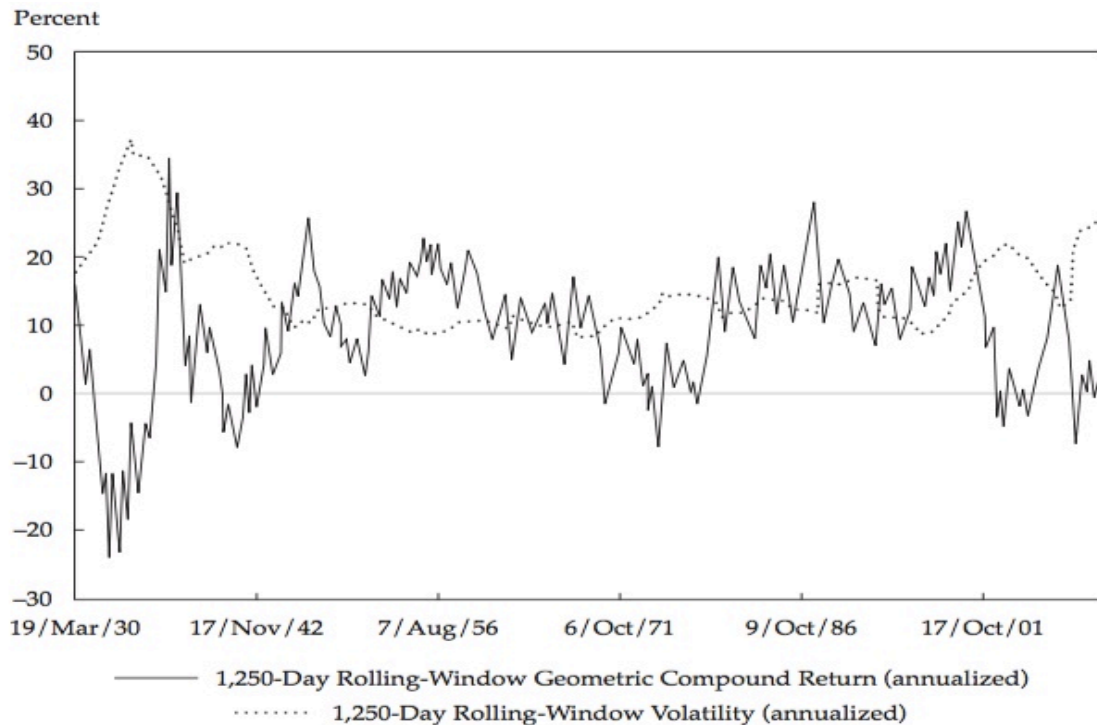


Figura 2: rendimenti e deviazioni standard annualizzati dell'indice CRSP *Value-Weighted Index Returns*, dati dal 19 marzo 1930 al 31 dicembre 2010 ed aggregati secondo rolling windows di 1250 giorni (Fonte: Lo, 2012).

Sotto le AMH, la spiegazione di tale fenomeno riguarda il comportamento che una consistente porzione degli investitori assume in periodo di galoppante volatilità azionaria: tali individui migrano le proprie risorse verso investimenti più sicuri o disinvestono i capitali dal mercato dell'equity, secondo un fenomeno finanziario definito *flight to safety*. Questo processo di disinvestimento produce un *downgrade* dei prezzi azionari e una pressione al rialzo nei prezzi degli asset meno rischiosi, causando così una rottura della relazione rischio-rendimento. Viene dunque dimostrato che è l'interazione tra gli agenti in gioco nell'ambiente di riferimento che determina la relazione tra i due fattori, risultato che non preclude alla relazione rischio-rendimento di seguire gli usuali binari della teoria classica in certi periodi ma che evidenzia come tale relazione sia mutevole ed adattiva.

Una seconda implicazione relativa all'asset management riguarda le opportunità di arbitraggio, assenti nel *framework* delle EMH ed invece qui, data la complessa dinamica

descritta dalle AMH, esistenti e sfruttabili. Come esposto precedentemente (v. pag. 28), già Grossman e Stiglitz avevano presentato aspre critiche alle EMH sull'impossibilità, da queste teorizzata, di ottenere extra profitti sfruttando tali opportunità, condizione invece esistente in un mercato attivo e liquido come quello dei capitali. Secondo la nuova prospettiva evuzionistica, un'attenta ed accorta *asset allocation*³⁴ che sfrutti tali opportunità può portare ad un *surplus* di valore per quegli investitori in grado di adattarsi al contesto economico. Tali opportunità sono mutevoli, si presentano e si esauriscono una volta sfruttate, seguendo i vari cicli che caratterizzano i mercati: periodi di forte trend, bolle azionarie, crolli finanziari, nascita e declino di nuove "specie", cambiamenti istituzionali e regolamentari. La realtà complessa dei mercati *impone* lo sviluppo di strategie adattive che sfruttino le opportunità di profitto, e solo le strategie più efficaci permetteranno la sopravvivenza nel sistema.

Una terza implicazione è l'efficacia relativa delle strategie di investimento, adatte e profittevoli in certi contesti ma dannose e sotto-performanti in determinati altri. Questo è un ulteriore punto di rottura rispetto al paradigma fondamentale di investimento, poiché ancora una volta viene ribadito come l'efficacia della strategia dipende incontrovertibilmente dalle condizioni dell'economia/ecologia nella quale è applicata, e se in certi periodi può condurre a considerevoli perdite, in certi altri può tornare a produrre profitti considerevoli. In Lo (2004, 2005) viene riportato uno studio a supporto di questa affermazione, presentando i coefficienti di autocorrelazione di primo ordine

³⁴ L'asset allocation è quel processo decisionale mediante il quale si distribuiscono le risorse a disposizione fra le varie possibilità di investimento (strumenti finanziari o attività reali). Vi sono tre possibili sottocategorie di asset allocation, ovvero:

- I. Asset allocation *strategica*, orientata su un orizzonte di medio-lungo periodo e finalizzata ad individuare quelle asset classes le cui performance presentano maggior stabilità in tale arco temporale;
- II. Asset allocation *tattica*, orientata su un orizzonte di breve periodo e contingente rispetto ad una visione strategica, adattando le scelte di portafoglio strategiche ai trend ed alle forze in atto nel mercato;
- III. Asset allocation *dinamica*, anch'essa orientata nel breve periodo ma finalizzata ad una completa adattabilità delle scelte di portafoglio in base alle condizioni di mercato, e sarà tanto più efficace quanto più sarà adeguato il *timing* di intervento del gestore di portafoglio.

(Fonte: <http://www.borsaitaliana.it/notizie/sotto-la-lente/assetallocation.htm>)

dei rendimenti mensili dello S&P *Composite Index*, considerando delle rolling windows di cinque anni e per il periodo compreso tra gennaio 1871 e aprile 2003 (figura 3 seguente).

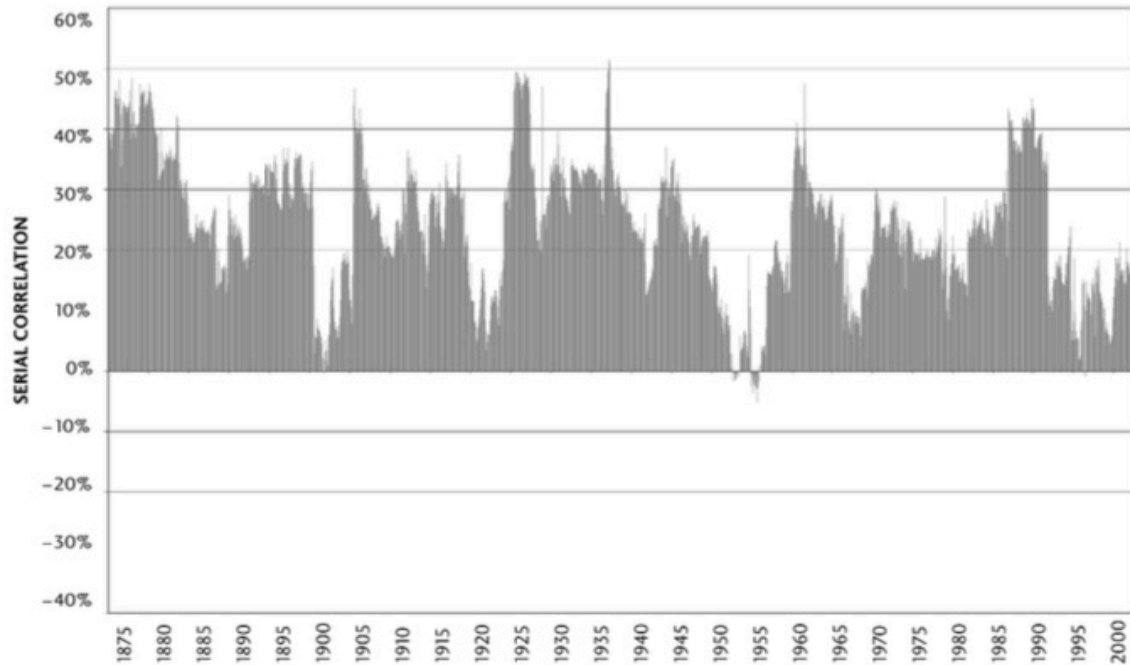


Figura 3: coefficiente di correlazione seriale dei rendimenti del S&P *Composite Index*, aggregati con rolling window di cinque anni, dati dal gennaio 1871 ad aprile 2003. (Fonte: Lo, 2012)

Seguendo i costrutti delle EMH (i prezzi seguono un modello *random walk* e sono serialmente indipendenti, $\hat{\rho}_1$ prossimo o uguale a zero a seconda del grado di efficienza del mercato), ci si aspetterebbe di osservare valori elevati di correlazione nei periodi iniziali, progressivamente tendenti a zero all'aumentare dell'efficienza del mercato statunitense. Al contrario, nella pratica, si apprezza un andamento ciclico della correlazione seriale, e l'efficienza secondo le EMH si raggiunge solo in alcuni periodi (primi anni del 1900, l'immediato dopoguerra degli anni '50, i primi anni '90 dove si registra una correlazione quanto meno prossima allo zero). Tale evidenza empirica conferma in pieno il concetto di ciclicità del mercato abbracciato dalle AMH ed evidenzia la necessità di ristrutturare ed adattare continuamente le proprie strategie di *asset allocation* al contesto di riferimento.

Un'ultima diretta implicazione delle AMH per la gestione dell'*asset allocation* è che i *fattori di rischio* non sono costanti nel tempo ma sono mutevoli, ed in determinate

situazioni caratteristiche quali crescita e valore possono essere considerati indicatori di rischio: se per le EMH esistono fattori di rischio e fattori *risk free*, il *framework* introdotto dalle AMH non pone restrizioni in tal senso. A stabilire quali fattori, in determinati contesti, siano da considerare rischiosi e dunque presentino un *risk premium* maggiore, è la natura della popolazione degli investitori e l'interazione tra le loro preferenze e le risorse presenti nei mercati. Ad esempio, durante la bolla tecnologica statunitense di fine anni '90, le preferenze degli investitori erano concentrate sulle azioni in forte crescita delle imprese operanti in campo informatico, salvo poi invertirsi una volta scoppiata la bolla speculativa. Se dunque una fetta consistente degli investitori esprime preferenze per i titoli in crescita, il prezzo di quei titoli incorporerà un *risk premium* sempre maggiore, salvo poi decrescere sia per l'entrata nel mercato di nuovi investitori con diverse preferenze sia per l'uscita del mercato degli investitori già presenti. Sebbene il lavoro di Lo non fornisca un metodo quantitativo per determinare quali e quante caratteristiche siano prezzate e considerate o meno *risk factors* e *risk premium* – ed è un fatto conseguente alla natura delle AMH – tuttavia lo studioso espone alcune linee guida qualitative (Lo, 2005): si rende necessario, secondo Lo, costruire delle misure di sintesi che colgano trasversalmente le preferenze degli individui e le condizioni del mercato esistenti ed ipotetiche, sviluppando così una descrizione completa dell'ambiente economico e dei fattori di rischio determinanti, sfruttati dagli individui che in tale ambiente operano.

2.6.4 Sopravvivenza e nuove tecnologie: sfide rinnovate per il *portfolio management*

Le AMH e le implicazioni che esse comportano aprono le porte a nuove sfide per il mondo dell'*investment management*: se da un lato abbracciare tali ipotesi significa dare freschezza ed interdisciplinarietà ad un contesto che non può prescindere dal mettere al centro delle valutazioni economiche le scelte, le preferenze ed i comportamenti degli individui, dall'altro la carenza di un *background* quantitativo sufficientemente ampio rende tale teoria ancora instabile. Le sfide che le AMH pongono, ad oggi, sono però molto meno scoraggianti di quanto si possa pensare: innovative tecnologie finanziarie e di *trading*, sviluppi nelle valutazioni statistiche riguardo a parametri più puntuali

nell'indicare il *market timing*, mercati globalizzati, liquidi ed interconnessi nei quali è più immediato reperire e trasmettere informazioni, sono tutti fattori che consentono la creazione di basi quantitative più adatte a descrivere un contesto finanziario che è innegabilmente sempre più complesso e non è più sufficiente qualche ipotesi approssimata e semplificativa per modellizzarlo. La chiave di tale ragionamento, che inequivocabilmente rappresenta anche l'ultima implicazione diretta del *framework* delle AMH, è «innovation is the key to survival» (Lo, 2005). Innovazione che non riguarda solamente tecnologie e strumenti finanziari, ma che indubbiamente concerne anche la preparazione e competenza necessarie che devono possedere i *portfolio managers*: un costante aggiornamento delle proprie conoscenze permette di primeggiare nella competizione e di sopravvivere anche durante repentini cambiamenti di mercato. Innovazione ed adattabilità devono essere i primi motori del comportamento nel mondo finanziario, a cui devono necessariamente aggiungersi flessibilità ed una mente costantemente aperta: tutti questi fattori possono determinare uno spartiacque tra sopravvivenza ed estinzione nei mercati finanziari.

In questo capitolo è stato fornito il *background* teorico di riferimento, le AMH, al quale si cercherà di dare una possibile forma quantitativa mediante la costruzione di un algoritmo adattivo a rinforzo con lo scopo di verificare se, effettivamente, seguendo il *framework* di Lo è possibile raggiungere un adeguato livello di profitto operando sul mercato. Il capitolo seguente presenta una serie di metodi computazionali, definiti *reinforcement learning*, utilizzati per risolvere problemi dinamici di ottimizzazione attraverso il criterio del *trial-and-error* e, per costruzione, efficaci nella ricerca di una selezione adattiva. Se dunque secondo Andrew Lo la nuova frontiera della teoria dei mercati si debba estendere ad abbracciare concetti quali apprendimento, adattamento, evoluzione, la filosofia alla base dei metodi *reinforcement learning* sembra combaciare perfettamente con quanto asserito dal professore statunitense: tali metodi infatti configurano il comportamento di un *agente* che apprende determinati comportamenti interagendo con l'ambiente circostante, e si candidano dunque ad essere una delle vie computazionali per tradurre *nella pratica* ciò che egli ha teorizzato.

3. REINFORCEMENT LEARNING

L'idea che l'individuo interagisca con l'ambiente circostante ed apprenda da esso è sicuramente il primo pensiero che sovrviene parlando della natura dell'apprendimento (Barto e Sutton, 2016). L'apprendimento per *rinforzo*, dall'inglese *reinforcement learning*, a pensarci bene, è il principale approccio comportamentale che l'individuo/principiante adotta nell'ambiente in cui vive: l'infante che impara a non avvicinarsi ai fornelli una volta ricevuta una bruciatura, l'adolescente che apprende come guidare un'automobile effettuando numerose prove, il giocatore di scacchi professionista che pianifica la propria mossa analizzando le possibili contromisure dell'avversario, sono tutti esempi di comportamenti adattivi e dipendenti dalle condizioni ambientali in cui l'*agente* (sia esso il neonato, l'adolescente o il giocatore professionista) è immerso. L'apprendimento per rinforzo (d'ora in avanti, RL) è il termine utilizzato per indicare sia il problema derivante dagli esempi appena esposti, sia il *corpus* di metodi computazionali adatti a massimizzare una certa funzione di ricompensa sia la disciplina che studia tali problemi e le possibili metodologie di risoluzione. Vi sono tre principali caratteristiche distintive che è necessario tenere a mente riguardo i problemi legati al RL:

- I. Essenzialmente, sono problemi *a ciclo chiuso* poiché le azioni intraprese dall'*agente* in un dato momento influenzeranno gli input ambientali futuri.
- II. Il raggiungimento della miglior ricompensa non avviene tramite *passi* che sono già conosciuti, ma viene guadagnato dall'*agente* procedendo per tentativi, scoprendo in itinere quali azioni conducono ai migliori risultati.
- III. In alcuni casi, le conseguenze delle azioni intraprese nel presente hanno effetti sostanziali anche nel lungo periodo.

Il presente capitolo si propone di argomentare le tematiche del RL. Preliminarmente è tuttavia necessario focalizzare attentamente l'aspetto chiave del problema che tali metodi computazionali cercano di risolvere, ovvero quello di un *agente* che interagisce con un *ambiente* dalle condizioni *incerte* al fine di raggiungere uno o più *risultati*, dipendenti anche dalle condizioni dell'ambiente nel quale egli agisce. L'agente in questione deve dunque essere in grado di avvertire i mutamenti ambientali, così come deve essere in grado di intraprendere azioni che influenzino siffatto ambiente. Il tritico *sensazione* (intesa come capacità di avvertire i mutamenti), *azione*, *obiettivo* è ciò che deve caratterizzare lo schema comportamentale di un *agente*, assieme ad un giusto bilanciamento tra *esplorazione* e *sfruttamento*³⁵: al fine di ottenere un profitto/ricompensa, l'agente deve perseguire quelle azioni che nel passato si sono rivelate efficaci (sfruttamento) ma, per scovare tali azioni, deve necessariamente intraprendere decisioni che ancora non ha selezionato (esplorazione). È chiara la necessità di un accorto *trade-off* tra i due comportamenti, com'è lampante che non vi è possibilità di perseguire gli obiettivi prefissati, intraprendendo tali azioni, senza incappare in errori: l'agente persegue quelle azioni che, in base all'interpretazione degli stimoli ricevuti, appaiono essere le migliori e se si riveleranno tali perseguirà nel proprio comportamento, se invece daranno esiti non desiderabili ne conserverà memoria ed intraprenderà scelte diverse. Dalle argomentazioni appena esposte, la connessione tra i metodi computazionali del RL e le AMH sviluppate da Andrew Lo si intreccia e connette ancor più intensamente: il *framework* presentato dall'accademico statunitense presenta proprio quelle peculiarità che i metodi RL mirano a descrivere e risolvere.

Questo capitolo presenta le principali caratteristiche dei metodi RL, precedute da alcuni interessanti ed utili spunti storici. In questa tesi, verrà considerato il metodo Q-Learning, un approccio RL appartenente alla branca dei sistemi basati sull'apprendimento per differenze temporali, di cui si avrà modo di apprenderne le specificità. Si analizzeranno dunque, in sequenza, aspetti generali dei metodi RL, aspetti più distintivi del

³⁵ I termini utilizzati da Barto e Sutton (2016) sono *exploration* ed *exploitation*; in questo caso, si è cercata la traduzione più fedele al contesto perdendo chiaramente, da un punto di vista sintattico, la portata linguistica derivante dall'accoppiamento dei due vocaboli.

sottoinsieme dei metodi per differenze temporali ed infine gli elementi fondamentali per la costruzione dell'algoritmo basato sul Q-Learning. D'ora in avanti, con il termine *agente*, si intenderà dunque non solo un individuo pensante ma soprattutto un agente artificiale strutturato secondo un algoritmo.

3.1 La storia del *reinforcement learning*: cenni

La storia del RL si può sostanzialmente raccogliere in tre voluminosi filoni, inizialmente sviluppati in maniera indipendente ed in periodi storici diversi, per poi convergere assieme attorno agli anni '80, formando così l'approccio moderno alle teorie dell'apprendimento per rinforzo. Il primo filone concerne il metodo *trial-and-error*, i cui iniziali sviluppi risalgono all'ambito della psicologia dell'apprendimento animale: la discussione degli anni '50 dell'800 dello studioso Alexander Bain riguardo l'apprendimento per "groping and experiment" (letteralmente *andare a tentoni e sperimentare*), e la descrizione che fece lo psicologo britannico Morgan nel 1894 per delineare il comportamento animale, sono due esempi. Tuttavia, la prima effettiva argomentazione del *trial-and-error* come principio base dell'apprendimento viene da Thorndike e dalla sua "Law of Effect" del 1911: lo studioso utilizza tale termine per descrivere l'effetto di rinforzo che producono gli eventi nella selezione dei comportamenti degli individui. Il termine *rinforzo*, applicato al contesto dell'apprendimento, prende definitivamente piede a seguito della teoria di Thorndike: tale concetto descrive il rafforzamento di uno schema comportamentale come risultato della ricezione di uno stimolo da parte dell'individuo, diretta conseguenza del cambiamento delle condizioni ambientali. Il *rinforzo* produce dunque cambiamenti nel comportamento, e l'individuo interpreta gli stimoli agendo di conseguenza, conservando memoria dei comportamenti passati e ricercando la risposta migliore allo stimolo che l'ambiente trasmette in quel dato istante temporale. Non tutti gli stimoli condurranno a reazioni efficaci ed efficienti, ma è proprio attraverso gli errori comportamentali che l'individuo migliorerà le proprie risposte future e la propria capacità adattiva. Tali idee e teorie trovarono applicazione, a partire dalla fine della Seconda Guerra Mondiale, anche in campo informatico e iniziarono ricerche e studi per

implementare il metodo di apprendimento *trial-and-error* in intelligenze artificiali. In ambito economico, invece, si deve *l'imprinting* delle ricerche a due studiosi, Bush e Mosteller, i quali nel 1973 applicarono la teoria dell'apprendimento ad una serie di modelli economici classici³⁶.

Il secondo filone concerne il problema del controllo *ottimo* attraverso l'utilizzo di funzioni di valore³⁷ e di una programmazione *dinamica*. L'utilizzo del termine "controllo ottimo" risale alla metà degli anni '50, con il quale si racchiudeva l'emergente problema dell'implementazione di un *controller* che minimizzasse la misura di un qualche comportamento di un sistema dinamico nel tempo³⁸. Uno degli approcci più famosi a tale problema venne da Richard Bellman, il quale sviluppò una funzione di valore, nota come *optimal return function* (oggi prende anche il nome di *equazione di Bellman*), per risolvere il problema di ottimizzazione in un ambiente dinamico: tale approccio viene definito *programmazione dinamica*, sistema che è anche considerato il metodo più efficace per risolvere problemi stocastici di controllo ottimale. Il terzo ed ultimo filone riguarda, infine, l'apprendimento per differenze temporali (dall'inglese *temporal-difference learning*), basato sulla differenza tra due stime temporalmente consecutive della stessa quantità (ad esempio, la probabilità di vittoria in una partita di scacchi a seguito di due mosse consecutive). Le origini di tale metodologia sono parzialmente derivanti dalla psicologia dell'apprendimento animale, in particolare attorno al concetto di *rinforzo secondario*: tale rinforzo è uno stimolo che viene accorpato ad un rinforzo primario, dal quale ne riceve le medesime proprietà. Minks, nel 1954, e Samuel, nel 1959, furono i primi pionieri dell'apprendimento per differenze temporali, implementando tali principi psicologici ai sistemi di apprendimento artificiali. Uno dei contributi più importanti al RL giunse dal Klopf nel 1972: lo studioso unì i principi del *trial-and-error* ad una rilevante componente dell'apprendimento per differenze

³⁶ Per approfondimenti si veda Barto e Sutton, 2016, cap. I.

³⁷ Una funzione di valore è una funzione che specifica il potenziale guadagno di lungo periodo, ad esempio il rendimento totale atteso che un investitore si aspetta di guadagnare considerando un orizzonte temporale lungo. Nel paragrafo seguente verrà trattato il tema in maniera esaustiva.

³⁸ Un sistema dinamico, com'è facilmente intuibile, è un sistema in costante cambiamento ed evoluzione, il cui stato presente può variare repentinamente e tali cambiamenti sono numerosi nel corso del tempo.

temporali da lui sviluppata, il *rinforzo generalizzato*, tale per cui ogni input e stimolo viene considerato in termini di rinforzo.

Questi tre filoni, inizialmente sviluppati singolarmente, a partire dagli anni '80 cominciarono definitivamente a confluire assieme costruendo così lo scheletro del RL moderno. Inoltre, i filoni dell'apprendimento per differenze temporali e del controllo ottimo vennero risolutivamente accorpati nel 1989 con lo sviluppo del Q-Learning da parte di Chris Watkins. Tale elaborato si focalizzerà su questo metodo di apprendimento per creare un algoritmo di trading finanziario. Verranno dunque di seguito esposte le principali caratteristiche di tale metodologia, partendo inizialmente dai concetti generali del RL, entrando poi nello specifico nel corso del capitolo.

3.2 Elementi base del *reinforcement learning*

Alla base di tutti metodi RL vi è, chiaramente, la relazione tra *agente* ed *ambiente* e le sue peculiarità. Tale interazione può evolversi lungo step temporali *discreti* ($t=0,1,2,3, \dots$) o *continui*, ed è rappresentata, in caso di un tempo discreto, nella *figura 4* riportata di seguito.

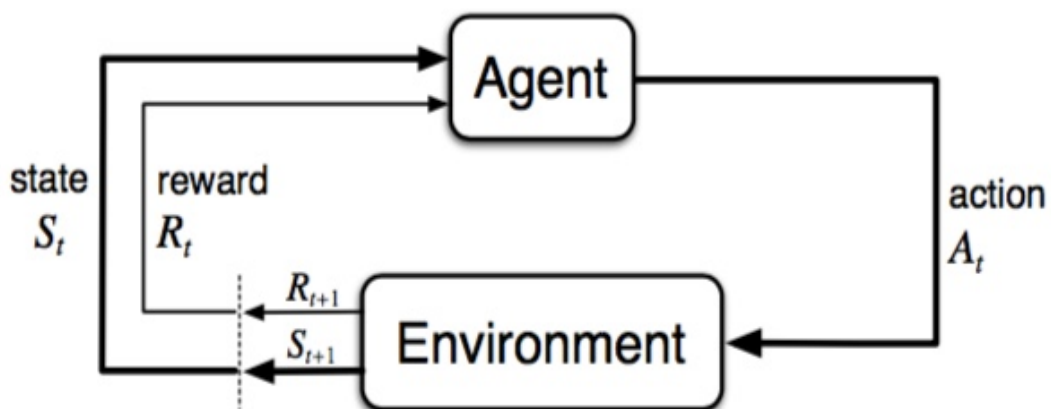


Figura 4: interazione tra agente ed ambiente nel framework RL (Fonte: Barto e Sutton, 2016)

Ad ogni istante di tempo t , l'agente riceve un input dall'ambiente, $s_t \in S$, dove S rappresenta l'insieme di tutti i possibili stati del mondo, e sulla base di questo intraprende un'azione $a_t \in A(S_t)$, dove $A(S_t)$ rappresenta l'insieme di tutte le possibili azioni dato lo stato del mondo s_t . Conseguentemente all'azione intrapresa, l'agente

riceverà una *ricompensa numerica* $r_{t+1} \in R(s_t) \subset \mathbb{R}$, dove $R(s_t)$ rappresenta l'insieme delle possibili ricompense condizionatamente ad s_t , e si troverà in un nuovo stato del mondo s_{t+1} , condizionato in parte anche dalla scelta dell'azione a_t . Il concetto di *ricompensa* permette di introdurre alcuni elementi fondamentali del *framework* RL; oltre all'interazione ambiente-agente, è possibile infatti individuare quattro sottocategorie di un sistema RL: un *criterio comportamentale*³⁹, un *sistema di ricompensa*, una *funzione di valore* e, in alcuni casi, un *modello* dell'ambiente circostante (Barto e Sutton, 2016).

Il *criterio comportamentale* caratterizza in ogni istante l'agente e rappresenta la mappatura dell'insieme di regole di risposta agli stimoli che questo sviluppa nel tempo. In alcuni casi può essere una semplice funzione, in altri casi invece è rappresentata da un sistema più complesso. Tale criterio è il fulcro di un agente RL poiché da solo è sufficiente a definire lo schema comportamentale che questo assume, ed è generalmente un criterio stocastico: ad ogni istante di tempo t , l'agente implementa una mappatura di tutti i possibili stati del mondo indicata con π ; nel caso di un ambiente stocastico la notazione esatta è $\pi(a|s)$, ovvero la probabilità che $a_t = a$ dato $s_t = s$, per ogni $s \in S$ ed $a \in A(s)$. Il metodo RL utilizzato specifica come l'agente modifichi la propria *policy* condizionatamente all'output ricevuto dall'ambiente e come dunque l'esperienza passata influisca sulla scelta delle azioni future.

Il *sistema di ricompensa* definisce l'obiettivo in un problema RL. Ad ogni istante di tempo considerato, l'ambiente rende all'agente un valore rappresentante la ricompensa per quel dato istante, r_t : ad esempio, un portafoglio di titoli fornisce giornalmente il rendimento raggiunto, e l'obiettivo dell'agente è quello di massimizzare tale ricompensa nel tempo⁴⁰. Il processo che genera la ricompensa non è in alcun modo alterabile né

³⁹ Il termine inglese utilizzato è *policy*. Si è qui cercato di trovare rendere efficace la traduzione del termine, tuttavia in alcuni passaggi seguenti verrà utilizzato il termine originale, per una maggiore coerenza linguistica.

⁴⁰ È bene precisare che non è sempre raggiunta la massimizzazione della ricompensa da parte di un agente, e che dunque *ottimizzazione* non è sinonimo di *ottimalità*, specialmente riferendosi ad un *framework* di tal genere. Il processo di ottimizzazione fa costantemente i conti con il processo ricorsivo dei metodi RL di "learning while interacting with the environment" (Barto e Sutton, 2016), conducendo a continue e

modificabile dall'agente: egli può alterare il *valore* della ricompensa modificando la sua policy di comportamento al fine migliorare r_{t+1} , ma non può in alcun modo modificare la funzione generatrice del sistema. Se, tornando all'esempio precedente, la selezione di portafoglio produce un basso o insoddisfacente livello di rendimento, l'agente agirà modificando i propri parametri di selezione per ottenere un miglior profitto futuro. Di fatto, la ricompensa è un immediato indicatore di *performance* che determina se il comportamento è efficace nel breve periodo, mentre una valutazione di lungo periodo la fornisce la funzione di valore.

La *funzione di valore* specifica, infatti, il valore di uno stato nel tempo ed equivale, in altre parole, al valore attuale dell'ammontare totale di ricompensa che l'agente si aspetta di ricevere nel futuro. Mentre dunque la ricompensa fornisce una valutazione istantanea ed immediata, la funzione di valore restituisce come output l'*appetibilità* dell'ambiente nel tempo. Chiaramente, non può esservi generazione di valore senza ricompense, e lo scopo principale che conduce alla stima del valore è la ricerca di migliori risultati; d'altro canto, le azioni dell'agente sono finalizzate ad ottenere maggior valore nel lungo periodo, non necessariamente maggiori ricompense nel breve periodo. Le ricompense derivano strettamente e direttamente dall'ambiente di riferimento, il valore che invece esso genera è una variabile che dev'essere costantemente monitorata e stimata e la funzione di valore di fatto è la chiave di qualsiasi sistema RL.

Il *modello ambientale*, infine, tenta di catturare lo schema comportamentale dell'ambiente di riferimento: ad esempio, considerato l'ambiente versante nello stato t e data un'azione dell'agente, tale modello potrebbe predire lo stato del mondo in t_1 e la ricompensa conseguente. Non tutti i metodi RL prevedono la modellizzazione dello schema comportamentale di un'ambiente al fine di *pianificare* accuratamente gli step di azioni future, quelli che la adoperano vengono definiti metodi *model-based* e si contrappongono ai metodi *model-free* che invece non ne prevedono l'utilizzo. Il metodo di RL considerato in questa tesi non prevede l'utilizzo di un modello ambientale, tema

inequivocabili modifiche al processo di ottimizzazione stesso verso soluzioni non sempre ottime ma *near-optimal*, come si approfondirà in seguito.

che dunque verrà tralasciato in questa sede rimandandone l'approfondimento ad altri testi⁴¹.

3.2.1 La funzione di ricompensa

I metodi RL, come si è avuto modo di apprendere dal paragrafo precedente, trattano lo scopo di raggiungimento di un *obiettivo* da parte dell'agente in termini di un segnale di ricompensa che viene trasmesso dall'ambiente all'agente. Formalmente, può essere definita una particolare *ipotesi di ricompensa* nel seguente modo (Barto & Sutton, 2016): lo scopo e l'obiettivo principale dell'agente è la massimizzazione del valore atteso della somma totale dei segnali scalari ricevuti, definiti ricompense, nel lungo periodo. La funzione di ricompensa può essere dunque espressa in termini di valore attuale dei rendimenti attesi futuri nel seguente modo:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^T \gamma^k r_{t+k+1}, \quad (15)$$

dove r_{t+k} è il rendimento al periodo $t + k$, γ è il tasso di sconto, compreso tra 0 e 1, estremi inclusi ($0 \leq \gamma \leq 1$), e T è lo step temporale finale. Quest'ultimo può essere considerato *finito* oppure *infinito*: nel primo approccio, si avranno interazioni agente-ambiente che verseranno in delle sotto sequenze, definite *episodi*, il cui *stato terminale* dell'episodio (nel tempo $t = T$) sarà seguito da un reset dell'interazione ed un ritorno ad uno stato iniziale; il secondo approccio, rappresentate un'interazione agente-ambiente infinita, è chiaramente più complesso poiché lo *stato terminale* sarà $T = \infty$ e non vi saranno sotto episodi ma l'interazione proseguirà senza interruzioni *a flusso continuo*. Il tasso di sconto γ determina il valore attuale dei rendimenti futuri: se $\gamma = 0$, l'agente è focalizzato solamente a massimizzare il rendimento di *breve periodo* r_{t+1} e tale comportamento può precludere l'accesso a rendimenti futuri più elevati. Più il valore di γ si avvicina all'unità, più l'agente terrà in maggior considerazione i rendimenti futuri, assumendo un comportamento più lungimirante.

⁴¹ In particolare, si veda Barto e Sutton, 2016.

3.2.2 Il processo decisionale à la Markov

Prima di addentrarsi ad esplicitare la funzione di valore, è necessario innanzitutto soffermarsi sul processo decisionale che l'*agente* intraprende e che è funzione di un segnale captato in base allo stato nel quale versa l'ambiente al dato istante indicato. Con *stato* si deve dunque intendere l'insieme di informazioni disponibili per l'agente, non alterabile in modo alcuno dallo stesso ma influenzabile dalle sue azioni, come descritto nella precedente *figura 4*. Sarebbe utopistico supporre che l'impulso derivante dall'ambiente fornisca all'agente la totalità delle informazioni, poiché vi sono *stati* ambientali celati, sicuramente utili al processo decisionale ma non ancora conosciuti dall'agente poiché questo non ha ancora ricevuto gli impulsi necessari per scoprirli, oppure non ha ancora intrapreso le azioni necessarie a condurlo in quei suddetti stati. Dunque, uno *stato* che riassume in maniera compatta tutte le informazioni passate evidenziando e conservando quelle più rilevanti sarebbe idealmente quello preferibile, poiché permetterebbe all'agente di ricevere il *segnale* necessario per intraprendere l'azione più appropriata al contesto di riferimento. Un *segnale* avente le caratteristiche appena descritte si definisce un *segnale à la Markov*, o equivalentemente un segnale possedente la proprietà Markoviana. Formalmente tale proprietà, applicata ad un problema RL, si sostanzia nel seguente modo⁴²: si consideri, innanzitutto, la possibile risposta di un ambiente al tempo $t+1$ a seguito dell'azione intrapresa dall'agente al tempo t . Questa, nel caso più generale, dovrebbe dipendere da tutto l'accaduto pregresso ed essere espressa in termini di probabilità condizionata nel seguente modo:

$$Pr\{s_{t+1} = s', r_{t+1} = r \mid s_0, a_0, r_1, \dots, s_{t-1}, a_{t-1}, r_t, s_t, a_t\}, \quad (16)$$

per ogni s', r e dati tutti i possibili valori degli stati passati a, s, r .

⁴² Per semplificazione, si considerano stati del mondo e valori di ricompensa numericamente finiti. Questo consente di lavorare in termini di probabilità e non in termini, più complessi, di funzioni di densità di probabilità e di integrali, nonostante poi il teorema esposto si possa facilmente estendere, da un contesto discreto e finito ($t=0,1,2, \dots, T$) ad un contesto continuo.

Se il segnale di risposta dell'ambiente possiede la proprietà Markoviana, allora la proprietà (16) può essere riscritta nel seguente modo:

$$p(s', r | s, a) = \Pr\{s_{t+1} = s', r_{t+1} = r | s_t = s, a_t = a\}, \quad (17)$$

per ogni s', s, r, a . Risulta evidente, dunque, la dinamica ad un passo insita nella proprietà Markoviana e la possibilità di calcolare, con relativa accuratezza, la probabilità di versare in uno stato futuro $s_{t+1} = s'$ ed ottenere una ricompensa $r_{t+1} = r$ essendo in possesso delle sole informazioni che l'ambiente fornisce al tempo t . Tale proprietà è dunque di fondamentale importanza poiché le decisioni e le ricompense future sono assunte derivanti solo dallo stato corrente t , il quale deve essere necessariamente informativo per l'agente.

Un approccio RL che soddisfa la proprietà Markoviana viene definito un *processo decisionale à la Markov* (MDP, acronimo di *Markov decision process*). Data la dinamica descritta dalla (17), si possono infine esplicitare facilmente le seguenti proprietà, ovvero:

- *sommatoria dei rendimenti attesi, data la coppia stato-azione:*

$$r(s, a) = E(r_{t+1} | s_t = s, a_t = a) = \sum_{r \in R} r \sum_{s' \in S} p(s', r | s, a) \quad (18)$$

- *probabilità di transizione:*

$$p(s' | s, a) = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\} = \sum_{r \in R} p(s', r | s, a). \quad (19)$$

3.2.3 La funzione di valore: computazione ed ottimizzazione

Tutti i metodi RL prevedono, nel loro algoritmo, l'implementazione di una funzione di valore che fornisca come output il livello di *bontà* di un particolare stato s_t per l'agente o, in egual misura, quanto risulta performante intraprendere una determinata azione a_t

in un certo stato s_t . Il valore, proporzionale alle ricompense future ottenibili dallo stato s_t , dipende necessariamente dal criterio comportamentale π adottato (definito precedentemente, si veda pag. 50); formalmente, dunque, il valore di uno stato $s_t = s$ sotto un criterio comportamentale $\pi(\cdot)$, indicato con $V^\pi(s)$ e definito *funzione del valore di stato s per il criterio π* , sarà:

$$V^\pi(s) = E_\pi(R_t | s_t = s) = E_\pi(\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s), \quad (20)$$

dove E_π indica il valore atteso condizionatamente alla selezione del criterio comportamentale π mentre T è lo *stato terminale* nel quale $V^\pi(s) = 0$, poiché chiaramente non vi è alcuno stato successivo possibile e di conseguenza non vi sarà alcuna ricompensa r_{T+1} .

In maniera analoga, è possibile definire la *funzione del valore dell'azione a per il criterio π* nel modo seguente:

$$\begin{aligned} Q^\pi(s, a) &= E_\pi(R_t | s_t = s, a_t = a) \\ &= E_\pi(\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s, a_t = a). \end{aligned} \quad (21)$$

L'equazione (21) rappresenta il valore determinato dalla scelta di perseguire l'azione $a_t = a$, versando in uno stato $s_t = s$ e seguendo un criterio comportamentale π , ovvero la sommatoria di tutti i rendimenti attesi futuri data quella particolare azione. Le funzioni V^π e Q^π possono essere stimate dall'agente in base all'esperienza nell'interazione con l'ambiente e, utilizzando l'una o l'altra funzione, questa dovrebbe essere mantenuta come *parametrizzata* per agire direttamente sulle variabili della stessa, *fittandole* ai rendimenti ottenuti, in modo da adattarsi meglio allo stato s_t e soprattutto agli stati futuri.

Una delle proprietà fondamentali delle funzioni di valore appena trattate è quella di seguire una particolare relazione ricorsiva, valida per ogni criterio π ed ogni stato s , ricorrente tra lo stato attuale $s_t = s$ ed ogni suo possibile stato futuro:

$$\begin{aligned}
V^\pi(s) &= E_\pi(R_t | s_t = s) \\
&= E_\pi(\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s) \\
&= E_\pi(r_{t+1} + \gamma \sum_{k=0}^T \gamma^k r_{t+k+2} | s_t = s) \\
&= E_\pi(r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s). \tag{22}
\end{aligned}$$

Tale uguaglianza si definisce *equazione di Bellman* per $V^\pi(s)$ e descrive la relazione tra s_t ed s_{t+1} . Analogamente, seguendo gli stessi passaggi dell'equazione (22), si giunge a definire l'*equazione di Bellman* per $Q^\pi(s, a)$, per ogni stato s , ogni azione a ed ogni criterio π come segue:

$$Q^\pi(s, a) = E_\pi(r_{t+1} + \gamma Q^\pi(s_{t+1}, a_{t+1}) | s_t = s, a_t = a). \tag{23}$$

I valori V^π e Q^π rappresentano le uniche soluzioni possibili rispettivamente per le equazioni (22) e (23). Tali equazioni, in sintesi, esprimono la proprietà che il valore dello stato di partenza s_t deve equivalere alla somma tra il valore atteso scontato dello stato successivo s_{t+1} e il rendimento ottenuto nel passaggio dallo stato s_t allo stato s_{t+1} .

Ricordando che la risoluzione di un problema RL consiste nell'individuare quel criterio π che permette di performare in maniera ottimale, si hanno dunque tutti gli elementi per poter introdurre l'ottimizzazione delle funzioni di valore. Innanzitutto, un criterio π' si definisce migliore di un criterio π qualora il suo rendimento atteso sia maggiore o uguale a quello ottenuto mediante il criterio π , per ogni stato possibile. Formalmente, $\pi' \geq \pi$ se e solo se $V^{\pi'}(s) \geq V^\pi(s)$, per ogni $s \in S$, o equivalentemente $Q^{\pi'}(s, a) \geq Q^\pi(s, a)$, per ogni $s \in S$ ed $a \in A(s)$. Quel criterio π^* che avrà un valore maggiore od uguale a tutti gli altri sarà *ottimo*: chiaramente, potrà esserci più di un criterio ottimale, ad ogni modo si considererà univocamente π^* come criterio ottimo, sia esso uno o più d'uno. Il criterio π^* permette di identificare i valori ottimi $V^*(s)$ e $Q^*(s, a)$ nel seguente modo:

$$V^*(s) = \max_{\pi} V^\pi(s) \tag{24}$$

$$Q^*(s, a) = \max_{\pi} Q^\pi(s, a) \tag{25}$$

per ogni $s \in S$ ed $a \in A(s)$. Entrambe le equazioni, essendo funzioni di valore, soddisfano le proprietà di Bellman enunciate in precedenza, rispettivamente la (22) e (23). Inoltre, dal momento che la (24) e la (25) sono funzioni ottime, si possono esprimere senza riferimento a criterio alcuno (π). In particolare, $V^*(s)$ si definisce *equazione ottima di Bellman* ed esprime l'uguaglianza tra il valore di uno stato sotto un criterio ottimo π^* ed il valore del rendimento atteso per l'azione migliore derivante dal suddetto stato, ovvero:

$$\begin{aligned}
V^*(s) &= \max_{a \in A(s)} Q^{\pi^*}(s, a) \\
&= \max_a E_{\pi^*}(R_t | s_t = s, a_t = a) \\
&= \max_a E_{\pi^*}(\sum_{k=0}^T \gamma^k r_{t+k+1} | s_t = s, a_t = a) \\
&= \max_a E_{\pi^*}(r_{t+1} + \sum_{k=0}^T \gamma^k r_{t+k+2} | s_t = s, a_t = a) \\
&= \max_a E(r_{t+1} + \gamma V^*(s_{t+1}) | s_t = s, a_t = a). \tag{26}
\end{aligned}$$

L'equazione ottima di Bellman per $Q^*(s, a)$ si esprime invece nel seguente modo:

$$Q^*(s, a) = \max_a E(r_{t+1} + \gamma V^*(s_{t+1}) | s_t = s, a_t = a), \tag{27}$$

ricordando che r_{t+1} è un valore dipendente da s_t, a_t, s_{t+1} e si può indicare anche con la notazione $r(s_t, a_t, s_{t+1})$. Per processi decisionali *à la Markov* di ordine finito, le equazioni (26) e (27) hanno un'unica soluzione indipendente dal criterio π selezionato. Una volta ottenuta $V^*(s)$, è relativamente semplice determinare un criterio ottimo: per ogni stato s , vi sarà un'azione (o più azioni) mediante la quale ottenere il valore massimo nell'equazione (26) e che condurrà allo stato successivo. In altri termini, ogni criterio *greedy*⁴³ relativamente al valore ottimo della funzione $V^*(s)$ è un criterio ottimo. Se in

⁴³ Si è scelto, in questo caso, di mantenere il termine originale utilizzato in Barto e Sutton (2016), senza addentrarsi in alcun tentativo di traduzione. Il termine, che letteralmente significa *ingordo, famelico*, viene utilizzato in informatica e programmazione per indicare una procedura di scelta che seleziona alternative sulla base solamente di

questo modo si selezionano azioni sulla base della loro efficacia nel breve periodo, $V^*(s)$ considera nella sua formulazione – e questa è la validità più grande dell’equazione ottima di Bellman – i rendimenti conseguenti a tutti i possibili comportamenti futuri, in un’ottica di lungo termine, dunque una ricerca dell’ottimo step-by-step utilizzando l’equazione di Bellman produce azioni ottime nel lungo periodo.

3.2.4 Iterazione e valutazione del criterio π

Giunti a questo punto, è possibile calcolare *iterativamente* la funzione di valore per uno stato, o equivalentemente per una coppia stato-azione (Corazza e Bertoluzzo, 2012). Si consideri un arbitrario valore iniziale $V_0^\pi(s_t)$, per ogni $s \in S$: ogni sua successiva approssimazione sarà ottenuta utilizzando l’equazione (22) come regola di aggiornamento, ovvero

$$V_{k+1}^\pi(s_t) = E_\pi(r_{t+1} + \gamma V_k^\pi(s_{t+1}) | s_t = s). \quad (28)$$

Se $V_{k+1}^\pi(s_t)$ esiste, allora la sequenza $\{V_k^\pi\}$ converge al suo valore ottimo V^* per $k \rightarrow \infty$, sotto la condizione che $\gamma < 1$; tale algoritmo viene definito *policy evaluation* (ovvero *valutazione del criterio π* (Barto e Sutton, 2016). Formalmente tale valutazione converge solo in limite al valore V^* , nella pratica tuttavia viene iterata solamente fino ad una condizione finale, tipicamente ottenuta testando la quantità $|V_{k+1}^\pi(s) - V_k^\pi(s)|$ dopo ogni iterazione e stoppando l’algoritmo quando questa approssima a zero. Uno possibile passo successivo è quello di ricercare, attraverso la valutazione effettuata dall’equazione (28), un criterio π migliore del precedente, se esiste, al fine di incrementare il valore atteso dei rendimenti futuri: tale processo viene definito *policy improvement* (letteralmente, *perfezionamento del criterio comportamentale*). Nell’ordine di implementare tale processo di perfezionamento, è sicuramente necessario partire dall’equazione (28) che, come visto, è un processo iterativo

considerazione *locali* o *immediate*. Applicato al contesto, il termine descrive quei criteri che selezionano azioni sulla base della loro conseguenza nel breve periodo (Barto e Sutton, 2016).

convergente al suo ottimo in limite; non è però necessario ricercare l'esatta convergenza, esistono infatti numerose tecniche di arresto del processo, una delle quali consiste nella combinazione tra il processo valutativo ed il processo di perfezionamento, stoppando la valutazione ad ogni istante t e perfezionando lo stesso. La fusione tra questi due processi genera il seguente algoritmo, definito *iterazione generalizzata del criterio π* (Corazza e Bertoluzzo, 2012) (Barto e Sutton, 2016):

$$V_{k+1}^{\pi}(s_t) = \max_a E(r_{t+1} + \gamma V_k^{\pi}(s_{t+1}) | s_t = s, a_t = a), \quad (29)$$

dove $V_{k+1}^{\pi}(s_t)$ è la stima aggiornata con il nuovo criterio perfezionato allo step $k + 1$, rispetto alla stima del valore ed al criterio dello step precedente k .

L'azione da intraprendere ad ogni step, al fine di valutare il miglior criterio π , può essere selezionata seguendo numerosi approcci differenti, uno dei quali prevede il seguente sistema:

$$a_t = \begin{cases} \pi'(s_t) & \text{con probabilità } 1 - \varepsilon \\ a \in A(s_t) & \text{con probabilità } \varepsilon \end{cases} \quad (30)$$

con $\varepsilon \in (0,1)$ e $\pi'(s_t)$ che è il criterio *greedy* che massimizza l'equazione $Q^{\pi}(s, a)$, quindi definibile come

$$\begin{aligned} \pi'(s_t) &= \operatorname{argmax}_a Q^{\pi}(s, a) \\ &= \operatorname{argmax}_a E(r_{t+1} + \gamma V^{\pi}(s_{t+1}) | s_t = s, a_t = a) \end{aligned} \quad (31)$$

dove con argmax_a si intende quell'azione a al cui livello l'espressione (31) è massimizzata. Dunque, la selezione di un'azione si basa su un particolare criterio che, ad ogni step, verrà valutato in base a due alternative: un criterio, definito *ε -greedy*, selezionato con una probabilità pari ad $1 - \varepsilon$ ed un criterio alternativo, selezionato con una probabilità pari ad ε . Questo ultimo criterio viene definito *policy di esplorazione* poiché andrà a selezionare un set di azioni ancora inesplorate, da cui potrebbero derivare performance migliori, peggiori o dei criteri a loro volta *ε -greedy*. Dal momento

che questa valutazione può essere fatta solo *a posteriori*, ovvero una volta che l'ambiente fornirà all'agente l'output, non si avrà la certezza della selezione corretta fin tanto che l'azione selezionata non verrà testata.

In sintesi, il processo di valutazione ed il processo di perfezionamento interagiscono tra loro fino a stabilizzarsi convergendo assieme ad un'unica soluzione, che per le proprietà enunciate finora sarà la soluzione ottima del processo. A livello grafico, con le dovute semplificazioni, è possibile apprendere tale processo osservando la *figura 5* riportata di seguito: ogni processo conduce la funzione di valore iniziale (V_0) o il criterio (π_0) verso una delle due linee rappresentanti la soluzione di uno dei due obiettivi (V^*, π^*). Inevitabilmente, il processo congiunto condurrà verso una soluzione comune, prossima o uguale alle condizioni di ottimo V^* e π^* : le frecce nel diagramma infatti rappresentano il comportamento del processo di interazione che tenta di raggiungere completamente e soddisfacentemente una delle due condizioni di ottimo, finendo per convergere ad una soluzione comune.

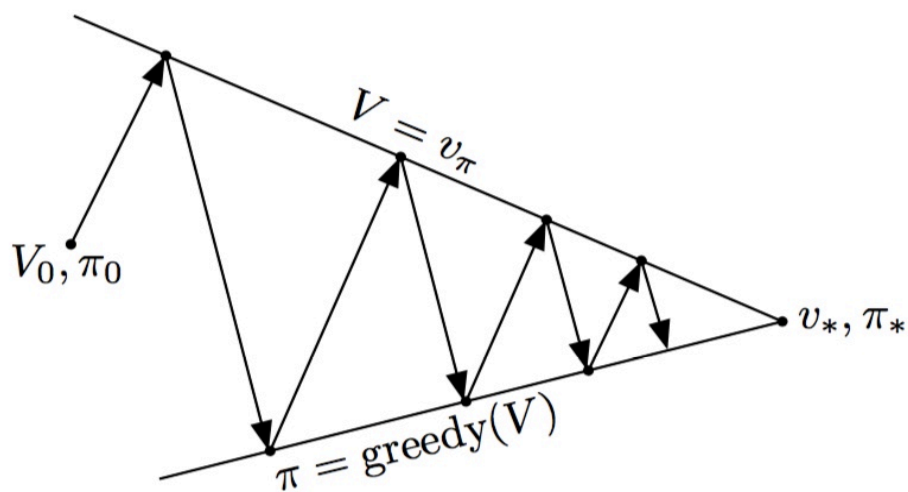


Figura 5: rappresentazione grafica del processo di iterazione generalizzata del criterio π (Fonte: Barto e Sutton, 2016)

Infine, dunque, sono stati introdotti tutti gli elementi essenziali di un approccio RL; al fine di calcolare il rendimento atteso dell'equazione (29) viene scelto di sviluppare un algoritmo Q-Learning, procedimento facente parte della classe di metodi per differenze temporali, di cui di seguito si presentano le principali ed essenziali caratteristiche, per poi addentrarsi nelle specifiche dell'algoritmo scelto e nelle sue implicazioni pratiche.

3.3 Apprendimento per differenze temporali: i metodi TD

I metodi per differenze temporali (*temporal-difference learning methods*, d'ora in avanti abbreviati con TD) sono una classe di metodi RL in grado di apprendere gli input dall'ambiente senza bisogno di un modello che ne descriva le dinamiche. L'appellativo TD proviene dall'impiego di tali metodi dei cambiamenti, o differenze, tra due step temporali consecutivi al fine di implementare il processo di apprendimento: la previsione ad ogni istante t sarà aggiornata per avvicinarla alla previsione dello stesso valore al successivo istante $t+1$, senza dover attendere l'*outcome* finale del processo e basandosi in parte su una stima del valore già esistente⁴⁴. I metodi TD, dunque, utilizzano l'esperienza passata e recente per risolvere ed aggiornare *step-by-step* il processo di valutazione: all'istante di tempo $t+1$, tali metodi permettono un istantaneo aggiornamento, utilizzando il rendimento ottenuto r_{t+1} e la stima $V_k(s_{t+1})$.

Il metodo TD più semplice, definito con $TD(0)$, è il seguente:

$$V_{k+1}(s_t) = V_k(s_t) + \alpha[r_{t+1} + \gamma V_k(s_{t+1}) - V_k(s_t)]. \quad (32.1)$$

Tale regola di aggiornamento è ricorrente nei metodi RL, come si avrà modo di appurare in seguito, e segue il seguente schema generale

$$Stima_{n+1} \leftarrow stima_n + step_size[target - stima_n]$$

Nel caso dell'equazione (32.1), il *parametro step-size*, definito anche *tasso di apprendimento*, si indica con $\alpha \in (0,1]$, γ è il già noto tasso di sconto e la differenza tra parentesi si definisce *errore TD* ed esprime il discostamento tra il valore $V_k(s_t)$ e il target $r_{t+1} + \gamma V_k(s_{t+1})$, auspicabilmente indicante la *direzione* verso la quale indirizzare il comportamento futuro. Questa quantità (indicata con δ_t) fornisce dunque, ad ogni istante di tempo t , l'errore di stima di quell'istante temporale; essendo dipendente dal valore dello stato e dal rendimento di uno step successivo, chiaramente tale quantità sarà calcolabile successivamente alle suddette realizzazioni e l'algoritmo implementato

⁴⁴ Tale comportamento si definisce *bootstrap*, tecnica che prevede l'implementazione di un processo complesso partendo da limitati input iniziali.

sarà in grado di aggiornare la stima non appena la quantità δ_t sarà disponibile. Dunque, l'algoritmo TD(0) ad ogni step, in base alle osservazioni r_{t+1} ed s_{t+1} , si aggiornerà come segue:

$$V_{k+1}(s_t) = \begin{cases} V_k(s_t) + \alpha\delta_t & \text{se } s = s_t \\ V_k(s_t) & \text{negli altri casi} \end{cases} \quad (33)$$

dove $s \in S$ indica ogni possibile segnale di input. Come già anticipato, i metodi TD non richiedono un modello ambientale e non necessitano delle distribuzioni di probabilità dei rendimenti o delle probabilità di transizione precedentemente enunciate. Tali metodi sono sviluppati in un modo *incrementale on-line* (dal momento che si aggiornano alla fine di ogni step senza dover attendere lo stato *terminale* T), caratteristica che li rende estremamente utili ed interessanti per la costruzione di sistemi di trading finanziario (Corazza e Bertoluzzo, 2012).

Partendo da questa ultima affermazione, si considererà certamente che i mercati finanziari si distinguono per presentare serie dei prezzi azionari non stazionarie, come si è ampiamente esposto nel Capitolo 2. In questi casi, sarebbe più sensato pesare in misura maggiore i rendimenti più recenti rispetto a quelli più lontani nel passato. Tale relazione si può ottenere utilizzando un parametro *step-size* costante; per dimostrare ciò si consideri l'equazione (32) dalla quale, sostituendo ricorsivamente a ritroso i valori, si giunge alla seguente relazione⁴⁵:

$$V_{k+1}(s_t) = (1 - \alpha)^k V_k(s_0) + \sum_{t=1}^k \alpha(1 - \alpha)^{k-t} [r_t + \gamma V_k(s_t)], \quad (32.2)$$

dove $V_k(s_0)$ è il valore arbitrario dello stato iniziale s_0 . L'equazione (34) viene definita *media pesata*, poiché la somma dei pesi $(1 - \alpha)^k + \sum_{t=1}^k \alpha(1 - \alpha)^{k-t} = 1$. Inoltre, il peso $\alpha(1 - \alpha)^{k-t}$ definito per il rendimento $R_t = r_t + \gamma V_k(s_t)$ dipenderà da quanti $k - t$ rendimenti passati sono stati osservati e, dal momento che $1 - \alpha < \alpha$, decrescerà esponenzialmente all'aumentare del numero dei rendimenti (Barto e Sutton, 2016).

⁴⁵ Si veda Barto e Sutton (2016), pag. 34, dove sono approfonditamente riportati tutti i passaggi ricorsivi per ottenere l'equazione 32.2.

Utilizzando un parametro variabile e non negativo ad ogni istante di tempo (ad esempio $\alpha_k(\alpha) = \frac{1}{k}$), in un ambiente stocastico entrambe le condizioni sufficienti di convergenza

$$\sum_{k=1}^{\infty} \alpha_k(\alpha) = \infty \text{ e } \sum_{k=1}^{\infty} \alpha_k^2(\alpha) < \infty \quad (34)$$

vengono rispettate con probabilità pari a 1 (Bertsekas e Tsitsiklis, 1996), dove con la prima condizione è assicurata l'elevata numerosità degli step da effettuare per superare eventuali condizioni di instabilità iniziali, la seconda garantisce invece la convergenza al valore ottimo all'approssimare di $\alpha_k(\alpha)$ verso un numero sempre più piccolo. Tali convergenze vengono rispettate nel caso di un parametro non costante poiché, nel caso in esempio, $\alpha_k(\alpha) = \frac{1}{k}$ decrescerà linearmente all'aumentare delle osservazioni realizzate. Se si considera $\alpha_k(\alpha)$ costante, la seconda condizione non verrà mai soddisfatta: la stima continuerà il suo processo di convergenza variando in continuo, in risposta ai valori dei rendimenti più recenti. Questa condizione è tipica in un ambiente non stazionario qual è quello dei mercati finanziari, nel quale è conveniente non appesantire il processo di ricerca includendo osservazioni dei rendimenti troppo lontane nel tempo per focalizzarsi sulle osservazioni più recenti. Chiaramente, se il parametro $\alpha_k(\alpha)$ viene settato sufficientemente piccolo, la convergenza di $V_k(s_t)$ al valore ottimo V^* sarà ottenuta sia in condizioni di stazionarietà che in condizioni di non stazionarietà. È dunque fondamentale ricercare il settaggio migliore per il parametro *step-size* costante, considerando comunque che più piccolo questo viene settato, più lento sarà il processo di convergenza al valore ottimo della funzione.

Fino a questo punto, si è argomentato come i metodi TD trattino il processo di valutazione per stimare il valore di V_k dato un qualsiasi criterio π . Per quanto concerne invece il problema di *controllo* dei metodi RL, ovvero quello di ricercare la *policy* ottima, i metodi TD utilizzano il processo di *iterazione generalizzata del criterio π* visto in precedenza approcciando il problema in un modo peculiare. Ancora una volta, si presenta il *leitmotiv* del *trade-off* tra esplorazione e sfruttamento (pag. 46), e nuovamente l'approccio si dirama in due sotto categorie: metodi *on-policy* e metodi *off-policy*. Dal momento che il presente lavoro è focalizzato sull'algorithm Q-Learning, nel

prossimo paragrafo ci si concentrerà sulle specificità di questo metodo *off-policy*, lasciando ad altri testi eventuali approfondimenti sui metodi *on-policy*⁴⁶.

3.4 Q-Learning: algoritmo *off-policy*

Il metodo Q-Learning appartiene alla classe di metodi per differenze temporali e si configura come un algoritmo di controllo *off-policy* che aggiorna *step by step* la stima della coppia stato-azione del valore ottimo $Q^*(s, a)$, in base al rendimento ottenuto nell'intervallo di valutazione considerato. Tale metodo si definisce *off-policy* poiché presenta la peculiarità di approssimare direttamente la funzione di valore $Q_k(s, a)$ al suo ottimo $Q^*(s, a)$ in misura indipendente rispetto al criterio π attuale (Barto e Sutton, 2016). Tale algoritmo, inoltre, si definisce sia *online* che algoritmo di *controllo*: *online* poiché permette di aggiornare la stima della funzione di valore ad ogni step senza dover attendere lo stato *terminale* (comportamento che assumono, ad esempio, le tecniche di Simulazione Montecarlo, non trattate in questa sede), dunque la stima di $Q^*(s, a)$ è disponibile dopo ogni step ed è già utilizzabile per lo stato seguente; viene definito algoritmo di *controllo* poiché intraprende azioni al fine di raggiungere il proprio scopo, ovvero l'ottimizzazione dell'ottimo della funzione di valore $Q^*(s, a)$. La formula dell'algoritmo si presenta come segue:

$$Q_{k+1}(s_t, a_t) = Q_k(s_t, a_t) + \alpha \left[r_t + \gamma \max_a Q_k(s_{t+1}, a) - Q_k(s_t, a_t) \right]. \quad (35)$$

La procedura dell'algoritmo Q-Learning, in linea generale, può invece essere rappresentata come segue (Barto e Sutton, 2016):

- Inizializzare $Q(s_0, a_0), \forall s \in S, a \in A(s)$, arbitrariamente
- Ripetere (per ogni stato)
 - Inizializzare S
 - Ripetere (per ogni stato)

⁴⁶ Si veda Barto e Sutton, pagg. 137 e seg.

- Selezionare a_t in base ad s_t mediante la policy derivante da Q
 - Intraprende l'azione a_t , osservare r_{t+1} ed s_{t+1}
 - $Q_{k+1}(s_t, a_t) \leftarrow Q_k(s_t, a_t) + \alpha[r_t + \gamma \max_a Q_k(s_{t+1}, a) - Q_k(s_t, a_t)]$
 - $s_t \leftarrow s_{t+1}$
- T stato terminale.

Finora le assunzioni fatte prevedevano l'insieme degli stati possibili del mondo S come insieme di variabili discrete, ed assunto un numero limitato e finito di valori. Chiaramente, se si è intenzionati a rappresentare un sistema dei possibili stati S dei mercati finanziari, questo sarà caratterizzato da stati continui nel tempo i quali assumono un numero infinito di possibili valori (Corazza e Bertoluzzo, 2012). Da qui, si evidenzia la necessità di ricercare non tanto una *policy* ottima oppure l'ottimo di una funzione di valore, ma l'obiettivo è la ricerca di una *buona soluzione approssimata*, utilizzando risorse computazionali limitate. La chiave risolutiva per tale *framework* è la combinazione tra i metodi RL ed i metodi di generalizzazione, ed il risultato sarà un'*approssimazione della funzione di valore*, la quale trarrà i valori da una funzione desiderata, tentando di generalizzarli al fine di creare un'approssimazione attendibile dell'intera funzione di valore stessa. Inizialmente dunque si descriverà tale processo di generalizzazione che verrà successivamente applicato all'algoritmo Q-Learning e fornirà la base su cui poi implementare il sistema di trading.

3.4.1 Approssimazione della funzione di valore

Il primo passo nella costruzione di una approssimazione della funzione di valore è la scelta di un'*architettura* di approssimazione, ovvero la scelta di una determinata funzione rappresentata da un numero definito di parametri. Nell'ordine di costruire un'efficace approssimazione, vi sono due importanti pre-condizioni (Bertsekas e Tsitsiklis, 1996): innanzitutto, è necessario che la struttura sia sufficientemente articolata da consentire un'accettabile approssimazione della funzione di valore che si sta stimando; in secondo luogo, è necessario che l'algoritmo implementato effettivamente calibri i parametri selezionati dalla struttura di approssimazione. Al fine

di approssimare la funzione di valore $Q(s_t, a_t)$, viene impiegata una forma di funzione parametrizzata mediante un vettore di pesi $\boldsymbol{\theta} \in \mathbb{R}^n$, composto da un numero fisso di numeri reali e rappresentabile nel seguente modo:

$$\boldsymbol{\theta}_t = [\theta_1 \quad \theta_2 \quad \theta_3 \quad \dots \quad \theta_n]'. \quad (36)$$

Utilizzando il vettore $\boldsymbol{\theta}$, è ora possibile riscrivere $Q(s_t, a_t)$ come

$$Q(s_t, a_t) \approx Q(s_t, a_t, \boldsymbol{\theta}_t), \quad (37)$$

dove $Q(s_t, a_t, \boldsymbol{\theta}_t)$ è una funzione differenziabile di $\boldsymbol{\theta}$ per ogni $s \in S$. Tipicamente, il numero dei n parametri del vettore, che varieranno *step by step*, sarà inferiore alla numerosità dei possibili stati del mondo ($n \ll S$) e, conseguentemente, la variazione di un solo componente del vettore può comportare la variazione della stima del valore di molti stati: tale generalizzazione rende il processo di apprendimento più potente ma anche molto più difficile da gestire. Data tale assunzione (più stati del mondo rispetto ai parametri del vettore $\boldsymbol{\theta}_t$), è necessario individuare un indicatore che *minimizzi* la distanza tra il valore ottenuto dall'approssimazione $Q(s_t, a_t, \boldsymbol{\theta}_t)$ ed il valore dell'effettiva funzione di valore $Q^*(s, a)$. Nella maggior parte dei casi, viene ricercata la minimizzazione dell'*errore quadratico medio* (MSE in acronimo)⁴⁷, ovvero della sommatoria delle differenze al quadrato tra le realizzazioni:

$$\min_{\boldsymbol{\theta}_t} \sum_{s \in S, a \in A(s)} [Q^*(s, a) - Q(s_t, a_t, \boldsymbol{\theta}_t)]^2. \quad (38)$$

L'obiettivo ideale, in termini di MSE, è quello di identificare un punto di *minimo globale*, ovvero quel vettore $\boldsymbol{\theta}^*$ per il quale $MSE(\boldsymbol{\theta}^*) \leq MSE(\boldsymbol{\theta})$ per ogni possibile settaggio di

⁴⁷ La letteratura non è ancora così sufficientemente concorde sull'utilizzo del MSE come strumento di performance per i metodi RL. In Barto e Sutton (2016), così come in Moody J. *et al.* (1998), viene infatti specificato come non sempre il valore migliore della funzione $Q(s_t, a_t)$, settato al fine di migliorare costantemente il criterio π , sia il valore migliore per la minimizzazione dell'equazione (38). Ad ogni modo, ad oggi, questo sembra essere il miglior strumento possibile da poter utilizzare.

θ . Tale risultato non viene facilmente raggiunto, poiché la funzione di valore $Q^*(s, a)$ non è nota: dal momento che il processo di ottimizzazione viene aggiornato *step-by-step*, l'equazione (38) potrebbe identificare nel corso del processo di ottimizzazione un punto di *ottimo globale* ed arrestare così la minimizzazione, essendo invece tale risultato un punto di *ottimo locale*. L'obiettivo dunque si riconduce all'individuazione di punti di *ottimo locali*, nel caso in esame quel vettore θ^* per il quale $MSE(\theta^*) \leq MSE(\theta)$ per un certo intorno di θ^* , iterando il processo di ricerca un numero elevato di volte al fine di ottenere un numero elevato di minimi locali, ricercando poi in quell'insieme l'*ottimo globale* che si configura come una *soluzione near-optimal*. D'altro canto, il processo di approssimazione della funzione, che è la chiave dei metodi RL, rimane ancora impregnato di una componente quasi *artistica* rispetto alla sua componente *scientifica* rilevante, ed è un passaggio obbligato assumere dei compromessi computazionali al fine di semplificare un processo che altrimenti risulterebbe troppo complesso da implementare (Gosavi, 2003). Nel prossimo paragrafo si entrerà nel dettaglio del processo di minimizzazione, introducendo il metodo della *discesa del gradiente*, innumerevolmente impiegato nei processi di approssimazione di funzioni e che permette, sotto certe condizioni di seguito esposte, di generare soluzione ottime.

3.4.2 Metodo della *discesa del gradiente*

Come riporta Gosavi (2003), il metodo della discesa del gradiente garantisce la generazione di una soluzione di *ottimo globale* solo in presenza delle seguenti condizioni:

- α , parametro *step-size* o *di apprendimento*, sufficientemente piccolo.
- Funzione di approssimazione che soddisfi certe condizioni relative alla *convessità*, *continuità*, *differenziabilità*.

Al fine di non scopercchiare, come si suole dire, un *vaso di Pandora* e dilungarsi nell'argomentare dimostrazioni e proprietà a supporto dei due punti sopra elencati, in questa sede si cercheranno di riassumere tutti i risultati chiave funzionali a dimostrare

come la scelta di tale metodo sia soddisfacente, non esimendosi dal fornire tuttavia tutti i riferimenti necessari da poter consultare per approfondire i tecnicismi che conducono a tali risultati.

Il metodo di *discesa del gradiente* opera, in linea generale, nel procedere da un'iterazione alla sua consecutiva, rispettando la seguente regola:

$$x(i) \leftarrow x(i) - \alpha \left. \frac{df(x)}{dx(i)} \right|_{x=x^m} \quad (39)$$

dove l'interesse è la minimizzazione della funzione f ; in particolare, $\left. \frac{df(x)}{dx(i)} \right|_{x=x^m}$ denota lo scalare della derivata parziale di f rispetto ad $x(i)$, alla m -esima iterazione, ed α è il già noto parametro step-size. Qualora, per ogni derivata parziale ottenuta ($i = 1, \dots, k$), il valore ottenuto sia pari a 0, l'algoritmo non è in grado di aggiornare ulteriormente le variabili $x(i)$ e viene raggiunto un punto di ottimo (Gosavi, 2003).

Richiamando le equazioni (36), (37) e (38) viste in precedenza, il metodo della discesa del gradiente, ad ogni step t , minimizza l'MSE tra il valore effettivo (e sconosciuto) Q^* e la sua approssimazione $Q(s_t, a_t, \theta_t)$, aggiustando gli elementi del vettore θ attraverso la seguente regola di aggiornamento:

$$\begin{aligned} \theta_{t+1} &= \theta_t - \frac{1}{2} \alpha \nabla_{\theta_t} [Q^*(s, a) - Q(s_t, a_t, \theta_t)]^2 \\ &= \theta_t + \alpha [Q^*(s, a) - Q(s_t, a_t, \theta_t)] \nabla_{\theta_t} Q(s_t, a_t, \theta_t), \end{aligned} \quad (40)$$

dove $\nabla_{\theta_t} f(\theta_t)$ rappresenta il vettore di derivate parziali delle componenti di θ_t , $\nabla_{\theta_t} f(\theta_t) = \left(\frac{df(\theta_t)}{d\theta_{t_1}}, \frac{df(\theta_t)}{d\theta_{t_2}}, \dots, \frac{df(\theta_t)}{d\theta_{t_n}} \right)'$, definito gradiente di $f(\theta_t)$. Tale metodo si definisce discesa del gradiente poiché ogni step θ_{t+n} è proporzionale all'opposto del gradiente dell'equazione (39), il quale fornisce la direzione verso la quale l'MSE tende a zero nella maniera più rapida. Il metodo del gradiente si definisce stocastico poiché l'aggiornamento della stima avviene solamente in una singola direzione, selezionata stocasticamente secondo l'equazione (30). Iterando il processo di ricerca più volte

l'effetto globale è la minimizzazione in media della misura di performance MSE (Barto e Sutton, 2016). Inoltre, se il parametro step-size rispetta le condizioni sufficienti di convergenza (34) viste in precedenza, il metodo di discesa del gradiente garantisce la convergenza verso un punto di minimo locale⁴⁸.

Utilizzando il metodo del gradiente, idealmente le aspettative sono quelle di generare una sequenza θ_t che converga ad un punto di *ottimo globale*. Sfortunatamente, tramite tale metodo, vengono individuati punti *stazionari* i quali, non conoscendo la forma esatta della funzione Q^* , potrebbero essere punti di *ottimo locale* o *globale*. Come anticipato in precedenza, affinché il punto stazionario raggiunto sia identificato come *ottimo globale* è necessario stabilire le condizioni di *convessità* della funzione; tali condizioni, spesso, non sono soddisfatte: in particolare, una funzione *non convessa* contiene multipli punti di *minimo locale*, verso ognuno dei quali il metodo di discesa del gradiente può essere *attratto* e perciò *intrappolato* (Gosavi 2003). Perciò, in presenza di funzioni non convesse, trattare con punti di minimo locale è una delle fondamentali debolezze del metodo di discesa del gradiente; un modo per eludere tale problema è lanciare l'algoritmo di ottimizzazione un numero sufficientemente elevato di volte, settando un punto arbitrario di partenza ad ogni diversa iterazione, e dichiarando il miglior punto di *ottimo locale* ottenuto come il risultato del processo. Chiaramente, tale approccio non garantisce che la miglior soluzione trovata sia un *ottimo globale*, ma frequentemente si configura come il miglior metodo applicabile in problemi di tal specie. Ciò che dunque occorre, infine, è garantire che, utilizzando un parametro *step-size* costante⁴⁹, il metodo di discesa del gradiente converga ad un punto *stazionario*: tale convergenza è stata dimostrata innumerevoli volte in letteratura, così come le sue condizioni iniziali sono incluse nella maggior parte degli algoritmi utilizzati per la risoluzione di metodi RL, tra cui l'algoritmo Q-Learning scelto in questo ambito⁵⁰.

⁴⁸ Si veda Bertsekas e Tsitsiklis (1996), pagg. 96-98, per la dimostrazione.

⁴⁹ Come si ricorderà, tale settaggio viene utilizzato in ambienti non stazionari quali sono i mercati finanziari.

⁵⁰ Si veda, per la dimostrazione, Gosavi (2003) pagg. 323-326 o anche Bertsekas e Tsitsiklis (1996) pagg. 94-96.

Tornando all'equazione (39), dal momento che il valore ottimo $Q^*(s, a)$ non è conosciuto, deve essere sostituito con la sua approssimazione $q_t = r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \theta_t)$, trasformando così l'equazione (39) nella seguente forma compatta:

$$\theta_{t+1} = \theta_t + \alpha \delta_t \nabla_{\theta_t} Q(s_t, a_t, \theta_t), \quad (41)$$

dove δ_t è il già noto *errore* TD, in questo caso corrispondente a $r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \theta_t) - Q(s_t, a_t, \theta_t)$. Se q_t è una stima *corretta*, ovvero $E(q_t) = Q^*(s, a)$ per ogni t , allora θ_t converge ad un punto di *ottimo locale*, sotto le condizioni di α citate in precedenza, sia esso costante o decrescente linearmente. L'equazione (41) nella sua forma finale prende il nome di metodo del *semi-gradiente*, poiché considera solo una parte della direzione del gradiente $\nabla_{\theta_t} Q(s_t, a_t, \theta_t)$ (basata sul valore del parametro α) ed un'approssimazione q_t del valore effettivo della funzione obiettivo $Q^*(s, a)$. La convergenza del vettore θ_t al suo ottimo θ^* è comunque garantita se la funzione $Q(s_t, a_t, \theta_t)$ è un'approssimazione di forma *lineare* rispetto al vettore dei pesi θ_t , caso preso in esame nel paragrafo seguente.

3.4.3 Forma lineare della funzione $Q(s, a, \theta)$

Uno dei metodi di approssimazione più utilizzati per rappresentare la funzione $Q(s, a, \theta)$ è una funzione lineare del vettore dei pesi θ . In corrispondenza di ogni stato s_t , esiste un vettore di valori reali $\phi(s_t) = [\phi_1(s_t), \phi_2(s_t), \dots, \phi_n(s_t)]$ con n numero di componenti del vettore equivalenti a θ_t . Da Barto e Sutton (2016), la funzione di valore stato-azione approssimata è ottenuta dunque dal prodotto interno tra θ' ed $\phi(s)$:

$$Q(s, a, \theta) = \theta' \phi(s) = \sum_{i=1}^n \theta_i (i) \phi_i(s), \quad (42)$$

dove $\phi_i(\cdot)$ si definisce funzione *squashing* (Corazza e Bertoluzzo, 2012) dello stato i -esimo e rappresenta un'adeguata trasformazione dello stato. Tale trasformazione è

ricercata nell'ordine di ottenere valori che siano in grado di aumentare la sensibilità del processo di calcolo della funzione di valore $Q(s, a)$. La proprietà fondamentale che ogni funzione $\Phi_i(\cdot)$ deve possedere è la *monotonia crescente*, per mantenere inalterato l'ordine dei valori trasformati mediante l'approssimazione. In letteratura sono presenti svariati esempi di equazioni lineari utilizzate per fini approssimativi (si veda, ad esempio, Barto e Sutton 2016, cap.9), in questo caso si riportano due esempi di funzioni utilizzate in Corazza e Bertoluzzo (2015) e in Corazza e Sangalli (2015), ovvero:

- $\tanh(x)$ tangente iperbolica, definita come il rapporto tra $\sinh(x)$ e $\cosh(x)$ ⁵¹, la quale è una funzione monotona strettamente crescente e limitata, con immagine $Im(f) = (-1; +1)$;
- funzione logistica, anch'essa strettamente crescente e definita, nel caso in esame, come

$$\Phi(x) = \frac{a}{1 + be^{-cx}} + d \quad (43)$$

con e numero di Nepero ed $a, b, c, d > 0$ coefficienti della funzione. Nel capitolo seguente, in particolare, verranno esposti settaggio e motivo della scelta di questa particolare funzione di approssimazione.

Tornando alla forma generale dell'equazione (42), dunque, la funzione di valore approssimata viene definita *lineare nei pesi*, o semplicemente *lineare*. Dal momento che è un processo di approssimazione semplice, il metodo del *semi-gradiente lineare* è uno dei più utilizzati nelle analisi matematiche, e quasi tutti i significativi risultati di convergenza di sistemi di apprendimento così strutturati passano attraverso l'utilizzo di metodi di approssimazione lineare di funzioni.

Si è dunque ampiamente dibattuto sulla struttura che debba avere un metodo RL basato sulle differenze temporali e implementato tramite un algoritmo Q-Learning. Nel prossimo capitolo si presenterà un caso pratico nel quale si utilizzerà tale struttura e tale

⁵¹ $\sinh(x) = \frac{e^x - e^{-x}}{2}$; $\cosh(x) = \frac{e^x + e^{-x}}{2}$

algoritmo, implementato utilizzando il software di programmazione *Matlab*[®], per verificare se si possano ottenere risultati di *performance* soddisfacenti tramite una configurazione innovativa com'è quella appena presentata, applicando tali concetti a dati reali di un paniere di titoli peculiarmente selezionato.

4. APPLICAZIONE PRATICA DEL QLa: STRUTTURA

In questo capitolo viene presentata la struttura pratica dell'algoritmo, cercando di dare risvolto empirico ai temi trattati nei capitoli due e tre. L'algoritmo Q-Learning è stato implementato in ambiente *Matlab*® ed applicato a 5 titoli reali, confrontando diversi settaggi, applicando tre diverse funzioni di *reward* e selezionando una funzione logistica come funzione *squashing*. Tutti gli elementi dell'algoritmo verranno di seguito presentati e descritti, per poi passare ad una breve presentazione delle serie storiche dei prezzi reali. Infine, nel capitolo 5, verranno presentati nel dettaglio risultati e commenti finali. L'algoritmo QL opera in quattro fasi distinte: una prima fase di *settaggio* nella quale vengono importati i dati storici, predisposti i parametri di riferimento e calcolati i rendimenti logaritmici giornalieri, una seconda fase di *computazione* tramite la quale si entra nel vivo del processo dell'algoritmo e vengono sondate un numero elevato di volte le possibili configurazioni delle variabili a_t , una terza fase *operativa* nella quale le azioni simulate vengono aggregate e viene indicata l'effettiva operatività del sistema di trading con la conseguente rappresentazione delle *equity line*, una quarta ed ultima fase di raccolta dati nella quale si ottengono statistiche e grafici finali di riferimento.

Ogni possibile configurazione viene testata un numero elevato di volte poiché, per i motivi esposti in precedenza, il processo di ottimizzazione viene approssimato utilizzando il vettore di pesi θ , il quale viene settato arbitrariamente ad ogni iterazione e perciò è necessario ripetere il processo svariate volte al fine di ottenere θ^* ; per lo stesso motivo, il processo di selezione di ogni azione a_t ad ogni step s_t viene iterato un numero elevato di volte al fine di ottenere un soddisfacente risultato che ponderi nella maniera più efficace possibile *exploration* ed *exploitation*. L'obiettivo è verificare se,

utilizzando dei settaggi e delle informazioni di input semplici, l'algoritmo riuscirà a fornire delle indicazioni di trading che portino a dei livelli di profitto soddisfacenti.

4.1 Elementi base dell'algoritmo

Si presentano di seguito gli elementi dell'algoritmo QL, ovvero le variabili utilizzate per descrivere lo stato s , le funzioni di *reward* utilizzate, i costi di transazione, la struttura che porta alla selezione dell'azione a e la funzione di *squashing*.

4.1.1 Descrittori dello stato s : rendimenti logaritmici

L'interesse principale nell'utilizzare una struttura QL applicata ad un sistema di *trading* finanziario è appurare se, anche utilizzando informazioni basilari per descrivere l'ambiente, l'algoritmo sia in grado di fornire indicazioni di operatività soddisfacenti. Inoltre, articolare una struttura troppo complessa allungherebbe i tempi di computazione, fattore non trascurabile in ambienti finanziari. Perciò, verranno utilizzati gli ultimi $N - 1$ rendimenti logaritmici di ogni titolo, il rendimento logaritmico corrente e l'ultima azione intrapresa a_{t-1} (Corazza e Sangalli 2015): ricordando che il rendimento logaritmico di un titolo è espresso dal seguente rapporto (con p_t si indica il prezzo del titolo all'istante t)

$$e_t = \log\left(\frac{p_t}{p_{t-1}}\right), \quad (44)$$

all'istante t lo stato s_t sarà descritto dal seguente vettore:

$$s_t = [e_{t-N+1}, e_{t-N+2}, \dots, e_{t-N+N-1}, e_t, a_t]. \quad (45)$$

È possibile assegnare diversi valori ad N al fine di considerare un numero più o meno elevato di rendimenti logaritmici passati. In questa applicazione si è scelto di considerare

due settaggi differenti, $N = 1$ ed $N = 5$, ovvero gli ultimi N rendimenti e l'ultima azione selezionata. La scelta di $N = 5$, ricorrente nella letteratura, è giustificata dal fatto che cinque sono i giorni settimanali nei quali è possibile operare in borsa, perciò rappresenta l'ultima settimana di rendimenti logaritmici giornalieri per ogni titolo.

4.1.2 La funzione di *reward* r_t

La funzione di ricompensa, o *reward*, fornisce il grado di soddisfazione dell'investitore ed è un feedback immediato che l'ambiente finanziario fornisce sul grado di efficienza delle azioni intraprese. In questa simulazione verranno prese in considerazione tre diverse funzioni di *reward*: lo Sharpe Ratio, largamente utilizzato nella letteratura empirica⁵² come misura di performance e perciò qui riproposto, e due misure *risk-adjusted* raramente considerate in letteratura, ovvero il Sortino Ratio e il Calmar Ratio, misure di performance molto in voga, specialmente nella *community* dei gestori di fondi investimento, ma non ancora approfonditamente testate in ambito di problemi RL applicati alla finanza.

Lo Sharpe Ratio è una delle più conosciute misure di rendimento *risk-adjusted*, e si ottiene dal seguente rapporto

$$SR_t = \frac{E_L(g_t)}{\sqrt{Var_L(g_t)}} \in \mathbb{R}, \quad (46)$$

dove SR_t è l'indice di Sharpe al tempo t e $E_L(g_t)$ e $Var_L(g_t)$ sono rispettivamente media e varianza campionarie calcolate sui rendimenti ottenuti negli ultimi L giorni di *trading* (sulla base delle azioni di volta in volta intraprese). Lo Sharpe Ratio si definisce misura *risk-adjusted* in quanto esprime la media dei rendimenti del periodo rapportati alla deviazione standard, esprimendo dunque una misura che penalizza periodi di alta volatilità nei rendimenti: più la volatilità è elevata, più lo SR_t diminuirà, fornendo una *reward* più bassa. Largamente utilizzato come indice di performance, SR_t risulta un

⁵² Si veda, ad esempio, Moody *et al.* (1998), o anche Corazza e Bertoluzzo (2012, 2015).

buon indice che permette di correggere i rendimenti rapportati alla volatilità. Tuttavia, applicando al denominatore la deviazione standard dei rendimenti del periodo considerato, vengono pesati in egual misura scostamenti positivi e negativi dalla media del periodo quando invece, in termini di rendimenti, gli scostamenti positivi non penalizzano i rendimenti ma sono invece preferibili poiché aumentano la performance del sistema di trading. Per tali motivi, viene confrontato l' SR_t con un'altra misura *risk-adjusted*, il Sortino Ratio, ottenuta dal seguente rapporto:

$$SOR_t = \frac{E_L(g_t)}{\sqrt{Var_L[\min(g_t - \bar{g}), 0]}} \in \mathbb{R}. \quad (47)$$

Al denominatore, al posto della deviazione standard degli ultimi L rendimenti, viene posto il *downside risk*, ovvero la deviazione standard campionaria che considera solo gli scostamenti negativi dalla media del periodo. In tal modo la performance sarà penalizzata solamente dalla volatilità di realizzazioni negative dei rendimenti, metodo che risulta essere più adatto in ambiente finanziario.

Il Calmar Ratio è un'altra misura *risk-adjusted* largamente utilizzata come indice di performance. Esistono diverse varianti in letteratura riguardo tale indice, in questa sede verrà utilizzata la versione presentata in Pederson e Rudholm-Alfvén (2003), ovvero il rapporto tra le seguenti quantità:

$$CAR_t = \frac{E_L(g_t)}{DDmax_L}, \quad (48)$$

dove con $DDmax_L$ si intende il valore massimo di *drawdown* delle ultime L realizzazioni della *equity line*, ovvero il massimo declino della serie da un punto di massimo ad un successivo punto di minimo all'interno del periodo considerato. Questa misura fornisce dunque la massima potenziale perdita di un periodo, valutando direttamente l'operatività dell'algoritmo. $DDmax$ va a sostituire, dunque, il *downside risk* dell'equazione (46) come misura di volatilità del titolo i -esimo considerato. Viste le similarità computazionali tra gli indici, l'obiettivo è dunque verificare quale delle misure di rischio sia più efficace nel cogliere la volatilità nell'andamento dei rendimenti di un

titolo e dunque di correggere con maggior accuratezza la media dei rendimenti $E_L(g_t)$. Dal momento che la letteratura empirica fa largo uso dello Sharpe Ratio come misura di performance, questo viene considerato in qualche maniera un *benchmark*, e si andrà ad appurare se ed in che modo l'indice di Sortino e il Calmar Ratio migliorino il processo di valutazione.

In sede computazionale, vengono testati due settaggi diversi di L , rispettivamente $L = 5$ ed $L = 22$, corrispondenti ad una settimana ed a circa un mese di rendimenti di mercato, settaggi per altro ricorrenti nella letteratura empirica analizzata. Le aspettative sono quelle di ottenere una qualche divergenza tra le funzioni di *reward* nel caso di $L = 22$, poiché nel caso di $L = 5$ il *sample* campionario non è così ampio da poter cogliere eventuali persistenze o oscillazioni dei rendimenti che provochino una differenza sostanziale tra i tre indici.

4.1.3 Costi di transazione δ

I costi di transazione vengono applicati se due operazioni di *trading* consecutive risultano diverse ($a_{t-1} \neq a_t$). L'ammontare di tali costi viene dimezzato poiché computato all'apertura e chiusura di ogni posizione e perciò i costi, indicati con δ , sono dati dalla seguente formula:

$$\delta = \frac{d}{2}, \quad (49)$$

dove con d si indica l'ammontare totale dei costi di transazione. Viene computata una sola tipologia di costi di transazione, ovvero una *percentuale* applicata al rendimento (dunque con impatto variabile sui potenziali guadagni) pari a 0,10%, valore applicato allo scambio di azioni, ETF e Fondi da svariate compagnie di brokeraggio italiane⁵³. I costi di transazione vengono applicati sia in fase operativa che in fase di computazione, in questo modo si rende l'algoritmo sensibile ai possibili costi applicati ai guadagni; in fase

⁵³ Fonte: <https://www.interactivebrokers.com/it/index.php?f=5688&p=stocks1>.

di computazione, l'ammontare dei costi viene applicato ai valori della *equity line*⁵⁴ utilizzando, dopo ogni step t , la seguente formula:

$$eq_line_{t+1} = eq_line_t * (1 + e_{t+1}a_t - \delta|a_t - a_{t-1}|). \quad (50)$$

Il valore assoluto $|a_t - a_{t-1}|$ viene utilizzato proprio per applicare la percentuale δ solamente in presenza di due azioni divergenti.

4.1.4 Struttura di selezione di a_t

Per quanto concerne il sistema di possibili azioni, si utilizza l'impostazione presentata in Corazza e Bertoluzzo (2012, 2015), ovvero il seguente sistema:

$$a_t = \begin{cases} -1 \\ 0 \\ 1 \end{cases} \quad (51)$$

dove con $a_t = -1$ si intende il segnale "vendita o posizione *short* nel mercato", con l'azione $a_t = 0$ si intende il segnale "star fuori dal mercato" e con l'azione $a_t = 1$ si intende "acquisto o posizione *long* sul mercato". L'azione a_t da intraprendere sarà fornita solamente alla conclusione dello step t , perciò una sua valutazione a fini di performance potrà essere effettuata solamente all'istante $t + 1$ una volta ottenuto, qualora non si sia intrapresa un'azione $a_t = 0$, il rendimento del titolo selezionato, r_{t+1} . Richiamando i costrutti teorici espressi al paragrafo 3.2.4, sarà chiaro che l'azione a_t non sarà selezionata solamente tramite l'equazione (45) ma anche tramite il metodo del *criterio ε -greedy*, ovvero

$$a_t = \begin{cases} \operatorname{argmax}_{a_t} Q(s_t, a_t, \theta_t) & \text{con probabilità } 1 - \varepsilon \\ a_t & \text{con probabilità } \varepsilon \end{cases}. \quad (52)$$

⁵⁴ L'*equity line* esprime l'ammontare del capitale iniziale investito all'istante di valutazione t .

Il parametro ε decide la frequenza dell'azione esplorativa a_t e viene selezionato secondo una distribuzione uniforme, mentre con una probabilità pari ad $1 - \varepsilon$ si selezionerà l'azione secondo il criterio *greedy* che massimizzerà la funzione di valore $Q(s_t, a_t, \theta_t)$. Tale processo porterà sicuramente ad alcune performance negative, poiché nel ε per cento dei casi verranno selezionate azioni su base randomica al fine di esplorare l'ambiente circostante e testare possibili miglioramenti di performance, sempre seguendo il pluricitato *trade off* tra *exploration* ed *exploitation*. Verranno testati tre diversi settaggi di ε , ovvero $\varepsilon \in \{5\%, 15\%, 25\%\}$, incrementando dunque l'incidenza dell'azione esplorativa ai fini di verificare se vengono riportate miglione di performance apprezzabili.

A fini computazionali, per ottenere la funzione di valore $Q(s_t, a_t, \theta_t)$ in base alla scelta di a_t non si utilizzano i valori $(-1, 0, 1)$ del sistema, ma la seguente matrice:

$$\mathbf{a} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (53)$$

nella quale la riga 1 corrisponde all'azione $a_t = -1$, la riga 2 corrisponde all'azione $a_t = 0$ e la riga 3 corrisponde ad $a_t = 1$. La ragione per la quale si introduce nel codice sviluppato in *Matlab*® una struttura di tal genere è molto semplice: differenziare le azioni possibili, ma assegnare loro il medesimo valore. Nel processo di ottimizzazione $\operatorname{argmax}_{a_t} Q(s_t, a_t, \theta_t)$ sia i pesi del vettore θ_t che il vettore s_t sono identici per ogni possibile azione a_t . Perciò, in questo caso, vengono stimate tre diverse possibili funzioni di valore per ogni possibile azione $a_t = -1, 0, 1$; essendo però richiesta la massimizzazione della funzione di valore $Q(s_t, a_t, \theta_t)$, si otterrebbe un sistema penalizzante per l'azione con valore $a_t = -1$ a discapito di un favore per l'azione $a_t = 1$ per una semplice ragione numerica: ricercando il massimo della funzione Q , questo sarà inteso in corrispondenza del massimo valore di a_t , ovvero $a_t = 1$. In questo modo si otterrebbe un sistema di trading sbilanciato verso una posizione *long* o di acquisto, con possibili ripercussioni anche in termini di performance. Richiamando l'equazione (43), è chiaro che la decisione dell'azione a_t comporterà solamente una modifica

dell'ultimo stato n della sommatoria e dunque è possibile riscrivere l'equazione nel seguente modo:

$$Q(s, a, \theta) = \theta' \Phi(s) = \sum_{i=1}^{n-1} \theta_i(i) \phi_i(s) + \theta(n) \sum_{j=1}^3 \Phi_n(\mathbf{a}_{j,:}) \quad (54)$$

dove con $\theta(n) \sum_{j=1}^3 \Phi_n(\mathbf{a}_{tj,:})$ si intende la moltiplicazione tra il vettore $\theta(n)$ e rispettivamente ogni riga j -esima della matrice \mathbf{a} . Si otterranno, in questo caso, tre diverse funzioni $Q(s, a, \theta)$ cui valore finale dipenderà solamente dal parametro θ_j del vettore dei pesi associato al valore non negativo della riga j -esima della matrice \mathbf{a} . Tale tecnica risulta essere efficiente poiché il valore della funzione $Q(s, a, \theta)$, ad ogni step t , dipenderà solamente dalla variabile θ , snellendo così il processo di computazione.

4.1.5 Funzione di *squashing*

La funzione di *squashing* viene utilizzata per trasformare gli input s derivanti dal vettore che descrive lo stato al fine di ottenere dei valori per accrescere la sensibilità del processo di valutazione della funzione $Q(s, a, \theta)$. Tale procedimento è simile ad un processo di *clusterizzazione* che raggruppa in *clusters* valori molto vicini di realizzazioni di variabili continue (Corazza e Sangalli, 2015). La scelta effettuata in questo lavoro è quella di utilizzare la funzione di *squashing* vista in Corazza e Sangalli (2015), ovvero la seguente funzione logistica:

$$\Phi(x) = \frac{a}{1 + be^{-cx}} + d, \quad (55)$$

sostituendo ai parametri della funzione i valori $a = 2$, $b = 1$, $c = 10^{15}$ e $d = -1$. La scelta di un valore di c così elevato rende la funzione $\Phi(x)$ graficamente comparabile ad un "gradino", per cui valori molto vicini tra loro vengono *clusterizzati* e distanziati in seguito a tale trasformazione. La *figura 6* (alla pagina seguente) riporta diverse funzioni logistiche con valori di c crescenti, dalla linea blu (ottenuta con $c = 1$) fino alla linea

rossa a scalino (ottenuta con $c = 10^{15}$). Come si evince dalla figura, questo processo amplifica la differenziazione tra gli elementi contenuti nel vettore s e dunque incrementa la sensibilità dell'algoritmo. Il motivo di tale scelta è insito nella caratteristica più comune e facilmente intuibile dei rendimenti di mercato: essi presentano limitati valori estremi e una elevata concentrazione intorno allo 0, per cui una elevata differenziazione come quella ottenuta tramite questo metodo permette una sensibilità maggiore dell'algoritmo anche in presenza di rendimenti con valori molto vicini tra loro.

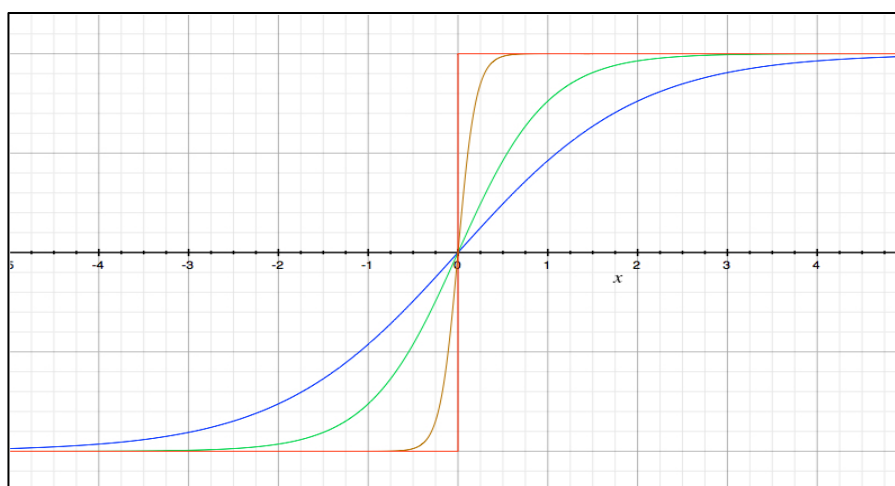


Figura 6: rappresentazione di funzioni logistiche settando $c=1$ (linea blu), $c=2$ (linea verde), $c=10$ (linea arancio) e $c=10^{15}$ (linea rossa).

4.2 Serie storiche dei prezzi

Sono stati selezionati cinque titoli quotati sul mercato italiano, ovvero UniCredit S.p.A. (codice UCG.MI), Assicurazioni Generali (codice G.MI), Eni S.p.A. (codice ENI.MI), Fiat Chrysler Automobiles NV (codice FCA.MI) e Telecom Italia S.p.A. (codice TIT.MI). I dati storici sono stati scaricati da *Bloomberg*® e ricoprono il periodo dal 1° gennaio 2000 al 31 dicembre 2017, ovvero 4569 prezzi giornalieri di chiusura, corrispondenti a circa 18 anni di quotazioni di mercato. La scelta dei titoli è stata effettuata in modo da considerare i maggiori rappresentanti dei principali settori dell'economia italiana, ovvero il settore bancario (UniCredit) ed assicurativo (Assicurazioni Generali), il settore energetico (Eni), il settore *automotive* (Fiat Chrysler) ed il settore delle telecomunicazioni (Telecom Italia). Di seguito vengono riportati i grafici rappresentanti l'andamento dei prezzi e dei rendimenti logaritmici giornalieri, nonché le principali statistiche descrittive, di ogni serie storica.

UniCredit S.p.A.

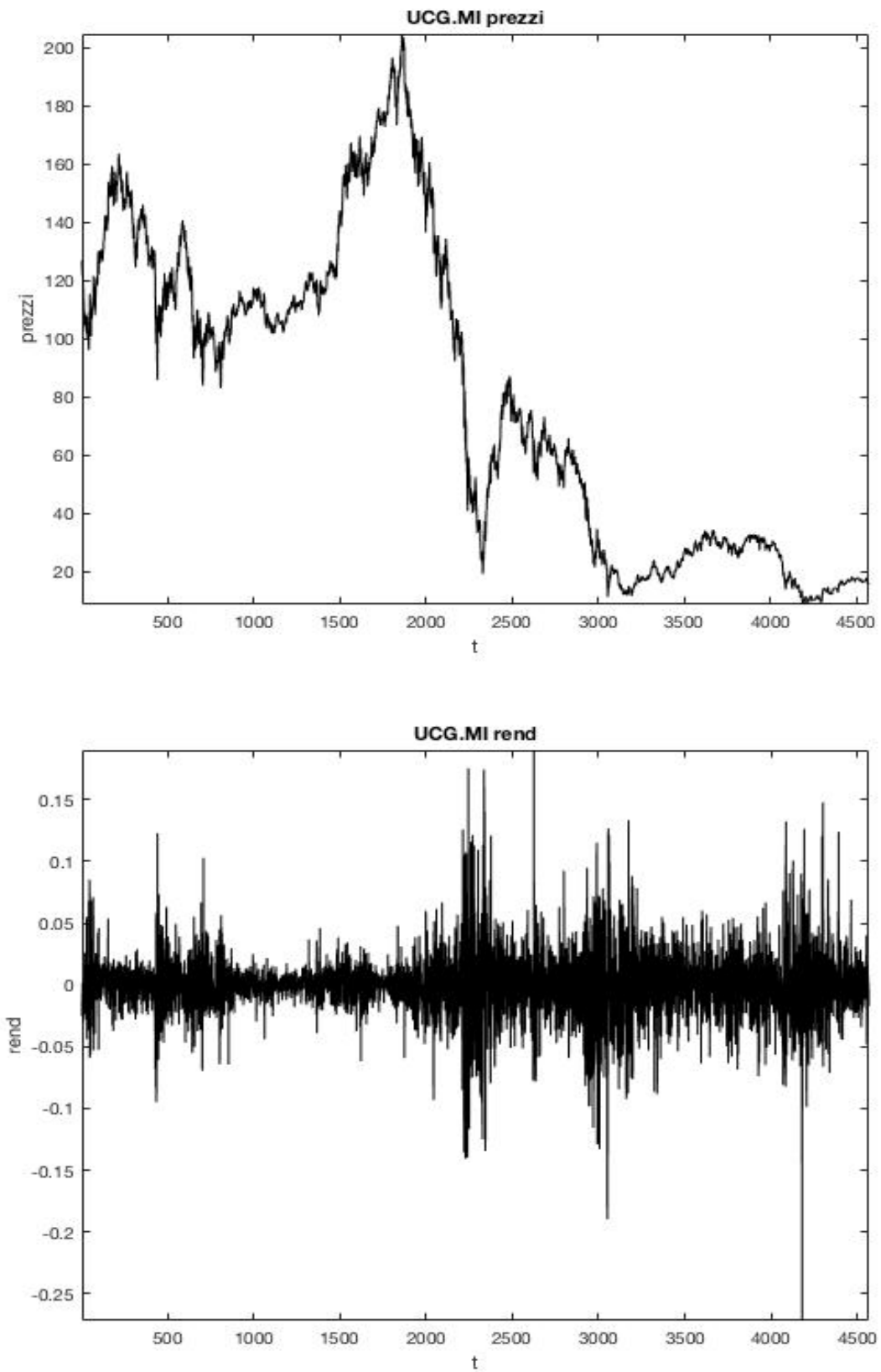


Figura 7: grafici dei prezzi e rendimenti giornalieri di Unicredit S.p.A. (elaborazione Matlab)

Assicurazioni Generali

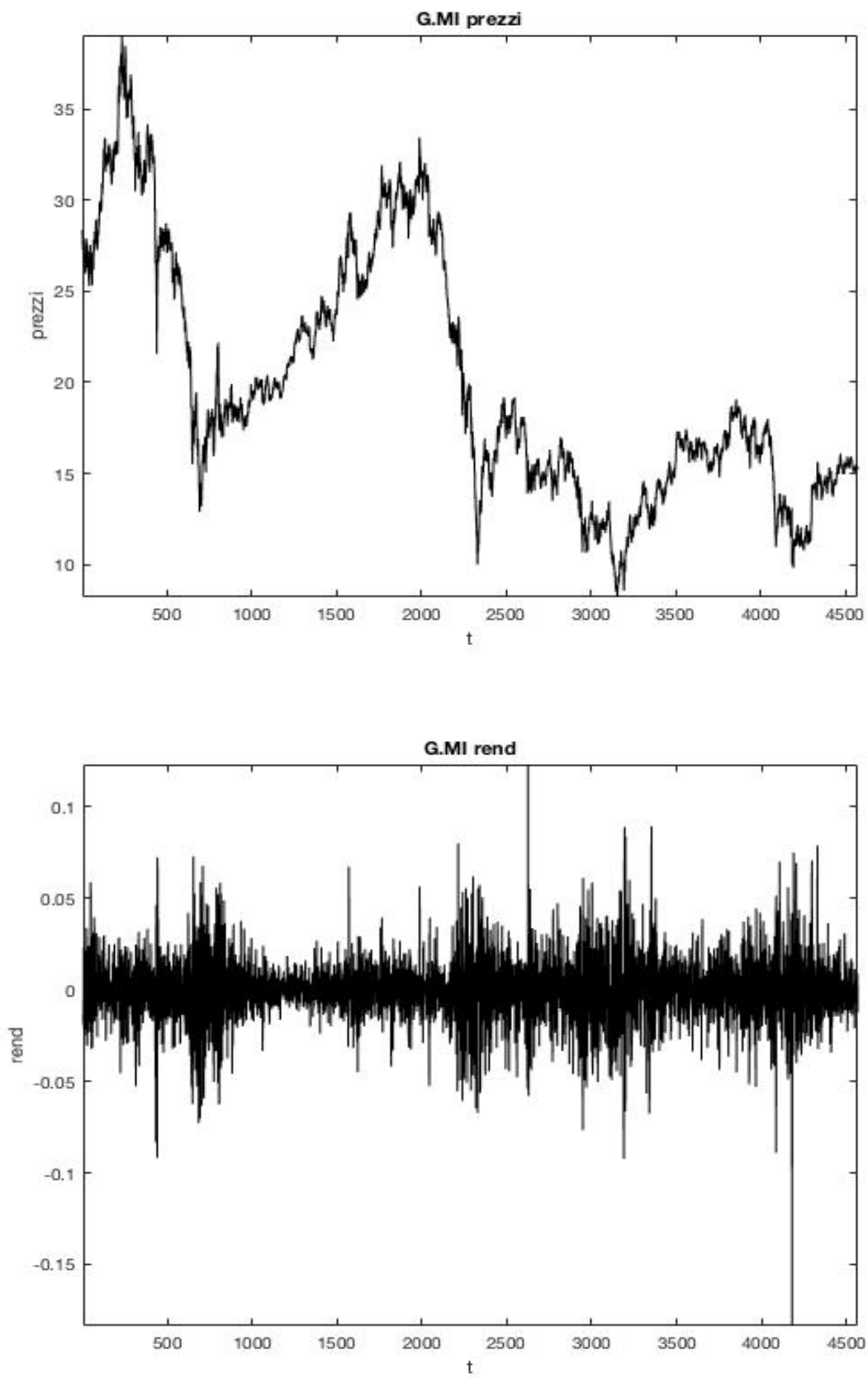


Figura 8: grafici dei prezzi e rendimenti giornalieri di Assicurazioni Generali (elaborazione Matlab)

ENI S.p.A.

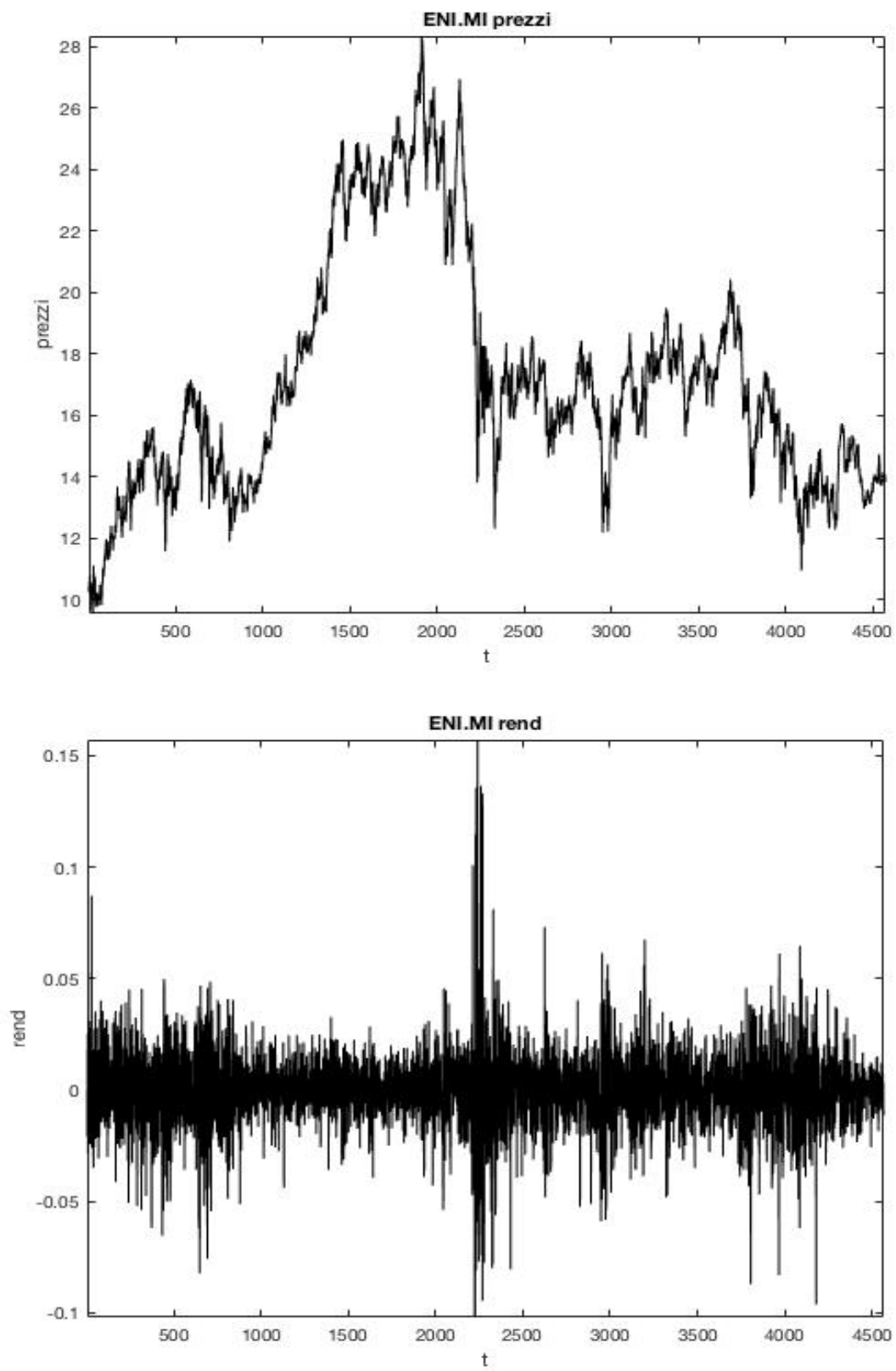


Figura 9: grafici dei prezzi e rendimenti giornalieri di Eni S.p.A. (elaborazione Matlab)

Fiat Chrysler Automotive S.p.A.

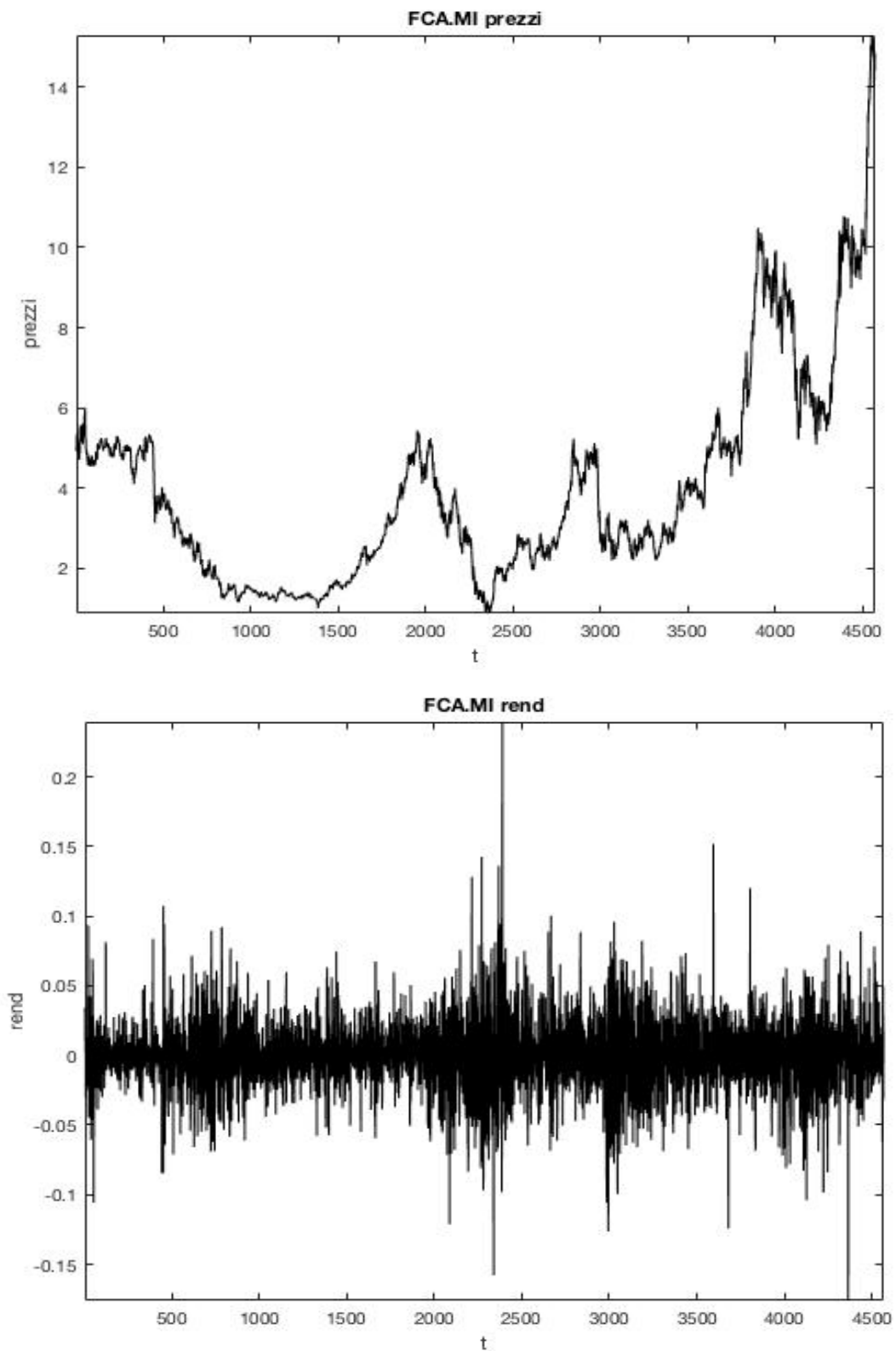


Figura 10: grafici dei prezzi e rendimenti giornalieri di Fiat Chrysler Automotive S.p.A. (elaborazione Matlab)

Telecom Italia S.p.A.

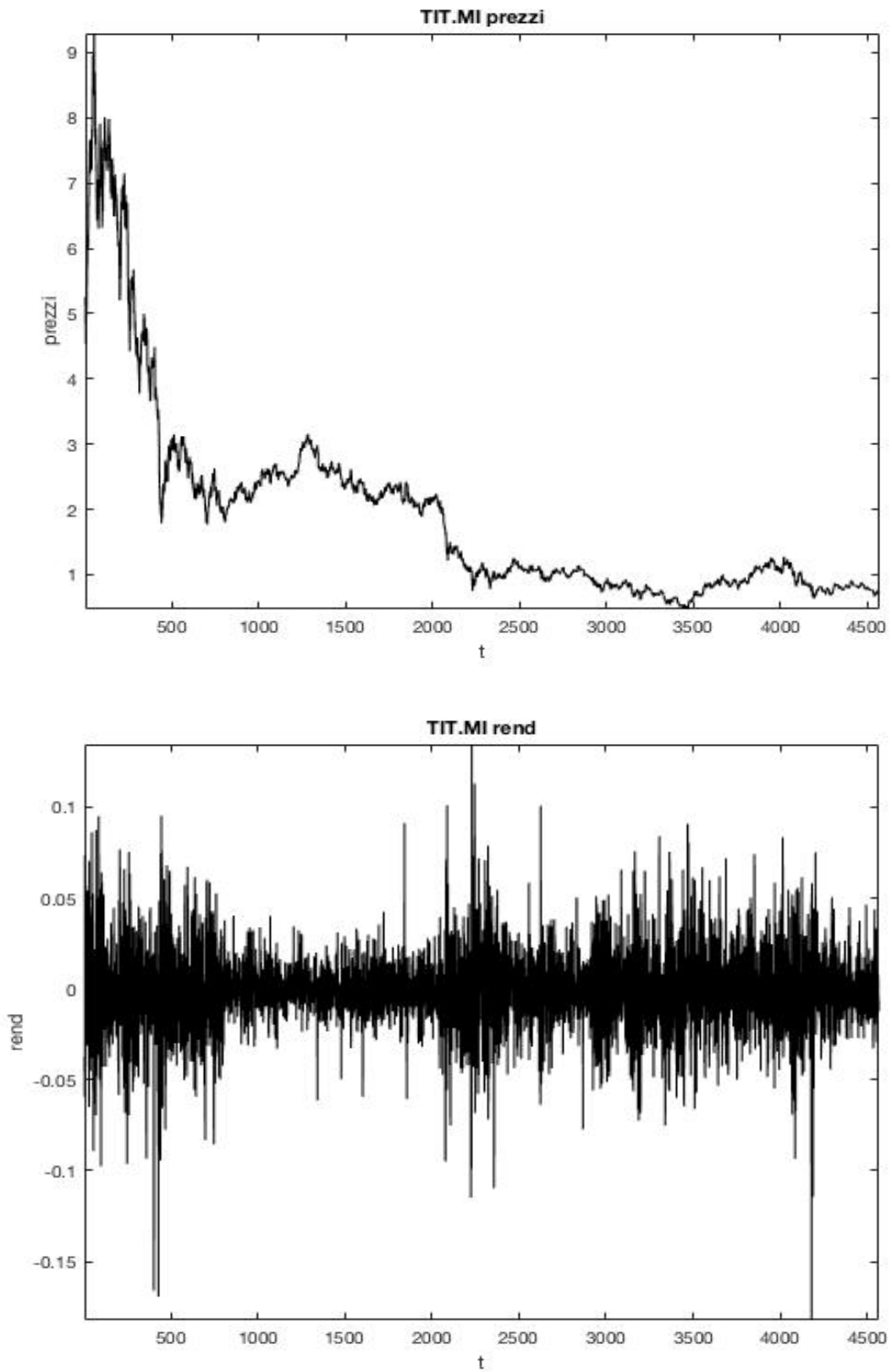


Figura 11: grafici dei prezzi e rendimenti giornalieri di Telecom Italia S.p.A. (elaborazione Matlab)

Statistiche descrittive

La *tabella 2* riporta le principali statistiche dei rendimenti dei titoli scelti, tutti i dati sono rappresentati su base giornaliera. Come si può appurare, le distribuzioni dei rendimenti si conformano ad essere leptocurtiche e leggermente asimmetriche, distanziandosi dall'ipotesi di normalità.

	UCG.MI	G.MI	ENI.MI	FCAMI	TIT.MI
rend medio	-0.0001994393	-0.0000593382	0.0000265459	0.0001042311	-0.0001889005
varianza	0.0001401218	0.0000584607	0.0000546023	0.0001263988	0.0000965871
dev.st	0.0118373069	0.0076459618	0.0073893401	0.0112427214	0.0098278739
max	0.0825454973	0.0534739393	0.0682338769	0.1042112537	0.0581728653
min	-0.1179796253	-0.0797111967	-0.0443559329	-0.0764519347	-0.0789189117
asimmetria	-0.1477264092	-0.2237035005	0.1324105407	0.1210993749	-0.2439253604
curtosi	7.8939125966	5.6136781262	7.5163422239	4.4356408603	4.4483660387

Tabella 2: statistiche descrittive dei titoli selezionati (elaborazione excel).

4.3 Fase di simulazione, fase operativa e rappresentazione dei risultati

La fase di simulazione rappresenta il primo step computazionale dell'algoritmo Q-Learning. Al fine di ottenere una soddisfacente approssimazione dei valori ottimi, e dal momento che il vettore θ , ad ogni iterazione, viene inizializzato sulla base di una distribuzione di probabilità uniforme compresa nell'intervallo $[-1,1]$, vengono computate 500 iterazioni per ogni settaggio diverso di parametri. In ogni differente settaggio, viene selezionato un *learning rate* $\alpha = 5\%$ ed un tasso di sconto $\gamma = 95\%$, entrambi valori ampiamente utilizzati in letteratura empirica ed in questo contesto riproposti. In particolare, seguendo quanto analizzato nel capitolo 3, il *learning rate* è sufficientemente piccolo per garantire la convergenza al valore ottimo $Q^*(s, a)$ ma non così tanto da rendere tale convergenza troppo lenta; il valore di γ è settato per tenere in maggior considerazione i rendimenti futuri rispetto ai rendimenti presenti, configurando l'algoritmo in maniera da assumere un comportamento più lungimirante. Riassumendo, vengono predisposti i seguenti settaggi: due diversi vettori per descrivere lo stato s_t ($N = 1, N = 5$), tre diverse funzioni di *reward* (Sharpe, Sortino, Calmar) considerando due diversi valori di L (5 e 22), una funzione logistica come funzione di

squashing, costi percentuali di transazione pari a 0,10% (applicati in fase di simulazione ed in fase operativa) e tre diversi valori di ε (5%,15%, 25%), per un totale di configurazioni pari a 36. I risultati sono raggruppati per numerosità dei titoli ($N = 1, N = 5$) e per funzione di reward impiegata, in modo tale da indagare se le due alternative allo Sharpe Ratio siano valide e forniscano indicazioni migliori all’algoritmo sulla ricompensa in relazione alle azioni di volta in volta intraprese.

In linea generale, l’algoritmo, in fase di simulazione, opera nel seguente modo: viene inizializzato il vettore θ ed il vettore s_t (sulla base dei parametri impostati). In seguito, ad ogni k -esima iterazione e per ogni titolo, viene selezionata l’azione *greedy* o l’azione *esplorativa*: impostando ad ogni t -esimo stato un’estrazione casuale (*rand*) da un intervallo $[0,1]$, se $rand < \varepsilon$ viene massimizzato il valore di $Q(s, a, \theta)$ estraendo casualmente un’azione a_t dalla matrice (53), contrariamente viene selezionata l’azione a_t secondo il criterio *greedy*, ovvero viene computata la funzione di valore $Q(s, a, \theta)$ per ognuna delle tre diverse azioni possibili ($a_t = -1, 0, 1$) ed i tre risultati vengono rappresentati in un vettore, selezionando poi il valore massimo tra i tre ottenuti. Nel caso in cui due o più risultati siano identici, è indifferente la scelta dell’azione da intraprendere poiché questa comunque massimizza la funzione di valore: ad esempio, $a_t = -1$ ed $a_t = 0$ potrebbero fornire gli stessi valori della funzione $Q(s, a, \theta)$ ed essere entrambi dei massimi. Dal momento che il vettore rappresentante i tre risultati delle funzioni di valori è il seguente:

$$\max Q L_t = \max [Q(s_t, a_{t,-1}, \theta_t), Q(s_t, a_{t,0}, \theta_t), Q(s_t, a_{t,1}, \theta_t)], \quad (56)$$

nel caso di due o più valori uguali questi verrebbero selezionati seguendo il sistema di preferenze $a_{t,-1} > a_{t,0} > a_{t,1}$, penalizzando sistematicamente le azioni $a_t = 0$ ed $a_t = 1$ rispetto all’azione $a_t = -1$, e l’azione $a_t = 1$ rispetto all’azione $a_t = 0$. Per ovviare a questo problema, in *Matlab*® viene computata la seguente equazione:

$$[\max Q L_t, \text{randi}(3)] = \max [Q(s_t, a_{t,-1}, \theta_t), Q(s_t, a_{t,0}, \theta_t), Q(s_t, a_{t,1}, \theta_t)], \quad (57)$$

dove $\text{randi}(3)$ indica un comando *Matlab*® che estrae casualmente un numero intero compreso tra 1 e 3; in questo modo, qualora vi fossero due o tre valori identici, la selezione dell'azione *greedy* verrà effettuata in maniera casuale, senza penalizzarne in questo modo la scelta. L'azione così selezionata determinerà il profitto del periodo successivo ($g_{t+1} = e_{t+1}a_t$), sulla base del rendimento giornaliero realizzato, e di conseguenza il valore della *equity line* (equazione 50). Il passaggio successivo consiste nel valutare la selezione tramite la funzione di reward, computata considerando gli ultimi L profitti percentuali ottenuti (g_t). Infine, si ottiene l'aggiornamento del vettore θ_{t+1} utilizzando l'equazione (41) e predisponendo, dunque, tutte le componenti necessarie al suo calcolo: $s_{t+1}, r_{t+1} + \gamma \max_a Q(s_{t+1}, a, \theta_t) - Q(s_t, a_t, \theta_t), \nabla_{\theta_t} Q(s_t, a_t, \theta_t)$.

Il numero di iterazioni, per le ragioni già precedentemente esposte, è elevato ($k = 500$), dunque per ogni step sono computate un numero pari a k azioni. Al fine di rendere l'algoritmo operativo è necessario che esso fornisca una sola indicazione di *trading* per ogni t -esimo periodo, dunque serve introdurre una regola per aggregare le k azioni in modo tale da ottenere un solo ed unico segnale operativo. Viene seguita dunque l'impostazione presente in Corazza e Sangalli (2015), determinando innanzitutto la media delle azioni per ogni periodo t -esimo come segue:

$$\bar{a}_t = \frac{\sum_{k=1}^K a_{t,k}}{K}, \quad (58)$$

dove \bar{a}_t è la media delle k azioni al periodo t -esimo, K è il numero totale di iterazioni impostate (500) ed $a_{t,k}$ è l'azione del periodo t -esimo alla k -esima iterazione. In seguito, ogni azione media \bar{a}_t viene tradotta in un segnale operativo utilizzando la seguente regola:

- se $-1 \leq \bar{a}_t < -1/3$ allora $a_t = -1$, ovvero "vendita o posizione *short* nel mercato".
- se $-1/3 \leq \bar{a}_t \leq 1/3$ allora $a_t = 0$, ovvero "star fuori dal mercato".
- se $1/3 < \bar{a}_t \leq 1$ allora $a_t = 1$, ovvero "acquisto o posizione *long* sul mercato".

In questo modo, il trend generale di tutte le *equity line* ottenute dalle k simulazioni viene raggiunto in media (Corazza e Sangalli, 2015). Sulla base dei segnali operativi ottenuti applicando il sistema appena esposto, vengono computate le *equity line* definitive, considerando il profitto al lordo (eq_lineG) ed al netto (eq_lineN) dei costi di transazione, secondo le seguenti equazioni:

$$eq_lineG_{t+1} = eq_lineG_t * (1 + g_{t+1}) \quad (59)$$

$$eq_lineN_{t+1} = eq_lineG_t * (1 + g_{t+1} - \delta|a_t - a_{t-1}|). \quad (60)$$

La quarta ed ultima fase dell'algorithmo riguarda, come già esposto, la valutazione delle *performance* del sistema di *trading*, applicato ai vari titoli, sulla base di determinate statistiche descrittive. In particolare, vengono inizialmente estratti, da ogni *equity line*, il capitale finale al lordo ed al netto dei costi di transazione. Sulla base dei valori estratti, si ottengono i rendimenti medi giornalieri (lordi e netti) nel seguente modo:

$$r_daily = \left(\frac{eq_lineT}{C} \right)^{\frac{1}{T}} - 1 \quad (61)$$

dove C è il capitale iniziale investito e T è l'ammontare totale dei giorni di trading considerati nel sistema ($T = 4569$). I rendimenti giornalieri così ottenuti vengono utilizzati per calcolare il rendimento medio annuo (lordo e netto):

$$r_annual = (1 + r_daily)^{252} - 1 \quad (63)$$

con 252 a rappresentare, notoriamente, il numero di giorni di trading all'anno. Viene inoltre calcolata la numerosità, espressa in percentuale, di volte nelle quali l'*equity line* (lorda e netta) è superiore al livello del capitale iniziale C , in maniera tale da valutare la performance e l'efficacia del sistema di *trading* non soltanto sulla base del profitto finale, ma anche sulla base dell'andamento del sistema nel corso di tutto il periodo considerato. Tali percentuali sono così ottenute:

$$perc_overG = \frac{\eta_G}{T} \times 100 \quad (65)$$

dove η_G ed η_N sono, rispettivamente, il numero di volte nelle quali $eq_lineG_t \geq C$ ed $eq_lineN_T \geq C$. Infine, viene calcolato il numero medio di azioni per anno, per indagare a quanto ammonta, mediamente, l'operatività indotta dal sistema di trading implementato:

$$\bar{a}_annual = \frac{a_{N,j}}{T/252} \times 100, \quad (66)$$

considerando $a_{N,j}$ come il totale di azioni di trading per il titolo j -esimo. Vengono infine costruiti due grafici: il primo raccoglie tutte le *equity line* ottenute dalle k iterazioni, il secondo invece viene suddiviso in tre pannelli rappresentanti rispettivamente i prezzi della serie storica rappresentata, le azioni intraprese dal sistema di *trading* ad ogni istante di tempo e l'andamento dell'*equity line* al lordo ed al netto dei costi di transazione confrontate con il livello di capitale iniziale investito C .

5. APPLICAZIONE PRATICA DEL QL_a: RISULTATI

In questo capitolo vengono presentati i risultati ottenuti settando le varie configurazioni presentate nel capitolo precedente. Tali risultati vengono esposti presentandoli per titolo, per numerosità del vettore dei rendimenti ($N = 1$ ed $N = 5$) e per funzione di reward impiegata, aggregando i dati e suddividendoli per configurazioni. Data la numerosità di settaggi (36), vengono presentati solo alcuni dei grafici ottenuti, adatti a comprendere sia il processo di computazione che la valutazione delle performance. Nella tabella sottostante è possibile apprezzare il tempo di calcolo impiegato dal software *Matlab*⁵⁵ per il computo di ogni settaggio e funzione di *reward* impostata⁵⁵.

rwd	L	ε	N=1		N=5	
			sec	min	sec	min
SR	5	5%	622.16	10:22	752.89	12:33
		15%	647.52	10:47	627.69	10:28
		25%	630.89	10:05	651.61	10:52
	22	5%	647.55	10:48	654.86	10:55
		15%	633.79	10:34	667.30	11:07
		25%	621.68	10:22	668.92	11:09
SOR	5	5%	985.49	16:26	1294.77	21:35
		15%	1001.74	16:42	1258.22	20:58
		25%	997.37	16:37	1022.07	17:02
	22	5%	2702.96	45:02	2766.63	44:07
		15%	2684.26	44:44	2902.20	44:22
		25%	2666.75	44:27	2641.43	44:02
CAR	5	5%	999.01	16:39	995.21	16:35
		15%	1023.42	17:03	976.21	16:16
		25%	968.70	16:08	996.84	16:36
	22	5%	1463.50	24:24	1493.34	24:53
		15%	2236.27	37:16	1391.16	23:11
		25%	1475.22	24:35	1459.70	24:20

Tabella 3: tempo di computazione di tutti i settaggi in ambiente Matlab.

Per ogni titolo e per ogni valore di N , vengono presentate le seguenti statistiche, al lordo (G) ed al netto (L) dei costi di transazione: valore dell'*equity line* finale, rendimento

⁵⁵ Il computer utilizzato per le computazioni è un MacBook Air 13', processore da 1.4 GHz (i-core i5), memoria RAM 4 GB 1600 MHz DDR3.

medio giornaliero ed annuo, percentuale di volte in cui l'*equity line* è maggiore del capitale iniziale investito ed operatività media annua. I valori ottenuti vengono confrontati in base alla funzione di *reward* impiegata, per verificare se una scelta diversa dallo Sharpe Ratio possa portare dei benefici al sistema di *trading* utilizzato. Di seguito si riportano, a titolo esemplificativo, una serie di grafici ottenuti in alcune simulazioni.

La *figura 12* rappresenta la totalità delle *equity line* ottenute in fase di computazione, corrispondente al numero di iterazioni impiegate ($k = 500$), del titolo UCG.MI settando $N = 1$, SR, $L = 22$, $\varepsilon = 25\%$, mentre la *figura 13* riporta le *equity line* del titolo TIT.MI ottenute settando $N = 5$, SOR, $L = 5$, $\varepsilon = 5\%$. In base a tali figure, il titolo UCG.MI risulta relativamente semplice da "apprendere" per l'algoritmo, poiché la maggior parte delle *equity line* in fase di computazione segue un andamento simile ed i segnali operativi sono dunque ben definiti. Il titolo TIT.MI, al contrario, presenta una elevata dispersione delle *equity line*; l'algoritmo così non è in grado di cogliere i segnali operativi in maniera efficace e questo inficia la performance della fase operativa. Le *figure 14 e 15* riproducono invece le performance operative dei titoli in base al settaggio utilizzato: il primo panel di ogni figura riporta l'andamento dei prezzi del titolo, il secondo panel rappresenta graficamente l'andamento della posizione assunta sul mercato (*long/buy, out, short/sell*), il terzo panel infine rappresenta l'*equity line* lorda (linea blu) e netta (linea verde) rispetto al capitale iniziale di partenza (*starting capital*, linea rossa). Nonostante il valore finale dell'*equity line* netta sia prossimo al valore del capitale iniziale per entrambi i titoli, il livello è costantemente maggiore del capitale investito, presentando anche notevoli picchi positivi. In questo caso, come in tutti i casi simili analizzati, si propende a considerare positivo il risultato dell'algoritmo ed efficace il processo di selezione delle azioni di *trading*. Le statistiche che indicano la percentuale di volte in cui l'*equity line* (lorda e netta) è maggiore del capitale iniziale investito sono, in questo senso, utili a valutare la bontà o meno delle performance di volta in volta analizzate. In particolare, si considerano positivi quei risultati che presentano una percentuale pari o superiore al 50%, in particolar modo valutando l'*equity line* netta, che determina l'effettivo profitto realizzato. Infine, si considera sufficientemente operativo un sistema di trading e di settaggi qualora presenti almeno una media di $\# = 4$ azioni annue, corrispondente ad un'azione ogni trimestre.

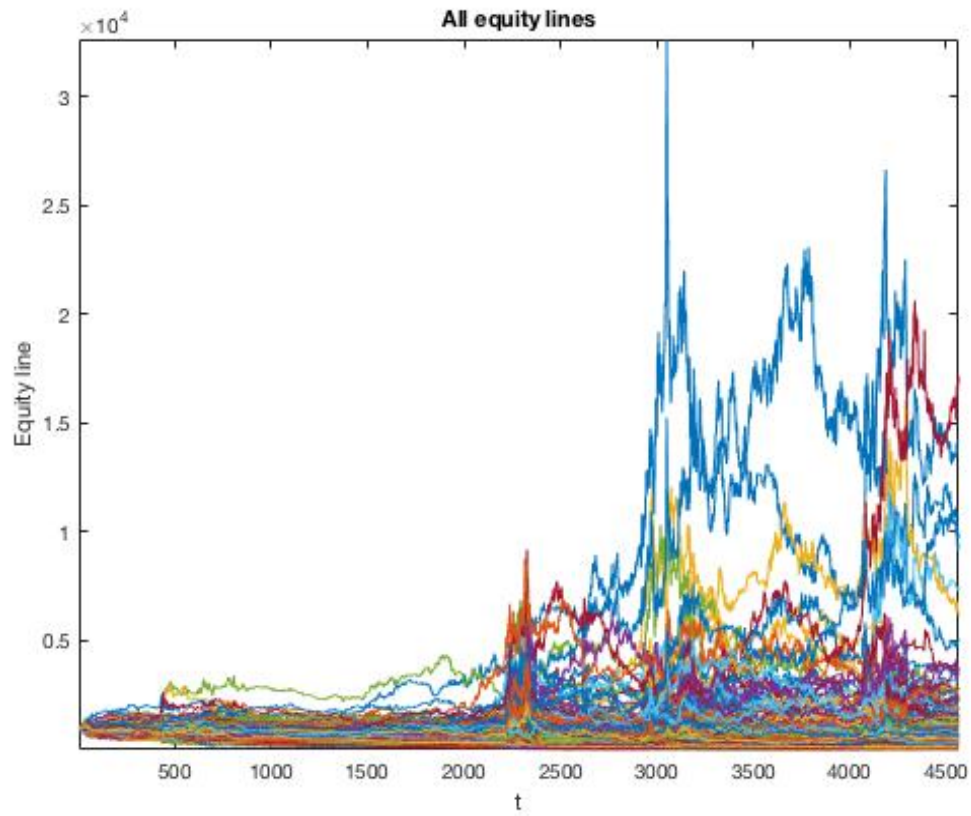


Figura 12: equity line totali, titolo UCG.MI ($N=1$, SR , $L=22$, $\varepsilon=0,25$).

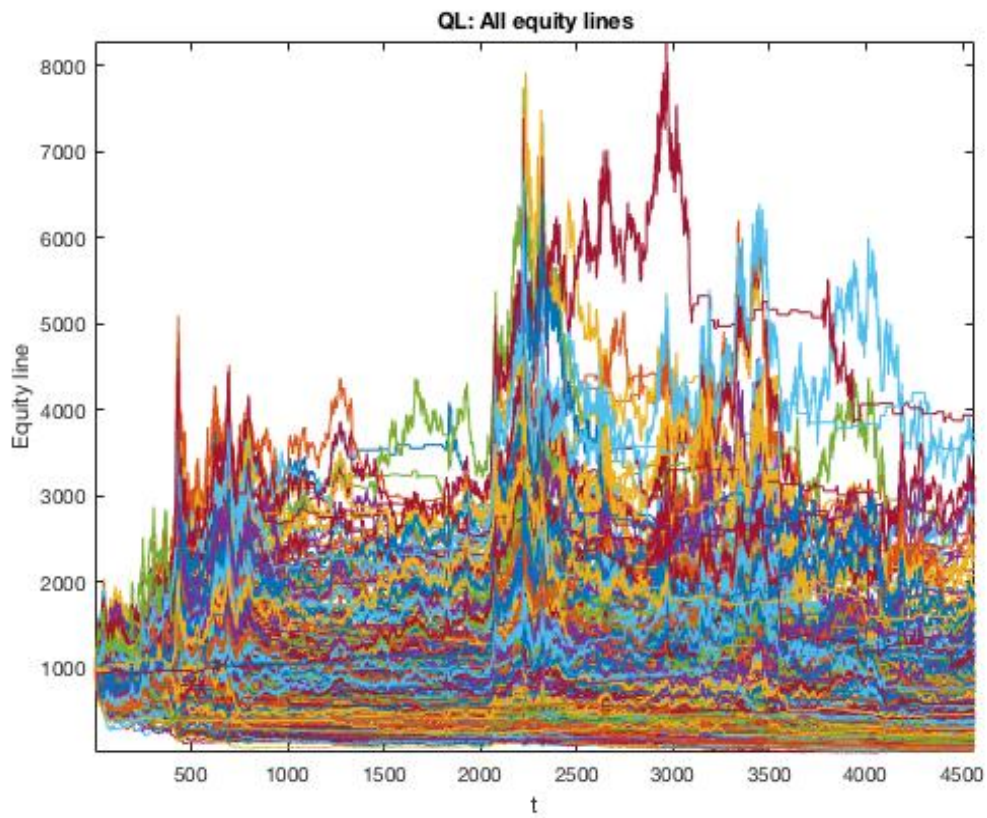


Figura 13: equity line totali, titolo TIT.MI ($N=5$, SOR , $L=5$, $\varepsilon=0,05$).

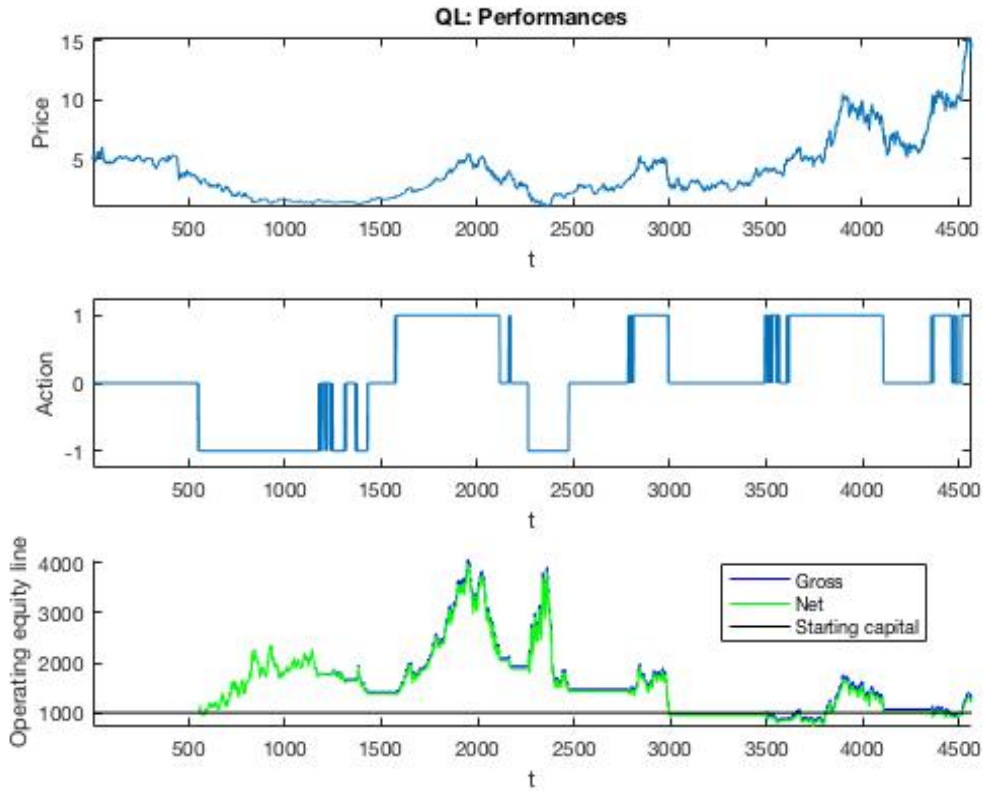


Figura 15: performance titolo FCA.MI ($N=1$, SOR , $L=22$, $\varepsilon=0,05$).

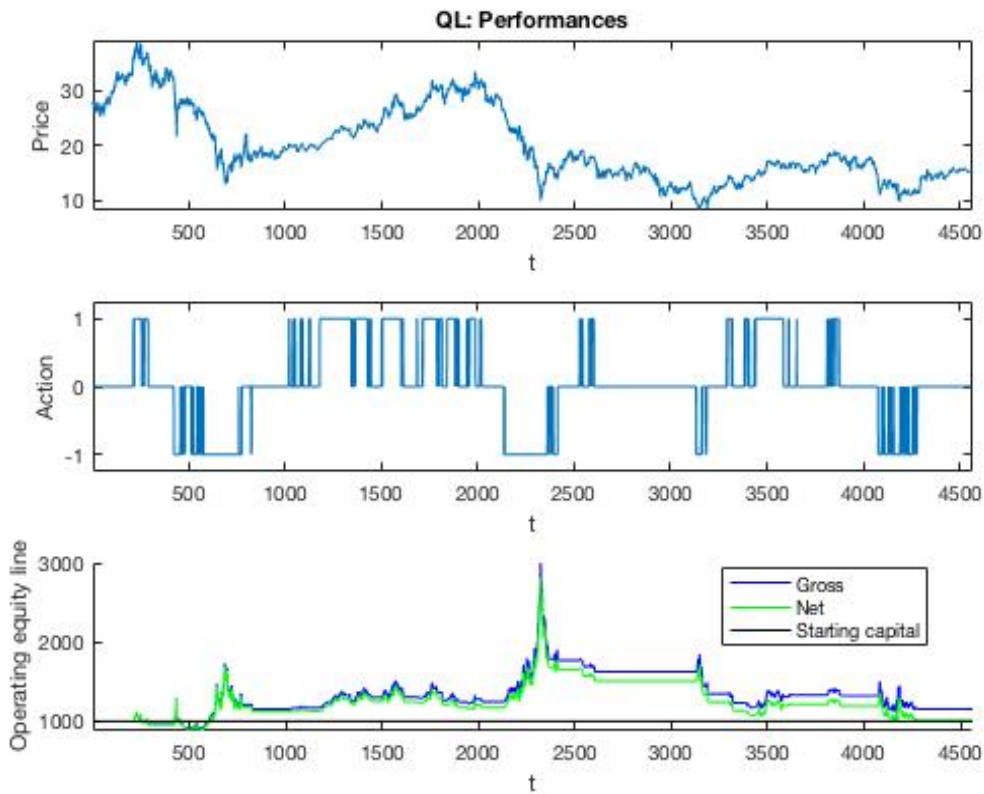


Figura 14: performance titolo G.MI ($N=5$, SR , $L=5$, $\varepsilon=0,15$).

5.1 Risultati: singoli titoli, settaggi $N = 1$, $N = 5$

UniCredit S.p.A.

N=1			eq_line		rend giorn		rend annuo		perc_over		#
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	
SR	5	5%	290.53	263.55	-0.027%	-0.029%	-6.59%	-7.09%	6.94%	6.79%	10.76
		15%	935.57	858.29	-0.001%	-0.003%	-0.37%	-0.84%	31.97%	26.32%	9.49
		25%	2505.43	2242.46	0.020%	0.018%	5.20%	4.56%	46.62%	44.41%	12.25
	22	5%	632.46	596.19	-0.010%	-0.011%	-2.50%	-2.81%	45.87%	35.21%	6.51
		15%	2750.31	2544.20	0.022%	0.020%	5.74%	5.29%	64.90%	63.76%	8.61
		25%	4256.98	3856.35	0.032%	0.030%	8.32%	7.73%	74.51%	72.37%	10.93
SOR	5	5%	378.19	343.05	-0.021%	-0.023%	-5.22%	-5.73%	6.98%	6.88%	10.76
		15%	1504.69	1367.09	0.009%	0.007%	2.28%	1.74%	45.26%	42.48%	10.59
		25%	1336.30	1151.17	0.006%	0.003%	1.61%	0.78%	41.80%	35.98%	16.44
	22	5%	800.27	740.19	-0.005%	-0.007%	-1.22%	-1.65%	30.94%	20.01%	8.61
		15%	2923.55	2690.48	0.023%	0.022%	6.10%	5.61%	61.02%	51.39%	9.16
		25%	3054.51	2629.54	0.024%	0.021%	6.36%	5.48%	49.70%	48.85%	16.55
CAR	5	5%	776.10	726.59	-0.006%	-0.007%	-1.39%	-1.75%	7.84%	6.94%	7.28
		15%	961.33	941.34	-0.001%	-0.001%	-0.22%	-0.33%	30.50%	30.02%	2.32
		25%	1027.34	1024.28	0.001%	0.001%	0.15%	0.13%	100.00%	100.00%	0.33
	22	5%	1249.08	1147.46	0.005%	0.003%	1.23%	0.76%	95.95%	85.02%	9.38
		15%	810.66	738.74	-0.005%	-0.007%	-1.15%	-1.66%	42.81%	41.32%	10.26
		25%	521.04	508.03	-0.014%	-0.015%	-3.53%	-3.67%	49.00%	49.00%	2.76

Tabella 4: statistiche riepilogative UCG.MI, N=1.

N=5			eq_line		rend giorn		rend annuo		perc_over		#
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	
SR	5	5%	391.69	353.76	-0.021%	-0.023%	-5.04%	-5.58%	7.50%	6.90%	11.27
		15%	726.73	662.78	-0.007%	-0.009%	-1.75%	-2.25%	12.34%	10.98%	10.16
		25%	1528.81	1350.55	0.009%	0.007%	2.37%	1.67%	45.45%	42.82%	13.70
	22	5%	488.39	466.44	-0.016%	-0.017%	-3.88%	-4.12%	67.92%	66.54%	5.08
		15%	2019.48	1838.40	0.015%	0.013%	3.96%	3.42%	48.39%	47.97%	10.38
		25%	1793.03	1571.69	0.013%	0.010%	3.28%	2.53%	60.97%	48.54%	14.58
SOR	5	5%	371.33	331.48	-0.022%	-0.024%	-5.32%	-5.92%	2.41%	1.80%	12.54
		15%	602.10	545.75	-0.011%	-0.013%	-2.76%	-3.29%	6.73%	5.57%	10.82
		25%	1157.73	1011.05	0.003%	0.000%	0.81%	0.06%	22.27%	13.46%	14.91
	22	5%	1534.08	1473.88	0.009%	0.009%	2.39%	2.17%	83.76%	83.32%	4.42
		15%	1420.36	1296.92	0.008%	0.006%	1.96%	1.45%	59.19%	57.68%	10.05
		25%	664.48	588.86	-0.009%	-0.012%	-2.23%	-2.88%	66.45%	64.50%	13.36
CAR	5	5%	245.28	234.96	-0.031%	-0.032%	-7.47%	-7.69%	44.53%	44.44%	4.75
		15%	740.32	725.57	-0.007%	-0.007%	-1.65%	-1.76%	39.16%	29.80%	2.21
		25%	1001.14	1000.14	0.000%	0.000%	0.01%	0.00%	100.00%	100.00%	0.11
	22	5%	889.26	820.79	-0.003%	-0.004%	-0.65%	-1.08%	69.78%	59.92%	8.84
		15%	971.96	945.33	-0.001%	-0.001%	-0.16%	-0.31%	36.66%	36.64%	3.09
		25%	1257.16	1246.01	0.005%	0.005%	1.27%	1.22%	99.98%	99.98%	0.99

Tabella 5: statistiche riepilogative UCG.MI, N=5.

Le performance del titolo UCG.MI sono mediamente buone. Nella maggior parte dei casi, il capitale finale risulta maggiore della somma iniziale. Nel caso di $N = 1$, la principale tendenza individuata è un miglioramento delle performance all'aumentare di ε , percentuale indicante il numero di volte nelle quali viene scelta l'azione esplorativa rispetto all'azione *greedy*. I tre ratio (Sharpe, Sortino e Calmar), impiegati come funzione di *reward*, si comportano quasi nello stesso modo: performance in perdita per $\varepsilon = 5\%$, in crescendo utilizzando $\varepsilon = 15\%$ ed $\varepsilon = 25\%$. Le migliori performance si ottengono con il settaggio $N = 1, \text{SOR}, L = 22, \varepsilon = 15\%$ (2690.48 € di capitale finale, 5,61% di rendimento netto medio annuo effettuando in media $\# = 9.16$ operazioni di trading annue); utilizzando invece il settaggio $N = 1, \text{SR}, L = 22, \varepsilon = 25\%$ si ottiene un rendimento netto medio annuo di 7,73% (3856.35 € di capitale finale, media di $\# = 10.93$ operazioni di trading annue). Il Calmar ratio non sembra essere, contrariamente a quanto ipotizzato in fase di costruzione delle funzioni di *reward*, un'adeguata misura di performance da poter applicare ad un algoritmo di trading finanziario così implementato. Sia nel caso di $N = 1$ che nel caso di $N = 5$, evidenzia elevati problemi di *tuning* dell'algoritmo e riduce drasticamente l'operatività del sistema di *trading* all'aumentare di N . Inoltre, la tendenza individuata è quella di ridurre l'operatività ed i profitti all'aumentare di ε , penalizzando l'azione esplorativa rispetto all'azione *greedy*. Nonostante tali problematiche, si è comunque deciso di proseguire con la valutazione del Calmar Ratio come funzione di *reward*, raccogliendo i risultati di tutti i settaggi per lasciare poi spazio alle considerazioni del caso, presentate nel paragrafo 5.2. Utilizzando il Calmar ratio, la performance migliore si ottiene settando $N = 1, \text{CAR}, L = 22, \varepsilon = 5\%$ e risulta essere pari a 1147.36 € (0.76% il rendimento netto annuo per un'operatività media pari a $\# = 9.38$). Per quanto concerne gli altri due indici (Sharpe e Sortino), nel caso di $N = 5$ vi è un leggero peggioramento delle performance dovuta al fatto che, aumentando il numero di rendimenti all'interno del vettore rappresentativo dello stato iniziale, vi è maggior persistenza dei rendimenti negativi in s_t . In entrambi i settaggi di N , si apprezzano miglioramenti di performance non solo all'aumentare del valore di ε ma anche aumentando il valore di L , poiché la funzione di *reward* viene maggiormente sensibilizzata alla volatilità del periodo in quanto si considera non più una settimana ($L = 5$) ma circa un mese di rendimenti di mercato ($L = 22$). I grafici rappresentanti le tre migliori performance individuate sono riportati nella *figura 16*.

La *tabella 6* riporta le percentuali medie, suddivise per funzione di reward, di volte in cui l'*equity line* presenta un valore maggiore od uguale al capitale inizialmente investito, considerati tutti i settaggi implementati: viene dunque ottenuta la percentuale media netta e lorda sommando tutti i settaggi di N, L, ε per ogni funzione di reward, per verificare quale dei tre indici presenta performance migliori. La percentuale media totale, in questa come in tutte le altre tabelle simili ottenute per gli altri titoli, viene calcolata sia includendo i risultati del Calmar Ratio (voce *ALL*) sia escludendoli (voce *ALL-CAR*) poiché, dati i problemi di operatività registrati, queste percentuali potrebbero alterare (come di fatto avviene) le medie totali, conducendo ad errate valutazioni.

rwd	N=1			N=5		
	perc_over	MEDIE		perc_over	MEDIE	
	G%	N%	#	G%	N%	#
SR	45.14%	41.48%	9.76	40.43%	37.29%	10.86
SOR	39.29%	34.26%	12.02	40.13%	37.72%	11.02
CAR	54.35%	52.05%	5.39	65.02%	61.80%	3.33
ALL	46.26%	42.60%	9.06	48.53%	45.60%	8.40
ALL - CAR	42.21%	37.87%	10.89	40.28%	37.51%	10.94

Tabella 6: percentuale media equity line > cap. investito per funzione di reward, titolo UCG.MI.

Considerando la percentuale totale (*ALL*), questa non è del tutto soddisfacente sia per quanto riguarda $N = 1$ che per $N = 5$. Entrando nei dettagli, la percentuale al netto del Calmar Ratio (*ALL-CAR*) si attesta attorno al 40%, con valori leggermente migliori per $N = 1$. Data comunque la positività dei rendimenti ottenuti nei vari settaggi, questo porta a considerare che l'operatività sul titolo UCG.MI sia stata efficace soprattutto nell'intervallo di tempo più recente, fattore che può essere verificato anche visionando i grafici ad esempio riportati in *figura 16*. Lo Sharpe Ratio ed il Sortino Ratio captano quasi gli stessi segnali operativi: il Sortino riesce a cogliere meglio l'operatività con valori di ε non elevati mentre, prediligendo l'azione esplorativa, lo Sharpe si comporta meglio, permettendo di raggiungere performance più elevate. Infine, l'operatività media (colonna contrassegnato con # della *tabella 6*) risulta più elevata per il Sortino Ratio rispetto allo Sharpe Ratio, sia per $N = 1$ che per $N = 5$, mentre per il Calmar Ratio si evidenziano i problemi di scarsa operatività già esposti in precedenza. Dunque, il Sortino Ratio si rivela essere una buona funzione di reward al pari dello Sharpe Ratio, il quale presenta comunque delle performance leggermente migliori.

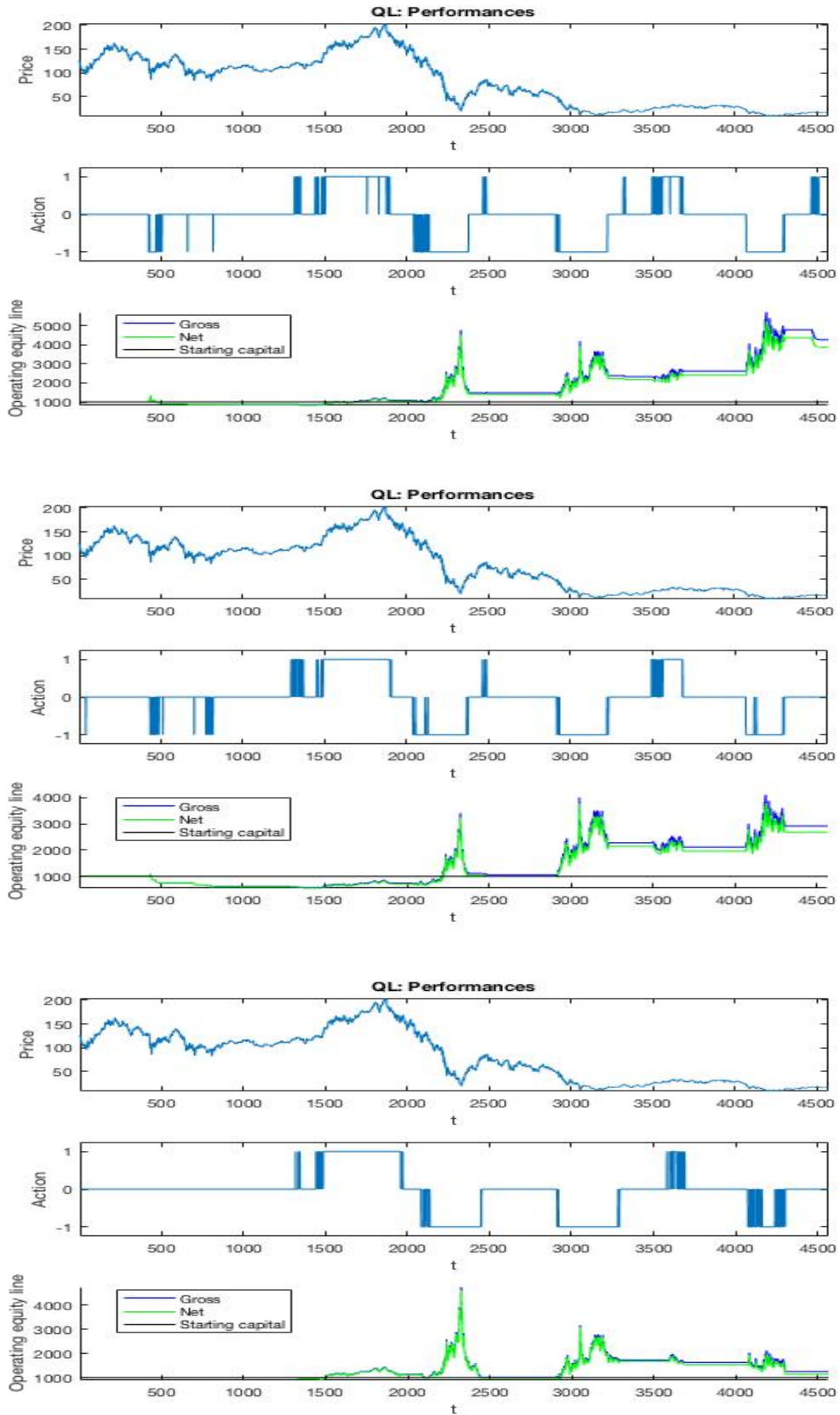


Figura 16: performance titolo UCG.MI con settaggi (rappresentati dall'alto verso il basso):

1. $N=1$, SR , $L=22$, $\varepsilon=0,25$.
2. $N=1$, SOR , $L=22$, $\varepsilon=0,15$.
3. $N=1$, CAR , $L=22$, $\varepsilon=0,05$.

Assicurazioni Generali

N=1			eq_line		rend giorn		rend annuo		perc_over		
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	1070.97	1018.72	0.002%	0.000%	0.38%	0.10%	81.80%	80.86%	5.52
		15%	766.08	691.78	-0.006%	-0.008%	-1.46%	-2.01%	85.42%	77.16%	11.26
		25%	708.93	626.91	-0.008%	-0.010%	-1.88%	-2.54%	52.11%	38.54%	13.57
	22	5%	929.35	890.27	-0.002%	-0.003%	-0.40%	-0.64%	76.51%	75.87%	4.75
		15%	1044.09	896.76	0.001%	-0.002%	0.24%	-0.60%	93.17%	84.65%	16.77
		25%	1180.75	1009.21	0.004%	0.000%	0.92%	0.05%	91.72%	83.29%	17.33
SOR	5	5%	1022.07	959.63	0.000%	-0.001%	0.12%	-0.23%	67.44%	50.73%	6.95
		15%	936.88	843.53	-0.001%	-0.004%	-0.36%	-0.93%	86.29%	81.80%	11.59
		25%	820.52	710.52	-0.004%	-0.007%	-1.09%	-1.87%	36.13%	17.25%	15.89
	22	5%	921.12	855.39	-0.002%	-0.003%	-0.45%	-0.86%	67.27%	59.62%	8.17
		15%	952.46	872.18	-0.001%	-0.003%	-0.27%	-0.75%	82.31%	64.92%	9.71
		25%	1030.55	963.67	0.001%	-0.001%	0.17%	-0.20%	77.84%	66.21%	7.39
CAR	5	5%	560.67	526.45	-0.013%	-0.014%	-3.14%	-3.48%	48.76%	40.00%	6.95
		15%	993.98	953.94	0.000%	-0.001%	-0.03%	-0.26%	78.52%	46.97%	4.52
		25%	989.78	987.79	0.000%	0.000%	-0.06%	-0.07%	27.74%	27.74%	0.22
	22	5%	785.70	710.92	-0.005%	-0.007%	-1.32%	-1.87%	60.08%	54.04%	11.04
		15%	983.36	912.35	0.000%	-0.002%	-0.09%	-0.50%	37.27%	17.65%	8.28
		25%	1066.49	1041.25	0.001%	0.001%	0.36%	0.22%	84.41%	84.30%	2.65

Tabella 7: statistiche riepilogative G.MI, N=1.

N=5			eq_line		rend giorn		rend annuo		perc_over		
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	1053.23	1012.91	0.001%	0.000%	0.29%	0.07%	79.64%	77.45%	4.31
		15%	954.65	870.87	-0.001%	-0.003%	-0.26%	-0.76%	84.26%	80.14%	10.16
		25%	1154.24	1014.65	0.003%	0.000%	0.80%	0.08%	93.25%	92.92%	14.25
	22	5%	366.21	335.35	-0.022%	-0.024%	-5.40%	-5.86%	58.14%	57.57%	9.72
		15%	975.51	890.66	-0.001%	-0.003%	-0.14%	-0.64%	79.64%	70.44%	10.05
		25%	1439.62	1318.31	0.008%	0.006%	2.03%	1.54%	91.78%	91.76%	9.72
SOR	5	5%	889.37	833.44	-0.003%	-0.004%	-0.65%	-1.00%	11.77%	9.66%	7.18
		15%	711.40	639.88	-0.007%	-0.010%	-1.86%	-2.44%	77.91%	69.93%	11.71
		25%	612.53	531.87	-0.011%	-0.014%	-2.67%	-3.43%	27.74%	26.50%	15.57
	22	5%	990.54	928.11	0.000%	-0.002%	-0.05%	-0.41%	61.36%	59.98%	7.18
		15%	1144.50	1038.78	0.003%	0.001%	0.75%	0.21%	74.47%	70.98%	10.70
		25%	1309.79	1219.98	0.006%	0.004%	1.50%	1.10%	99.41%	99.28%	7.84
CAR	5	5%	804.55	779.27	-0.005%	-0.005%	-1.19%	-1.37%	42.82%	41.77%	3.53
		15%	724.93	682.72	-0.007%	-0.008%	-1.76%	-2.09%	51.08%	50.98%	6.63
		25%	1027.01	1023.96	0.001%	0.001%	0.15%	0.13%	98.47%	98.44%	0.33
	22	5%	974.99	898.33	-0.001%	-0.002%	-0.14%	-0.59%	73.64%	49.90%	9.06
		15%	1126.22	1073.54	0.003%	0.002%	0.66%	0.39%	99.47%	99.47%	5.30
		25%	867.42	859.65	-0.003%	-0.003%	-0.78%	-0.83%	50.89%	50.89%	0.99

Tabella 8: statistiche riepilogative G.MI, N=5.

Le performance del titolo G.MI sono mediamente al di sotto del capitale iniziale, ad ogni modo è possibile trarne alcune riflessioni positive. La tendenza individuata, in questo caso, risulta essere una predilezione per l'azione *greedy* nel caso di $L = 5$, al contrario una preferenza per l'azione esplorativa nel caso di $L = 22$. Tali risultati possono dipendere dai seguenti fattori: poiché il titolo presenta in molte fasi un andamento dei prezzi che non segue una fase *trending*⁵⁶, considerare nella computazione della funzione di reward un numero di rendimenti maggiore rende l'algoritmo meno sensibile alle continue oscillazioni dei rendimenti di mercato e di conseguenza la volatilità viene *lisciata* e valutata considerando un periodo più ampio; in questo caso, l'azione esplorativa risulta essere più efficace, come da attese ed in quanto in presenza di un ambiente fortemente stocastico, e si raggiungono alcuni risultati positivi. Per le stesse ragioni appena esposte, si ottiene maggior efficacia del sistema di trading settando $N = 5$ rispetto ad $N = 1$: infatti, le due migliori performance si ottengono con i setting $N = 5, SR, L = 22, \varepsilon = 25\%$ ed $N = 5, SOR, L = 22, \varepsilon = 25\%$ (rispettivamente 1,54% con ed 1,10% di rendimento netto medio annuo, per un capitale finale di 1318.31€ operando mediamente $\# = 9.72$ azioni annue nel primo caso e di 1219.98€ con $\# = 7.84$ nel secondo). Confrontando queste due performance si può notare come, sebbene con lo Sharpe Ratio si raggiunga un rendimento finale maggiore, utilizzando il Sortino Ratio si ottenga un miglior andamento dell'*equity line* durante tutto il periodo: questa infatti si mantiene costantemente al di sopra del livello del capitale inizialmente investito (99,41% lorda, 99,28% netta) in maniera migliore rispetto all'*equity line* ottenuta con lo Sharpe Ratio (91,78% lorda, 91,76% netta). I risultati sono riportati nei grafici della *figura 17*. Permane, invece, il problema legato all'operatività utilizzando il Calmar Ratio: aumentando l'incidenza dell'azione esplorativa, questa crolla a livelli vicini allo zero, ad esempio $\# = 0,33$ settando $N = 5, L = 5, \varepsilon = 25\%$, oppure $\# = 0,99$ settando $N = 5, L = 22, \varepsilon = 25\%$. Tali risultati confermano le problematiche già individuate analizzando le performance del titolo UCG.MI, aggravate in questo caso dalla diversa struttura della serie storica del titolo G.MI che rende, nei due casi ad esempio, l'operatività media totale pressoché nulla. Sebbene le performance ottenute non siano

⁵⁶ L'andamento dei prezzi si definisce in fase *trending* se segue un trend definito ed individuabile. I prezzi, questo caso, non seguono un trend definito e vi sono numerose oscillazioni in successione.

pienamente soddisfacenti, dalla *tabella 9* (di seguito) si possono comunque trarre conclusioni positive sull'operatività individuata dall'algoritmo.

rwd	N=1			N=5		
	perc_over	MEDIE	#	perc_over	MEDIE	#
SR	80.12%	73.40%	11.53	81.12%	78.38%	9.70
SOR	69.55%	56.76%	9.95	58.78%	56.06%	10.03
CAR	56.13%	45.12%	5.61	69.40%	65.24%	4.31
ALL	68.60%	58.42%	9.03	69.76%	66.56%	8.01
ALL - CAR	74.83%	65.08%	10.74	69.95%	67.22%	9.87

Tabella 9: percentuale media equity line > cap. investito per funzione di reward, titolo G.MI.

Considerando i dati al netto del Calmar Ratio, la percentuale totale (voce *ALL*) di volte nelle quali l'*equity line* è maggiore del capitale iniziale si attesta intorno al 70% per le *equity line* lorde, mentre per quanto riguarda le *equity line* nette vi è un miglioramento di 7 punti percentuali passando da un settaggio $N = 1$ ad un settaggio $N = 5$, avvalorata da una riduzione di operatività media di circa un'azione annua (# passa da 10.77 a 9.87) che diminuisce l'incidenza dei costi di transazione e che consente dunque di ottenere performance migliori agendo meno sul mercato. Lo Sharpe Ratio si conferma essere una buona funzione di reward per il titolo G.MI, poiché con essa si ottengono performance migliori. Ciò nonostante, il Sortino Ratio presenta, per $N = 5$, una percentuale superiore al 50%, il che comporta una valutazione comunque positiva del suo utilizzo. La maggior efficacia dello Sharpe rispetto al Sortino può dipendere dal fatto che, in periodi di oscillazioni senza un preciso trend (come si configura il secondo periodo temporale del campione considerato), viene in qualche modo penalizzata la sua efficacia rispetto allo Sharpe e quest'ultimo, considerando scostamenti positivi e negativi dal valor medio del periodo e non solamente il *downside risk*, riesce a mappare in misura più completa la volatilità fornendo indicazioni più efficaci all'algoritmo. Per quanto concerne il Calmar Ratio, questo porta le percentuali dell'*equity line*, nel caso di $N = 5$, a risultati migliori rispetto al Sortino, riducendo però l'operatività di quasi 6 azioni medie annue e non candidandosi dunque ad essere un adeguato metro di paragone con l'indice che utilizza il *downside risk*.

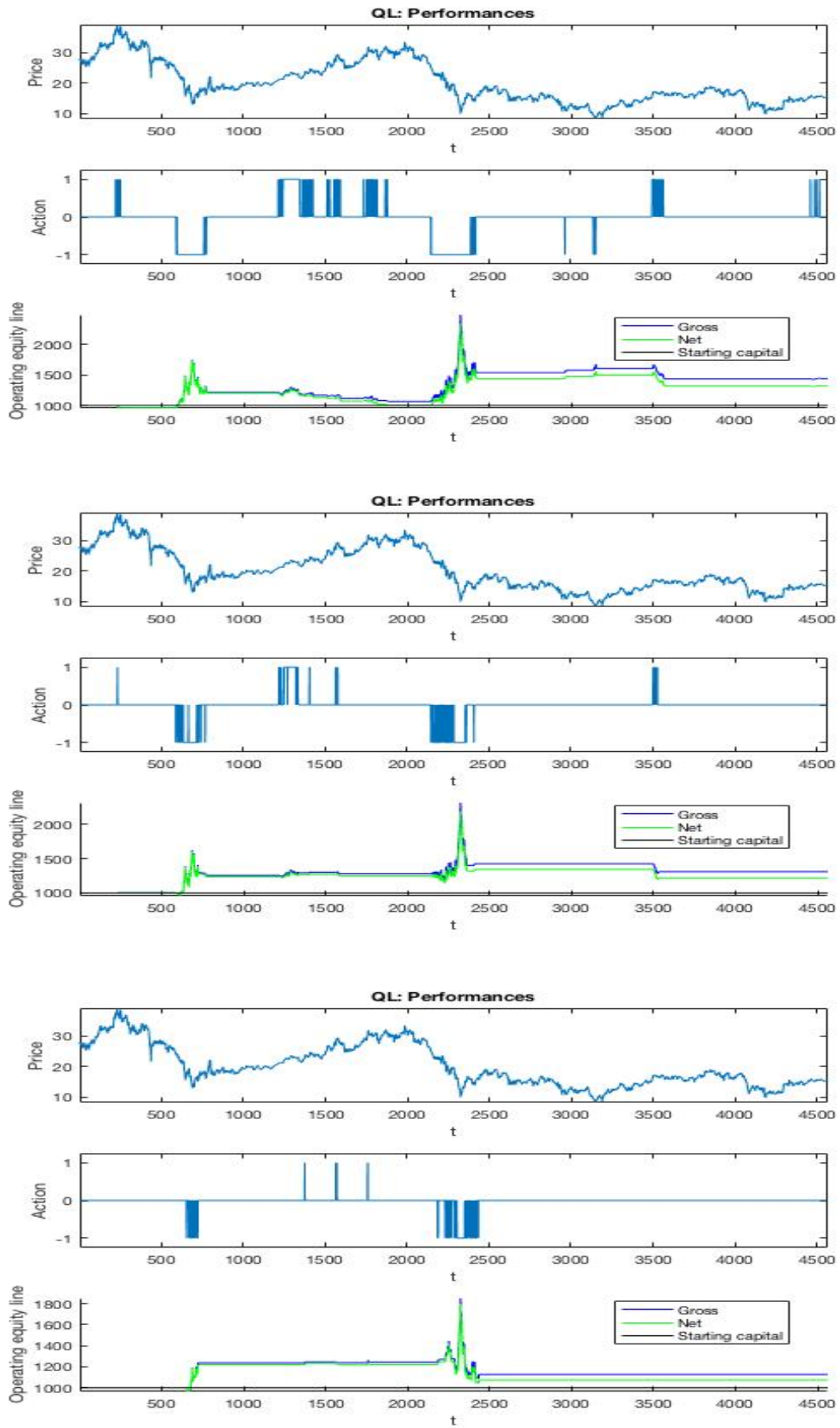


Figura 17: performance titolo G.MI con settaggi (rappresentati dall'alto verso il basso):

1. $N=5$, SR , $L=22$, $\varepsilon=0,25$.
2. $N=5$, SOR , $L=22$, $\varepsilon=0,25$.
3. $N=5$, CAR , $L=22$, $\varepsilon=0,15$.

Eni S.p.A.

N=1			eq_line		rend giorn		rend annuo		perc_over		
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	435.93	411.74	-0.018%	-0.019%	-4.48%	-4.78%	8.71%	8.30%	6.29
		15%	282.95	255.18	-0.028%	-0.030%	-6.73%	-7.26%	6.70%	6.63%	11.37
		25%	353.88	312.52	-0.023%	-0.025%	-5.57%	-6.22%	30.04%	25.60%	13.68
	22	5%	627.87	603.26	-0.010%	-0.011%	-2.54%	-2.75%	28.03%	27.09%	4.41
		15%	460.24	410.66	-0.017%	-0.019%	-4.19%	-4.79%	18.15%	12.61%	12.58
		25%	945.01	888.12	-0.001%	-0.003%	-0.31%	-0.65%	80.18%	79.53%	6.84
SOR	5	5%	520.33	489.00	-0.014%	-0.016%	-3.54%	-3.87%	17.98%	15.28%	6.84
		15%	381.75	350.59	-0.021%	-0.023%	-5.17%	-5.62%	12.04%	11.08%	9.38
		25%	501.21	453.56	-0.015%	-0.017%	-3.74%	-4.27%	7.25%	7.18%	11.04
	22	5%	935.90	876.98	-0.001%	-0.003%	-0.36%	-0.72%	56.97%	26.78%	7.17
		15%	693.98	645.61	-0.008%	-0.010%	-2.00%	-2.39%	30.15%	27.44%	7.95
		25%	914.76	848.61	-0.002%	-0.004%	-0.49%	-0.90%	8.96%	8.15%	8.28
CAR	5	5%	501.16	484.88	-0.015%	-0.016%	-3.74%	-3.92%	26.14%	26.12%	3.64
		15%	1013.17	993.10	0.000%	0.000%	0.07%	-0.04%	91.96%	72.32%	2.21
		25%	993.98	992.98	0.000%	0.000%	-0.03%	-0.04%	41.76%	41.76%	0.11
	22	5%	404.72	370.19	-0.020%	-0.022%	-4.87%	-5.34%	49.18%	40.51%	9.82
		15%	835.74	796.53	-0.004%	-0.005%	-0.99%	-1.25%	24.68%	24.68%	5.30
		25%	998.91	991.93	0.000%	0.000%	-0.01%	-0.04%	29.28%	26.87%	0.77

Tabella 10: statistiche riepilogative ENI.MI, N=1.

N=5			eq_line		rend giorn		rend annuo		perc_over		
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	356.32	332.20	-0.023%	-0.024%	-5.54%	-5.90%	8.92%	8.77%	7.73
		15%	428.96	392.00	-0.019%	-0.021%	-4.57%	-5.04%	28.99%	27.99%	9.94
		25%	497.70	447.62	-0.015%	-0.018%	-3.78%	-4.34%	32.87%	31.32%	11.71
	22	5%	670.91	643.26	-0.009%	-0.010%	-2.18%	-2.41%	22.70%	19.99%	4.64
		15%	606.12	557.77	-0.011%	-0.013%	-2.73%	-3.17%	51.94%	34.67%	9.17
		25%	1307.16	1255.94	0.006%	0.005%	1.49%	1.27%	100.00%	100.00%	4.42
SOR	5	5%	583.15	538.81	-0.012%	-0.014%	-2.93%	-3.36%	8.02%	6.40%	8.73
		15%	438.80	397.86	-0.018%	-0.020%	-4.45%	-4.96%	13.30%	12.23%	10.82
		25%	765.73	688.69	-0.006%	-0.008%	-1.46%	-2.04%	34.80%	32.59%	11.71
	22	5%	799.03	766.90	-0.005%	-0.006%	-1.23%	-1.46%	27.53%	19.46%	4.53
		15%	1016.33	959.98	0.000%	-0.001%	0.09%	-0.23%	78.70%	58.19%	6.30
		25%	922.48	866.16	-0.002%	-0.003%	-0.44%	-0.79%	28.73%	25.58%	6.96
CAR	5	5%	418.69	401.86	-0.019%	-0.020%	-4.69%	-4.91%	41.95%	41.95%	4.53
		15%	983.81	954.72	0.000%	-0.001%	-0.09%	-0.26%	72.74%	71.62%	3.31
		25%	981.02	979.05	0.000%	0.000%	-0.11%	-0.12%	30.86%	30.86%	0.22
	22	5%	1369.91	1332.11	0.007%	0.006%	1.75%	1.60%	97.90%	97.90%	3.09
		15%	1034.02	1007.52	0.001%	0.000%	0.18%	0.04%	96.60%	96.32%	2.87
		25%	989.72	988.73	0.000%	0.000%	-0.06%	-0.06%	80.69%	80.69%	0.11

Tabella 11: statistiche riepilogative ENI.MI, N=5.

Le performance del titolo ENI.MI non sono mediamente positive. Indipendentemente dai settaggi utilizzati, l'algoritmo non riesce a sfruttare i *feedback* ambientali per generare una strategia di trading efficace che conduca a dei profitti soddisfacenti: la serie dei prezzi presenta un numero elevato di oscillazioni e l'algoritmo non è in grado di adattarsi velocemente al mercato, producendo così un *downgrade* dei profitti nella quasi totalità dei casi. Come infatti è possibile appurare dal grafico (*figura 18*), le 500 *equity line* ottenute nel caso ad esempio presentano un'elevata dispersione ed orientamento decrescente (tale andamento si registra comunque in quasi tutti i settaggi impostati), sintomo questo di una difficoltà nell'apprendere completamente il comportamento del titolo, difficoltà che si traduce nell'individuare, in fase operativa, la corretta operazione di trading da intraprendere.

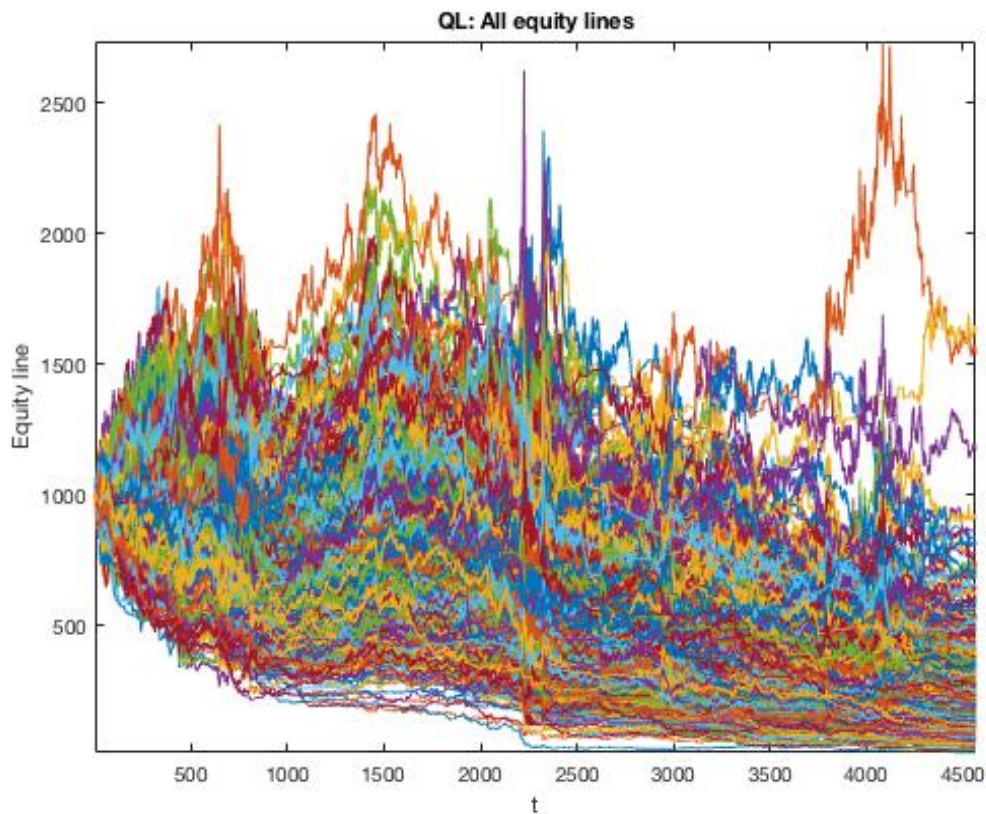


Figura 18: equity line totali, titolo ENI.MI ($N=1$, SOR , $L=5$, $\varepsilon=0,15$).

In questo caso, il *drawdown* si configura essere una misura più *conservativa* del capitale inizialmente investito e leggermente più efficace: è infatti con questa funzione di reward che si ottiene il profitto più elevato (1332.11€ di capitale netto pari 1.60% di rendimento medio annuo effettuando $\# = 3.09$ operazioni medie all'anno, settando $N = 5, CAR, L = 22, \varepsilon = 5\%$), evidenziando comunque scarsa operatività

dell'algoritmo (questa decade a zero aumentando ε). D'altro canto, la tendenza generale individuata, utilizzando le altre due funzioni di reward, è quella di ridurre l'operatività all'aumentare delle informazioni rese disponibili all'algoritmo: infatti, aumentando N (da 1 a 5) ed L (da 5 a 22), l'operatività in base allo Sharpe Ratio passa da una media di $\# = 10.45$ ($N = 1, L = 5$) ad una media di $\# = 6.07$ ($N = 5, L = 22$) azioni annue; allo stesso modo, per il Sortino Ratio, l'operatività passa da $\# = 7.95$ ($N = 1, L = 5$) a $\# = 5.39$ ($N = 5, L = 22$). Questa riduzione non porta ad un peggioramento dei profitti, bensì ad un miglioramento: le migliori performance per le due funzioni di reward citate si ottengono settando $N = 5, SR, L = 22, \varepsilon = 25\%$ ed $N = 5, SOR, L = 22, \varepsilon = 15\%$, con risultati pari a 1.27% di rendimento netto medio annuo (1307.16€ di capitale finale, $\# = 4.42$) nel primo caso e -0.23% (959.98€ di capitale finale, $\# = 6.30$); a pag. 117 vengono riportati i grafici con le tre migliori performance (*figura 19*). Il Sortino Ratio presenta, rispetto allo Sharpe, perdite più contenute poiché penalizza solamente il *downside risk* e solo una volatilità al ribasso dei prezzi, ma non riesce a performare in nessun caso in maniera positiva. Uno dei motivi principali può essere, oltre al numero di iterazioni, il sistema di trading operativo settato (pag. 97): individuata la tendenza a ridurre il numero di azioni all'aumentare dei settaggi di N ed L nel caso di titoli difficili da apprendere, si è implementato per il titolo in esame un numero maggiore di iterazioni ($k = 1000$) variando il sistema operativo di trading aumentando l'intervallo dell'azione *stay out of the market* ($a = 0$) da $[-1/3, +1/3]$ a $[-1/2, +1/2]$. Il settaggio utilizzato con questa nuova impostazione è $N = 1, SOR, L = 5, \varepsilon = 15\%$, il quale aveva comportato una performance negativa pari a -5.62% di rendimento netto annuo. I risultati della simulazione sono i seguenti: il capitale netto finale ottenuto è pari a 1016.99€ per un rendimento netto medio annuo pari a 0.09% ed un numero medio di azioni di $\# = 5.63$, l'equity line netta risulta superiore al capitale inizialmente investito per il 95.1% del periodo. Chiaramente questi valori, in termini assoluti, non sono propriamente positivi, ma se confrontati con il precedente settaggio, ovvero $k = 500$ ed azione $a = 0$ compresa nell'intervallo $[-1/3, +1/3]$, assumono connotati decisamente positivi, poiché la performance finale è quadruplicata ed è migliore l'efficacia del sistema di trading (*figura 19* alla pagina seguente).

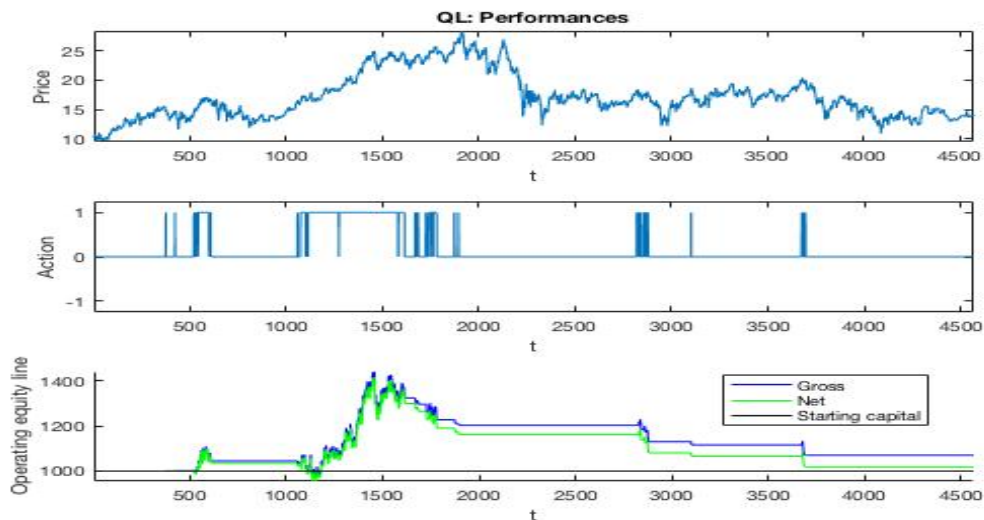


Figura 19: performance titolo ENI.MI con settaggio $k=1000$, $N=1$, SOR , $L=5$, $\varepsilon=0,15$.

Nel caso di titoli di difficile apprendimento per l' algoritmo, una soluzione adeguata è dunque quella di aumentare il numero di iterazioni ed aggiustare il sistema di trading, in maniera tale da renderlo più efficace al contesto di riferimento ed ottenere così migliori performance. Una riflessione, di carattere generale, riguardo questo tema viene affrontata al paragrafo seguente.

Tornando ai risultati generali, la *tabella 12* riepiloga le percentuali medie, per funzione di reward, di volte nelle quali l'equity line è maggiore di C . I risultati si conformano a quelli delle performance ottenute, evidenziando scarsa efficacia del sistema di trading. Le percentuali migliorano aumentando il numero di rendimenti inclusi nel vettore s_t , ma sono comunque insoddisfacenti poiché inferiori al 50% per lo Sharpe Ratio e per il Sortino Ratio. Il Calmar presenta buone percentuali nel caso di $N = 5$ ma un'operatività media non sufficientemente elevata da consentirne una valutazione positiva.

rwd	N=1			N=5		
	perc_over MEDIE		#	perc_over		#
	G%	N%		G%	N%	
SR	28.64%	26.63%	9.20	40.91%	37.12%	7.93
SOR	22.22%	15.98%	8.44	31.85%	25.74%	8.17
CAR	43.83%	38.71%	3.64	70.12%	69.89%	2.36
ALL	31.56%	27.11%	7.09	47.62%	44.25%	6.15
ALL - CAR	25.43%	21.31%	8.82	36.38%	31.43%	8.05

Tabella 12: percentuale media equity line > cap. investito per funzione di reward, titolo ENI.MI.

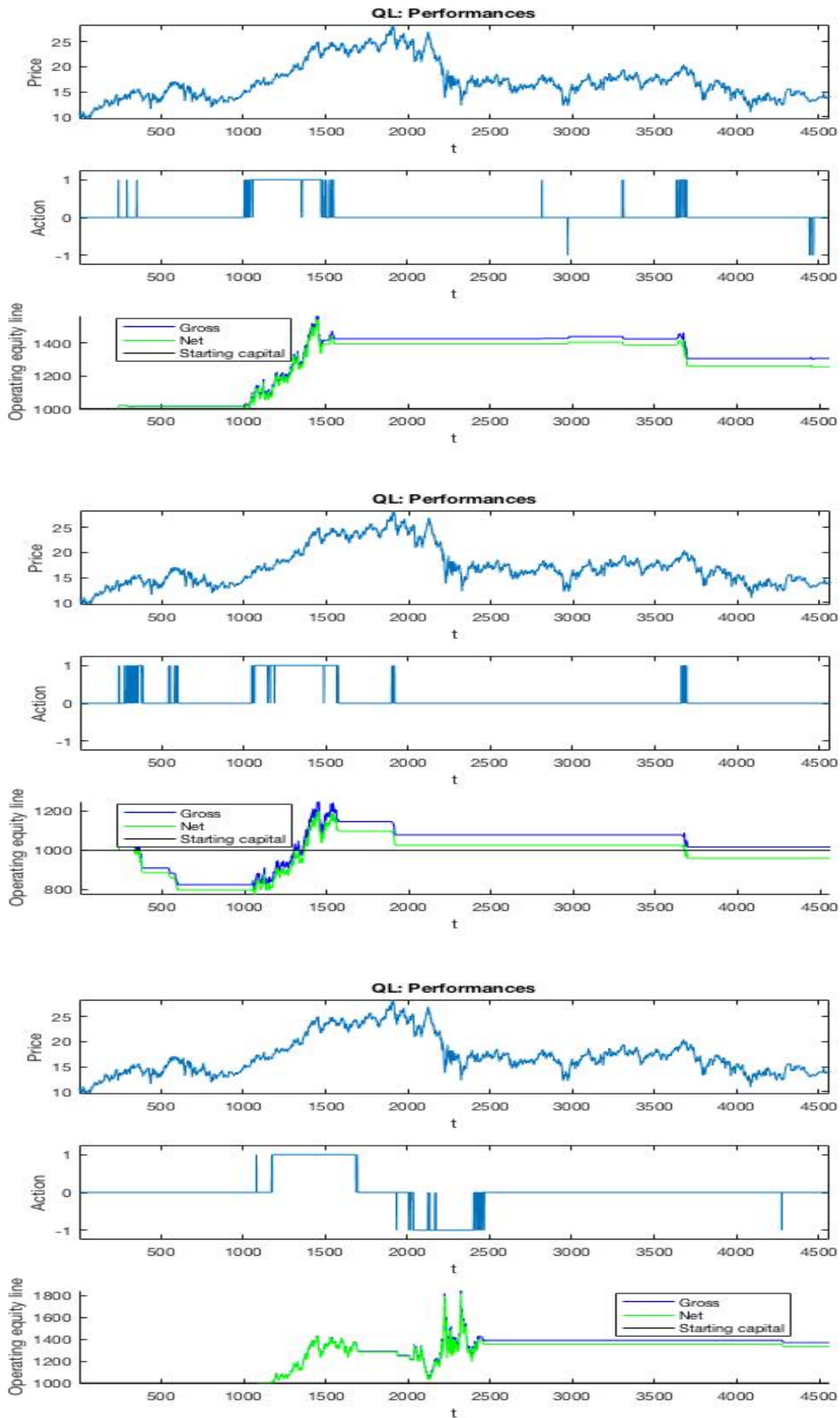


Figura 20: performance titolo ENI.MI con settaggi (rappresentati dall'alto verso il basso):

1. $N=5$, SR , $L=22$, $\varepsilon=0,25$.
2. $N=5$, SOR , $L=22$, $\varepsilon=0,15$.
3. $N=5$, CAR , $L=22$, $\varepsilon=0,05$.

Fiat Chrysler Automotive S.p.A.

N=1			eq_line			rend giorn		rend annuo		perc_over	
rwd	L	e	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	1145.63	1081.48	0.003%	0.002%	0.75%	0.43%	82.88%	82.20%	6.35
		15%	1882.27	1727.70	0.014%	0.012%	3.55%	3.06%	93.61%	91.96%	9.44
		25%	4276.49	3805.74	0.032%	0.029%	8.35%	7.65%	94.18%	93.74%	12.86
	22	5%	1027.85	991.92	0.001%	0.000%	0.15%	-0.04%	71.12%	70.66%	3.92
		15%	2448.56	2248.01	0.020%	0.018%	5.07%	4.57%	99.98%	99.96%	9.44
		25%	3295.96	2995.95	0.026%	0.024%	6.80%	6.24%	99.69%	99.67%	10.54
SOR	5	5%	1284.14	1198.06	0.005%	0.004%	1.39%	1.00%	86.80%	86.69%	7.67
		15%	2317.72	2134.09	0.018%	0.017%	4.75%	4.27%	96.08%	96.01%	9.10
		25%	2519.48	2231.06	0.020%	0.018%	5.23%	4.53%	92.47%	92.01%	13.41
	22	5%	1344.65	1268.17	0.006%	0.005%	1.65%	1.32%	80.97%	79.02%	6.46
		15%	3082.91	2879.04	0.025%	0.023%	6.41%	6.01%	98.86%	98.75%	7.56
		25%	2460.38	2174.57	0.020%	0.017%	5.09%	4.38%	98.86%	98.71%	13.63
CAR	5	5%	1377.16	1324.60	0.007%	0.006%	1.78%	1.56%	85.81%	85.40%	4.30
		15%	1407.95	1370.37	0.007%	0.007%	1.91%	1.75%	97.79%	97.59%	2.98
		25%	1081.17	1064.03	0.002%	0.001%	0.43%	0.34%	99.80%	99.80%	1.77
	22	5%	2114.66	2018.73	0.016%	0.015%	4.22%	3.95%	100.0%	100.00%	5.13
		15%	1266.33	1171.35	0.005%	0.003%	1.31%	0.87%	80.20%	78.74%	8.60
		25%	1130.54	1080.76	0.003%	0.002%	0.68%	0.43%	95.14%	95.01%	4.97

Tabella 13: statistiche riepilogative FCA.MI, N=1.

N=5			eq_line			rend giorn		rend annuo		perc_over	
rwd	L	e	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	1867.24	1798.74	0.014%	0.013%	3.51%	3.30%	93.38%	93.27%	4.14
		15%	1592.99	1468.12	0.010%	0.008%	2.60%	2.14%	81.17%	78.22%	9.00
		25%	3249.29	2926.76	0.026%	0.024%	6.72%	6.11%	92.20%	91.58%	11.54
	22	5%	1225.25	1177.70	0.004%	0.004%	1.13%	0.91%	72.45%	72.10%	4.36
		15%	1644.30	1492.90	0.011%	0.009%	2.78%	2.24%	99.41%	98.49%	10.66
		25%	1627.68	1472.14	0.011%	0.008%	2.73%	2.16%	99.45%	99.43%	11.10
SOR	5	5%	1275.22	1207.53	0.005%	0.004%	1.35%	1.05%	54.61%	53.28%	6.02
		15%	2243.10	2055.41	0.018%	0.016%	4.56%	4.06%	90.07%	89.96%	9.66
		25%	2489.15	2249.25	0.020%	0.018%	5.17%	4.58%	90.01%	84.97%	11.21
	22	5%	1171.54	1096.24	0.003%	0.002%	0.88%	0.51%	94.00%	86.50%	7.35
		15%	2925.11	2709.77	0.024%	0.022%	6.11%	5.66%	99.76%	99.76%	8.45
		25%	1474.04	1280.76	0.009%	0.005%	2.17%	1.38%	73.77%	73.46%	15.52
CAR	5	5%	707.05	673.57	-0.008%	-0.009%	-1.90%	-2.16%	47.47%	47.10%	5.36
		15%	1595.91	1506.07	0.010%	0.009%	2.62%	2.29%	100.0%	100.0%	6.41
		25%	953.76	945.20	-0.001%	-0.001%	-0.26%	-0.31%	42.19%	42.17%	0.99
	22	5%	1268.73	1220.21	0.005%	0.004%	1.32%	1.11%	83.34%	83.32%	4.31
		15%	1371.22	1308.27	0.007%	0.006%	1.76%	1.50%	100.00%	100.00%	5.19
		25%	979.04	946.25	0.000%	-0.001%	-0.12%	-0.30%	36.18%	35.74%	3.76

Tabella 14: statistiche riepilogative FCA.MI, N=5.

Le performance relative al titolo FCA.MI sono molto consistenti. L’algoritmo riesce ad apprendere velocemente l’ambiente di riferimento, in special modo utilizzando lo Sharpe Ratio ed il Sortino Ratio, e permette di ottenere performance positive per ogni settaggio utilizzato. Il Calmar Ratio presenta le caratteristiche già evidenziate in precedenza, conservando in questo caso una certa operatività impostando preferenza per l’azione *greedy* ($\varepsilon = 5\%$) e settando una numerosità pari ad $N = 1$, operatività che decresce all’aumento di ε e di N . Le performance ottenute con il Calmar Ratio non si discostano molto rispetto a quelle ottenute utilizzando gli altri due indici, pur non presentando gli *exploit* ottenuti con le due precedenti funzioni di reward: il motivo risiede nel fatto che, dato il carattere *conservativo* dell’indice che utilizza il *drawdown*, in presenza di serie di prezzi le cui opportunità di profitto sono relativamente semplici da apprendere per l’algoritmo la riduzione di operatività può comportare un impatto negativo sul livello di profitto finale di ogni settaggio. La relativa semplicità d’apprendimento dell’algoritmo si può appurare dal seguente grafico (*figura 21*), nel quale sono rappresentate le 500 *equity line* ottenute impostando il settaggio $N = 5, SR, L = 22, \varepsilon = 25\%$.

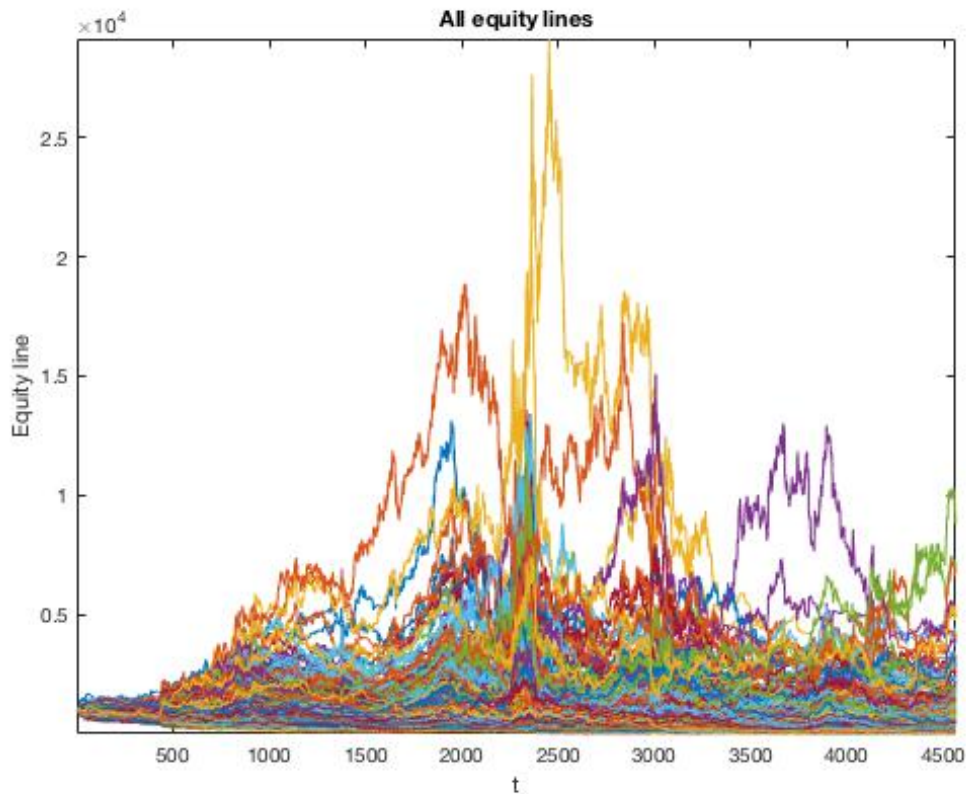


Figura 21: equity line totali, titolo FCA.MI ($N=5, SR, L=22, \varepsilon=0,25$).

Nonostante una predilezione per l'azione esplorativa ($\varepsilon = 25\%$), vi è estrema compattezza nella distribuzione delle *equity line*: queste, salvo qualche serie isolata, si raccolgono in un intorno ben definito, segnale questo che il processo di selezione delle azioni operative riesce a cogliere un trend ben definito nella serie dei rendimenti ed a fornire una operatività idonea.

I settaggi più performanti per il titolo FCA.MI (riportati in *figura 22* a pag. 122) risultano essere i migliori dell'intero *pool* di titoli e settaggi considerati in questo lavoro: impostando $N = 1, SR, L = 5, \varepsilon = 25\%$ si ottiene un capitale netto finale di 3805.74€ (pari al 7,65% di rendimento netto medio annuo operando $\# = 12.86$ azioni di trading l'anno), impostando $N = 1, SOR, L = 22, \varepsilon = 15\%$ si ottiene un capitale netto di 2879.04€ (6,01% di rendimento medio annuo, $\# = 7.56$ azioni operative medie), impostando infine $N = 1, CAR, L = 22, \varepsilon = 5\%$ si ottiene un capitale netto pari a 2018.73€ (3,95% di rendimento medio annuo, $\# = 5.13$ azioni operative). Si riscontra, in generale, un miglioramento medio delle performance aumentando l'incidenza dell'azione esplorativa, tendenza che accomuna dunque le serie di prezzi di apprendimento relativamente semplice per l'algoritmo; inoltre, si ottengono performance migliori considerando $N = 1$ rendimenti descrittivi dello stato s_t rispetto ad $N = 5$, tendenza in linea con le aspettative iniziali. In entrambi i casi di settaggio di s_t , l'indice di Sortino si rivela essere migliore dell'indice di Sharpe per valori di ε contenuti (5%, 15%), con alcune differenze anche sostanziali: ad esempio, impostando $N = 5, L = 22, \varepsilon = 15\%$, la differenza di rendimento netto medio annuo è di 3 punti percentuali (5,56% utilizzando il Sortino Ratio contro i 2,26% utilizzando lo Sharpe Ratio). Impostando invece $\varepsilon = 25\%$ la tendenza si inverte: l'algoritmo performa in modo più efficiente utilizzando una funzione di reward che mappa totalmente la volatilità e non solamente il *downside risk*. Il Calmar Ratio fornisce problemi operativi e computazionali in maniera ridotta rispetto agli altri titoli: si riscontra una tendenza inversa rispetto allo Sharpe ed al Sortino, ovvero quella di prediligere azioni *greedy* rispetto ad azioni esplorative, ad ogni modo l'operatività media, sebbene diminuisca in maniera più che evidente all'aumentare di ε . Si registra comunque, in quasi la totalità di casi, un aumento di operatività da $\varepsilon = 5\%$ ad $\varepsilon = 15\%$, operatività che poi tende ad annullarsi per $\varepsilon = 25\%$; ciò nonostante, per alcuni settaggi, risulta essere significativa

(ad esempio, impostando $N = 1$, CAR , $L = 22$, $\varepsilon = 15\%$ questa è pari a $\# = 8.60$, in linea con le operatività medie ottenute, mediante gli stessi settaggi, con le altre due funzioni di reward) e comporta, in alcuni casi, anche il raggiungimento di un profitto maggiore (ad esempio, nel caso $N = 1$, $L = 5$, $\varepsilon = 5\%$, il profitto ottenuto utilizzando il CAR è più elevato rispetto agli altri due).

La *tabella 15* conferma i buoni risultati analizzati finora, evidenziando come l'*equity line*, sia lorda che netta, si mantenga in maniera ampiamente soddisfacente al di sopra del capitale iniziale C . La percentuale totale netta per $N = 1$, non considerando il Calmar Ratio (il quale, in questo caso, non ha comunque un impatto rilevante sulla media totale poiché le percentuali delle tre diverse funzioni di reward sono quasi equivalenti) si attesta al 90.78%, risultato pienamente soddisfacente. Il Sortino Ratio performa leggermente meglio dello Shapre Ratio nel caso di $N = 1$, esito che si inverte passando ad $N = 5$; ad ogni modo, la differenza è nell'ordine di qualche punto percentuale e, visti i valori elevati ottenuti con entrambe le funzioni di reward, il giudizio finale è positivo per entrambe. Perdurando, valutando le medie in tabella, la bassa operatività legata al Calmar Ratio anche se, come approfondito in precedenza, vi sono alcuni settaggi che conducono a risultati positivi; per $N = 1$ le *equity line* (lorda e netta) presentano valori superiori al 90%, mentre per $N = 5$ la situazione peggiora individuando che, nonostante la bassa operatività, il Calmar Ratio riesce a condurre a performance positive nel caso di $N = 1$ mentre la sua efficacia declina nel caso di $N = 5$.

	N=1			N=5		
	perc_over	MEDIE		perc_over	MEDIE	
	G%	N%	#	G%	N%	#
ALL	91.90%	91.44%	7.67	80.53%	79.41%	7.50
ALL - CAR	91.29%	90.78%	9.20	86.69%	85.08%	9.08
SR	90.24%	89.70%	8.75	89.68%	88.85%	8.47
SOR	92.34%	91.87%	9.64	83.70%	81.32%	9.70
CAR	93.12%	92.76%	4.63	68.20%	68.05%	4.34

Tabella 15: percentuale media equity line>cap. investito per funzione di reward, titolo FCA.MI.

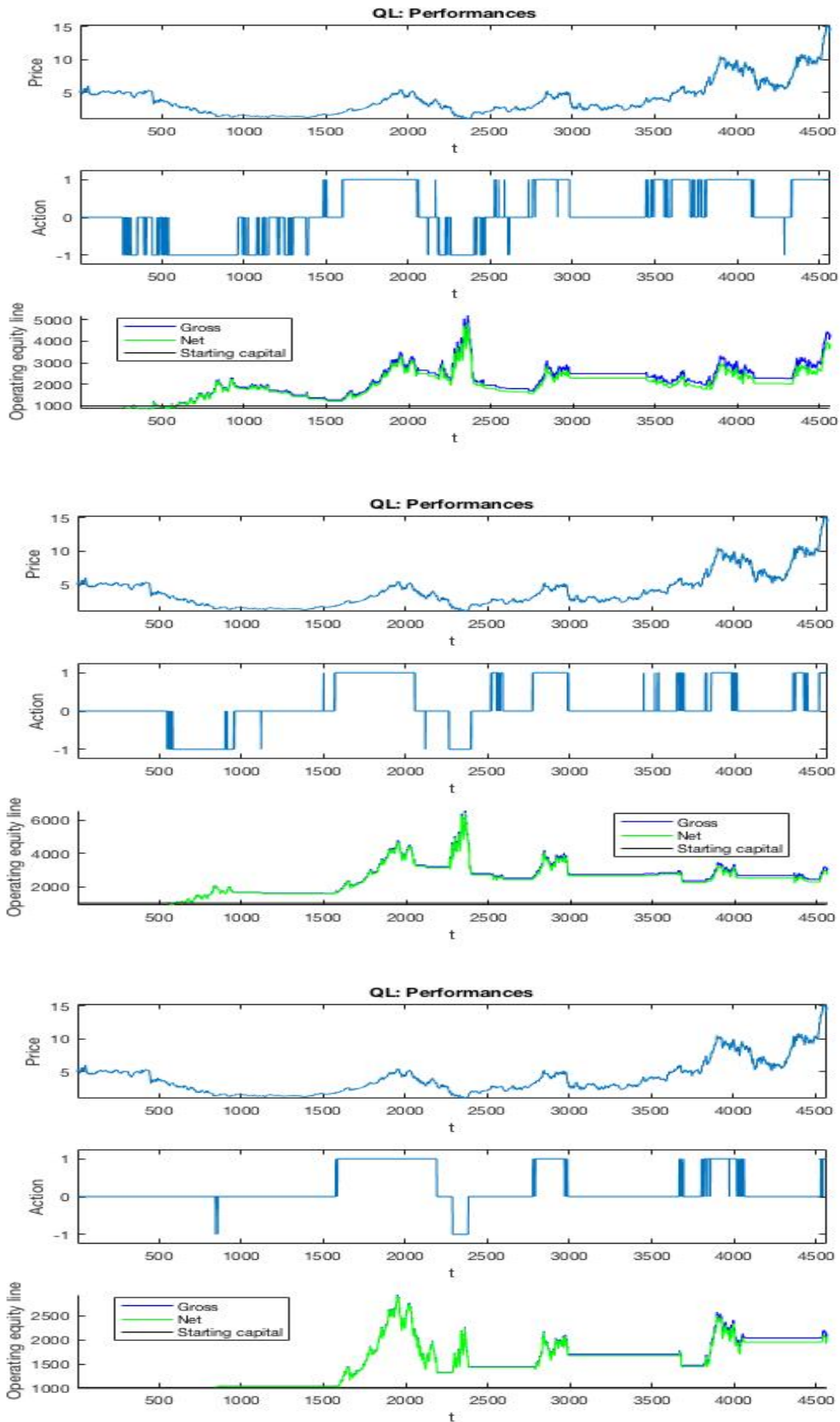


Figura 22: performance titolo FCA.MI con settaggi (rappresentati dall'alto verso il basso):

1. $N=1$, SR , $L=5$, $\varepsilon=0,25$.
2. $N=1$, SOR , $L=22$, $\varepsilon=0,15$.
3. $N=1$, CAR , $L=5$, $\varepsilon=0,05$.

Telecom Italia S.p.A.

N=1			eq_line		rend giorn		rend annuo		perc_over		
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	727.33	665.68	-0.007%	-0.009%	-1.74%	-2.22%	50.71%	39.44%	9.77
		15%	477.93	419.05	-0.016%	-0.019%	-3.99%	-4.69%	18.79%	12.20%	14.51
		25%	217.49	185.76	-0.033%	-0.037%	-8.07%	-8.87%	6.09%	6.07%	17.38
	22	5%	366.93	340.57	-0.022%	-0.024%	-5.38%	-5.77%	26.03%	24.70%	8.22
		15%	176.98	146.57	-0.038%	-0.042%	-9.11%	-10.05%	7.99%	7.88%	20.80
		25%	422.32	382.82	-0.019%	-0.021%	-4.65%	-5.16%	8.96%	8.41%	10.81
SOR	5	5%	321.18	282.39	-0.025%	-0.028%	-6.07%	-6.74%	2.04%	1.69%	14.18
		15%	204.23	172.54	-0.035%	-0.038%	-8.39%	-9.24%	1.95%	1.47%	18.60
		25%	278.64	238.23	-0.028%	-0.031%	-6.81%	-7.61%	1.49%	1.45%	17.27
	22	5%	352.09	326.26	-0.023%	-0.025%	-5.60%	-5.99%	0.94%	0.94%	8.44
		15%	312.84	274.09	-0.025%	-0.028%	-6.21%	-6.89%	9.66%	9.66%	14.57
		25%	591.49	550.37	-0.011%	-0.013%	-2.86%	-3.24%	9.70%	9.70%	7.95
CAR	5	5%	573.47	549.33	-0.012%	-0.013%	-3.02%	-3.25%	28.71%	28.27%	4.75
		15%	671.45	652.82	-0.009%	-0.009%	-2.17%	-2.33%	27.15%	27.02%	3.09
		25%	984.27	982.29	0.000%	0.000%	-0.09%	-0.10%	47.95%	48.00%	0.22
	22	5%	758.20	721.19	-0.006%	-0.007%	-1.52%	-1.79%	56.45%	56.10%	5.52
		15%	1185.24	1166.51	0.004%	0.003%	0.94%	0.85%	99.34%	99.21%	1.77
		25%	962.40	952.78	-0.008%	-0.001%	-0.21%	-0.27%	49.40%	48.92%	1.10

Tabella 16: statistiche riepilogative TIT.MI, N=1.

N=5			eq_line		rend giorn		rend annuo		perc_over		
rwd	L	ε	G €	N €	G%	N%	G%	N%	G%	N%	#
SR	5	5%	643.37	601.32	-0.010%	-0.011%	-2.41%	-2.77%	13.43%	10.59%	7.46
		15%	378.23	328.95	-0.021%	-0.024%	-5.23%	-5.96%	27.13%	13.28%	15.41
		25%	386.47	329.15	-0.021%	-0.024%	-5.12%	-5.95%	6.47%	6.38%	17.73
	22	5%	272.98	255.66	-0.028%	-0.030%	-6.92%	-7.26%	11.86%	11.51%	7.23
		15%	308.11	264.90	-0.026%	-0.029%	-6.29%	-7.07%	9.38%	9.09%	16.68
		25%	735.58	684.35	-0.007%	-0.008%	-1.68%	-2.07%	9.31%	9.12%	7.95
SOR	5	5%	364.04	332.89	-0.022%	-0.024%	-5.43%	-5.89%	0.61%	0.55%	9.89
		15%	181.74	156.12	-0.037%	-0.041%	-8.99%	-9.75%	0.77%	0.77%	16.73
		25%	288.03	252.00	-0.027%	-0.030%	-6.64%	-7.33%	9.64%	8.46%	14.75
	22	5%	128.17	116.46	-0.045%	-0.047%	-10.73%	-11.20%	0.75%	0.75%	10.55
		15%	514.13	472.23	-0.015%	-0.016%	-3.61%	-4.06%	7.08%	7.08%	9.39
		25%	555.68	530.58	-0.013%	-0.014%	-3.19%	-3.44%	9.80%	9.77%	5.08
CAR	5	5%	784.00	763.95	-0.005%	-0.006%	-1.33%	-1.48%	56.15%	55.91%	2.87
		15%	1033.48	1014.03	0.001%	0.000%	0.18%	0.08%	78.63%	78.57%	2.10
		25%	718.63	674.70	-0.007%	-0.009%	-1.81%	-2.15%	24.9%	21.1%	0.96
	22	5%	801.10	771.19	-0.005%	-0.006%	-1.22%	-1.42%	56.15%	34.19%	4.20
		15%	545.69	510.78	-0.013%	-0.015%	-3.29%	-3.64%	45.63%	45.15%	7.28
		25%	870.23	867.56	-0.003%	-0.003%	-0.76%	-0.78%	48.61%	48.61%	0.33

Tabella 17: statistiche riepilogative TIT.MI, N=5.

Le performance del titolo TIT.MI sono le peggiori del *pool* selezionato. L'andamento della serie dei prezzi, soprattutto nel secondo periodo temporale, non registra alcun trend rialzista o ribassista e rimane pressoché immutato per tutto il periodo. Nonostante questo, l'algoritmo riporta un'operatività media molto elevata, utilizzando lo Sharpe Ratio ed il Sortino Ratio, la quale si rivela essere, in media, superiore rispetto agli altri titoli analizzati, con picchi anche notevoli ($\# = 20.8$ impostando $N = 1, SR, L = 22, \varepsilon = 15\%$, oppure $\# = 18.6$ impostando $N = 1, SOR, L = 5, \varepsilon = 15\%$). Questo fatto porta alla considerazione che l'algoritmo riceva dei segnali operativi corretti ma non riesca ad avere un buon *market timing* di entrata ed uscita dal mercato. La volatilità della serie è molto bassa, soprattutto nel secondo periodo temporale: questo spinge le funzioni di reward basate sulla deviazione standard come misura di rischio (sia essa totale o solo computata come *downside risk*) ad assumere valori più elevati, dunque a fornire dei *feedback* all'algoritmo non completamente aderenti a quello che è l'ambiente di riferimento. Il Calmar Ratio, in questo caso, risulta una miglior funzione di reward: basata sul massimo drawdown dell'equity line, quest'ultimo funge da una sorta di *stop loss* per l'algoritmo e l'indicazione *stay out of the market*, frutto della natura conservativa di tale indice di performance, risulta essere in questo contesto la più efficace. Le tre performance migliori si ottengono settando $N = 1, SR, L = 5, \varepsilon = 5\%$ (-2.22% il rendimento netto medio annuo, $\# = 9.77$), $N = 1, SOR, L = 22, \varepsilon = 25\%$ (-3.24% , $\# = 7.95$), $N = 5, CAR, L = 22, \varepsilon = 5\%$ (si è scelto di considerare la miglior performance quella ottenuta con una certa operatività minima superiore a $\# = 4$, ed in questo caso il rendimento netto medio annuo è -1.42% con $\# = 4.2$), i grafici sono riportati in *figura 24*. La tendenza generale, utilizzando lo Sharpe ed il Sortino Ratio, è quella di prediligere azioni *greedy* per $L = 5$ (le performance infatti diminuiscono all'aumentare di ε), attitudine che si inverte per $L = 22$: calcolando la volatilità su un periodo temporale più ampio, questa è in grado di valutare in misura più completa l'ambiente e fornire migliori *feedback* all'algoritmo, anche se questi non sono comunque soddisfacenti.

Considerando l'andamento della serie dei prezzi, ed individuata la difficoltà dell'algoritmo ad ottenere performance significative, si è agito sul sistema di scelta dell'azione operativa sbilanciandolo a favore dell'azione $a_t = -1$ nel seguente modo:

- se $-1 \leq \bar{a}_t < -1/6$ allora $a_t = -1$,
- se $-1/6 \leq \bar{a}_t \leq 1/3$ allora $a_t = 0$,
- se $1/3 < \bar{a}_t \leq 1$ allora $a_t = 1$.

In questo caso viene aumentata l'incidenza dell'azione *sell/short position* rispetto all'azione *stay out of the market*, in maniera tale da indirizzare una porzione di queste ultime scelte verso una partecipazione attiva al mercato (*sell/short*) e non più passiva (*stay out*). Il settaggio utilizzato per questo test è $N = 1, k = 500, SR, L = 22, \varepsilon = 25\%$ il quale aveva realizzato un rendimento medio annuo netto pari a -5.16% per un capitale finale netto di 382.82 €. Impostando il nuovo sistema di trading, i risultati sono soddisfacenti (la figura 23, di seguito, riporta i grafici di entrambi i casi).

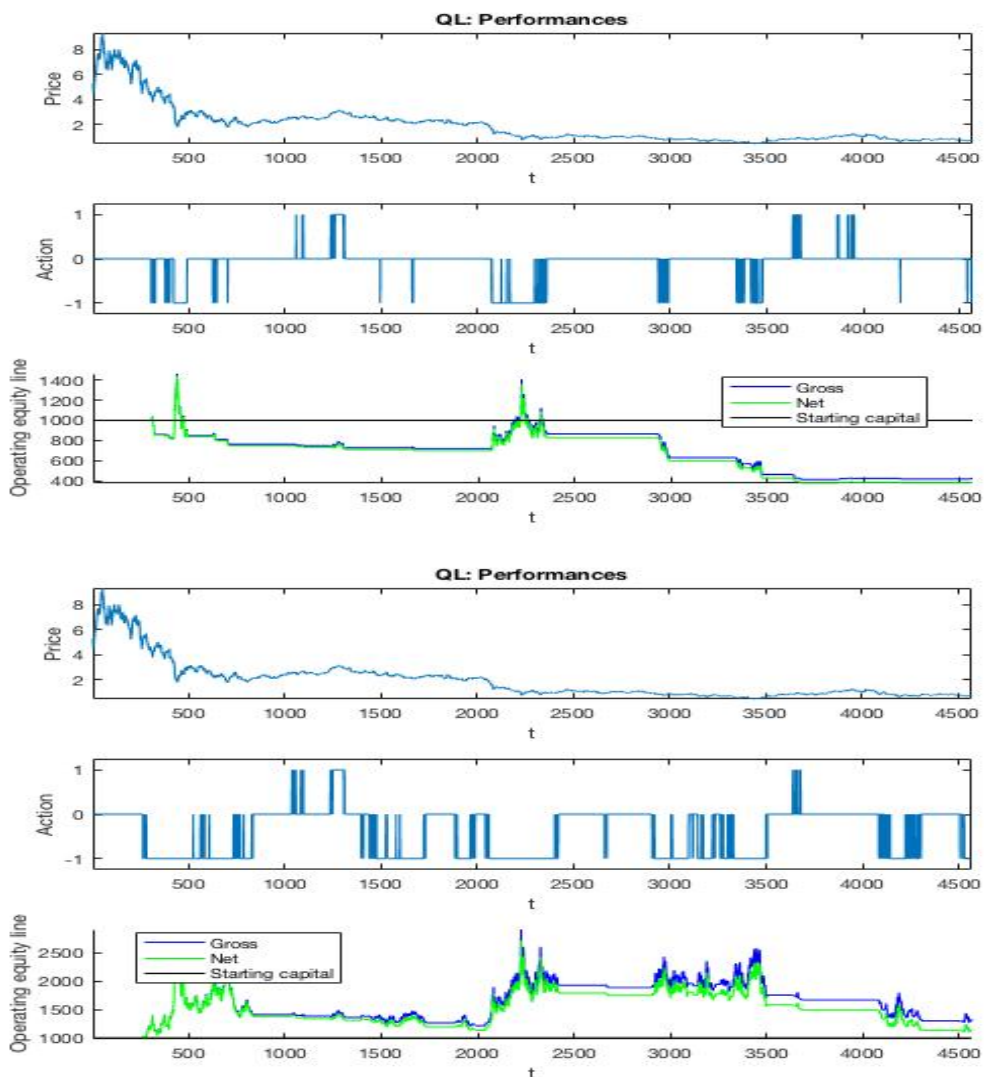


Figura 23: Dall'alto verso il basso:

1. sistema di trading di partenza, titolo TIT.MI con settaggio $k=500, N=1, SR, L=22, \varepsilon=0,25$.
2. nuovo sistema di trading testato, titolo TIT.MI con settaggio $k=500, N=1, SR, L=22, \varepsilon=0,25$.

In termini numerici, il capitale netto finale ammonta a 1154.9 € per un rendimento netto medio annuo di 0.80%, il numero medio di operazioni ammonta a # = 15.73 (fattore che incide sul profitto finale poiché più aumenta il numero di azioni maggiori saranno i costi di transazione), l'equity line netta è superiore al capitale iniziale investito nel 99.82% dell'intervallo temporale. Se tale performance, in termini assoluti, può essere considerata non eccessivamente positiva, se rapportata al precedente risultato indica un netto miglioramento, in termini sia di rendimento netto medio annuo che in termini di capitale finale ottenuto. Tale risultato porta alla considerazione che, adeguando il sistema operativo di selezione dell'azione di trading all'andamento dei prezzi, questo riesca ad essere molto più efficace ed a condurre ad incrementi di performance.

Infine, la *tabella 18* sottolinea, ancora una volta, l'inefficacia dell'algorithm a scovare una strategia di *trading* che risulti profittevole: il Calmar Ratio ottiene percentuali superiori al 50% sia per $N = 1$ che per $N = 5$ ma la sua operatività non risulta essere significativa poiché inferiore a tre azioni medie annue (rispettivamente, # = 2.74 settando $N = 1$ e # = 2.96 settando $N = 5$). Lo Sharpe Ratio ottiene percentuali leggermente migliori del Sortino Ratio, ma in nessun caso supera il 20% (la miglior percentuale è pari al 19.76%, media della numerosità di volte nelle quali l'*equity line* lorda è superiore al capitale iniziale, settando $N = 1$) delineando dunque una valutazione finale negativa per entrambe le misure *risk-adjusted*.

rwd	N=1			N=5		
	perc_over	MEDIE	#	perc_over	MEDIE	#
SR	19.76%	16.45%	13.58	12.93%	9.99%	12.08
SOR	4.30%	4.15%	13.50	4.77%	4.56%	11.06
CAR	51.50%	51.25%	2.74	51.68%	47.26%	2.96
ALL	25.19%	23.95%	9.94	23.13%	20.61%	8.70
ALL - CAR	12.03%	10.30%	13.54	8.85%	7.28%	11.57

Tabella 18: percentuale media equity line > cap. investito per funzione di reward, titolo TIT.MI.

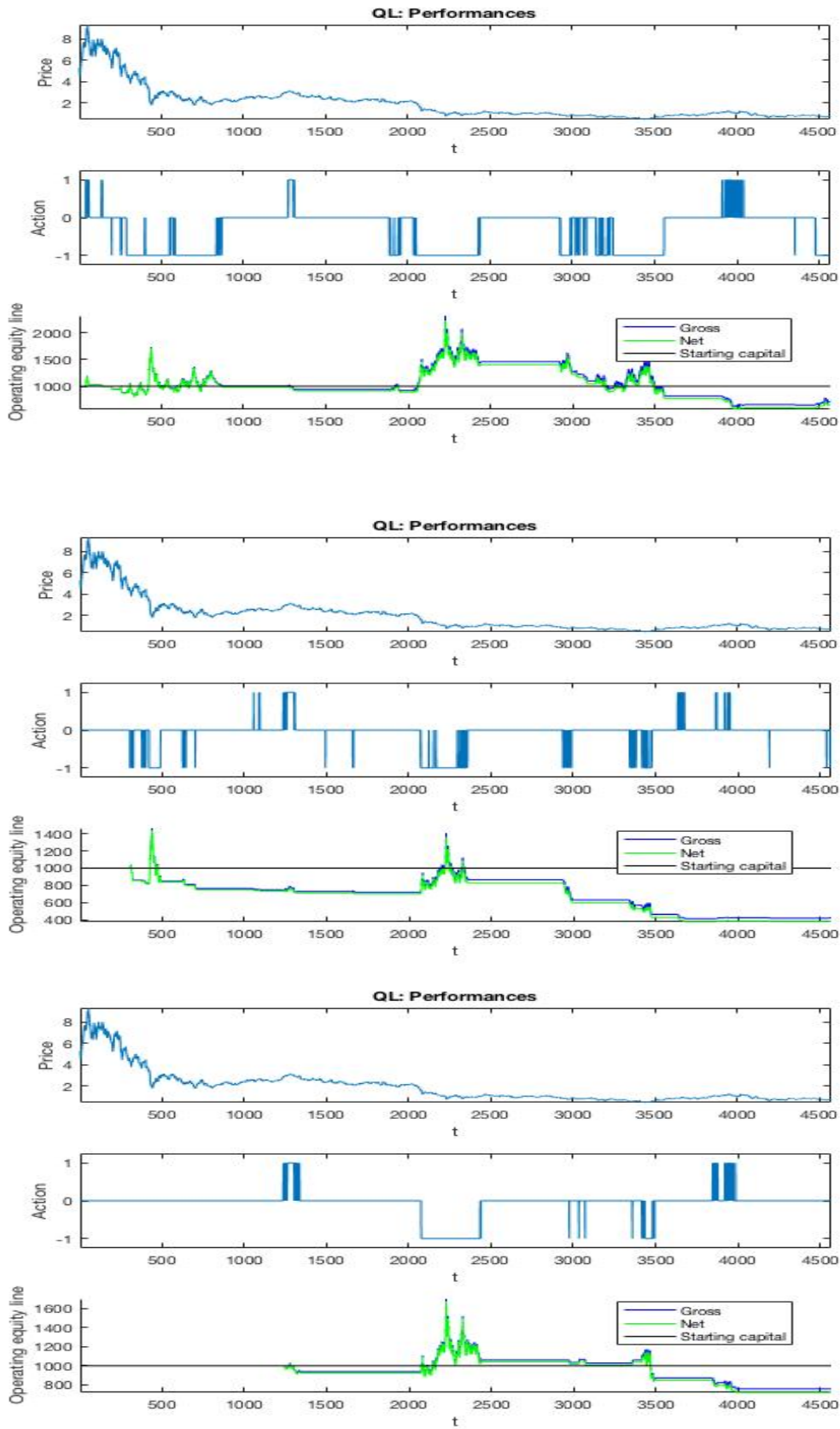


Figura 24: performance titolo TIT.MI con settaggi (rappresentati dall'alto verso il basso):

1. $N=1$, SR , $L=5$, $\varepsilon=0,05$.
2. $N=1$, SOR , $L=22$, $\varepsilon=0,25$.
3. $N=5$, CAR , $L=22$, $\varepsilon=0,05$.

5.2 Considerazioni finali

In questo paragrafo si riportano le analisi finali riguardanti i risultati ottenuti utilizzando l'algoritmo Q-Learning, soffermandosi in particolare su alcune riflessioni e fornendo qualche spunto operativo per eventuali sviluppi futuri. Innanzitutto, dai risultati esposti precedentemente appare chiaro come il titolo FCA.MI sia il titolo meglio sfruttato dall'algoritmo, mentre i titoli TIT.MI e ENI.MI presentano le performance peggiori. Un fattore determinante è sicuramente la *facilità* con la quale l'algoritmo Q-Learning è in grado di apprendere e di fornire efficaci input di trading. La *figura 25* e la *figura 26* riportano un confronto tra le *equity line* ottenute in fase di computazione tra il titolo FCA ed i titoli TIT.MI ed ENI.MI. Se per il titolo FCA.MI le *equity line* seguono dei trend definiti e rialzisti, vi è una dispersione evidente negli altri due casi, orientata per altro verso il ribasso e dunque verso una situazione di sistematica perdita di capitale.

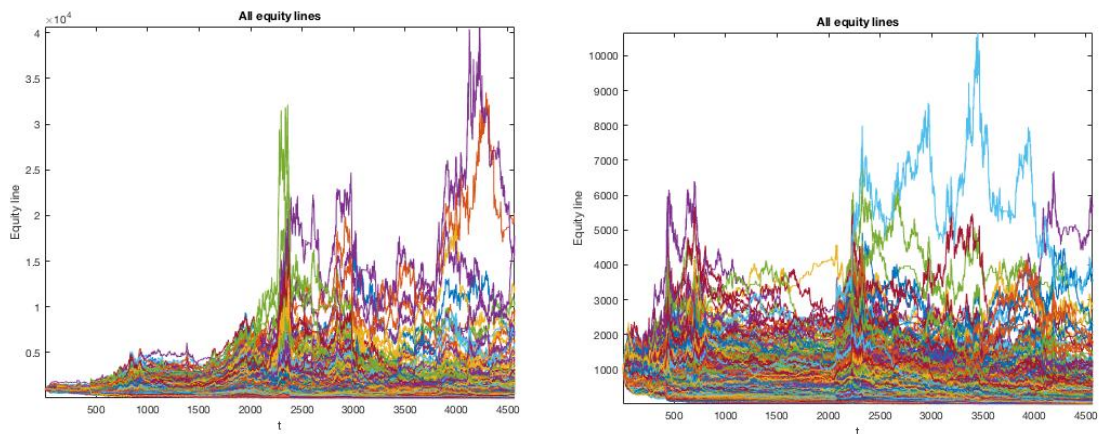


Figura 25: equity line totali, titolo (da sinistra verso destra) FCA.MI e TIT.MI, settaggio $N=1$, SR , $L=22$, $\varepsilon=15\%$.

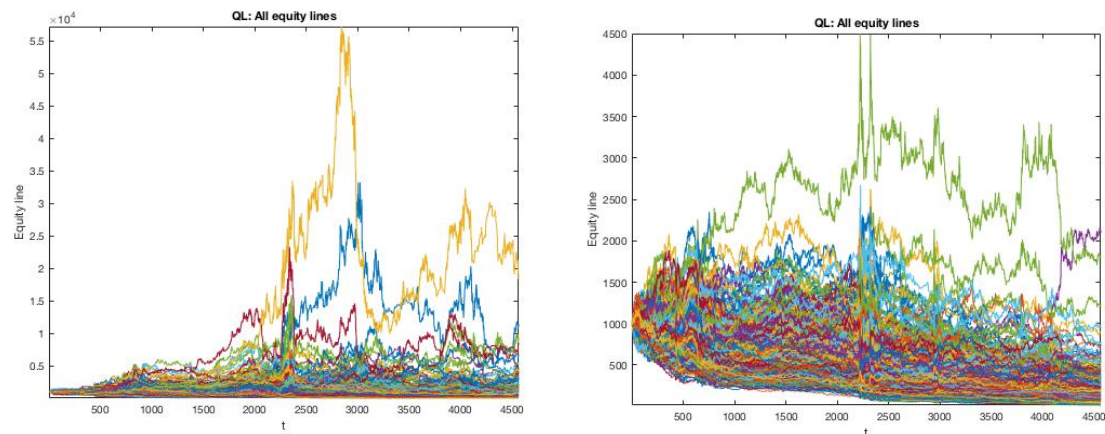


Figura 26: equity line totali, titoli (da sinistra verso destra) FCA.MI ed ENI.MI, settaggio $N=5$, SOR , $L=5$, $\varepsilon=0.25$.

Questo fattore è determinante in fase operativa, quando vengono aggregati in media le indicazioni delle 500 iterazioni e definito il sistema di azioni di trading: non essendovi un segnale operativo ben consolidato e confermato dalla maggior parte delle *equity line* ottenute nelle varie iterazioni, non è garantito che gli intervalli di operatività scelti siano efficaci per quei titoli che presentano caratteristiche simili ad ENI.MI e TIT.MI, ovvero numerose oscillazioni di prezzo in brevi periodi di tempo per il primo, oppure curva dei prezzi che si mantiene pressoché piatta per un lungo periodo per il secondo. Altra evidenza riscontrata nei risultati ottenuti è infatti la difficoltà a cogliere segnali operativi in fasi di mercato nelle quali non è evidente un trend definito ed il prezzo oscilla attorno a valori molto vicini tra loro. Un esempio viene riportato nella seguente *figura 27*, rappresentante le performance ottenute investendo sul titolo G.MI settando $N = 5, SOR, L = 5, \varepsilon = 15\%$.

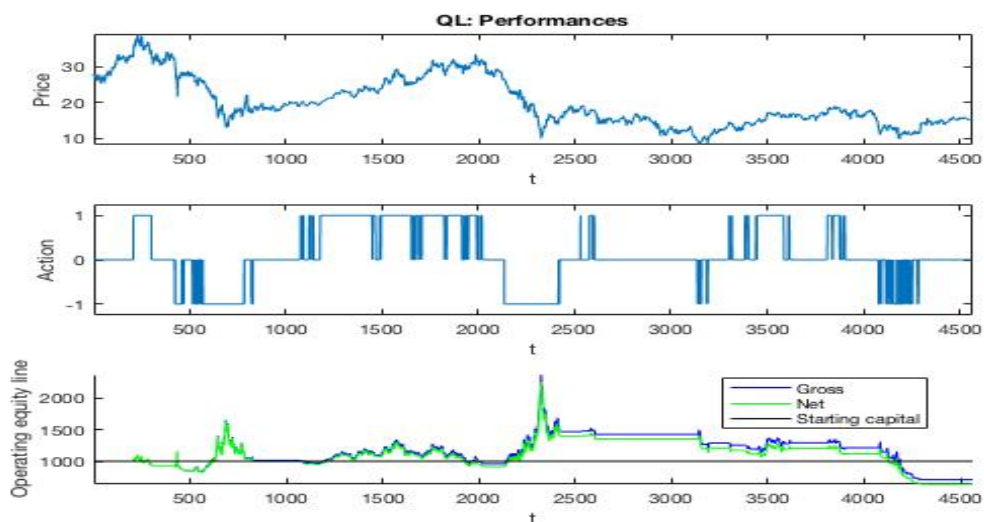


Figura 27: performance titolo G.MI ($N=5, SOR, L=5, \varepsilon=0,15$).

Confrontando l'*equity line* (terzo panel della figura) con la serie dei prezzi (primo panel) si può notare come, soprattutto nei periodi di numerose oscillazioni di prezzo in intorno di valore contenuti, l'algoritmo colga segnali di ingresso ed uscita dal mercato in tempi non sufficientemente rapidi da ottenere un profitto consistente; per questo motivo, l'*equity line*, nell'ultimo periodo, decresce fino a valori inferiori al capitale iniziale. Questo avviene non solo per il titolo G.MI ma anche per il titolo ENI.MI (*figura 28*), il

quale presenta, in generale in tutto l'*holding period* ma soprattutto nella seconda parte, oscillazioni più marcate del titolo G.MI senza assumere un trend definito.

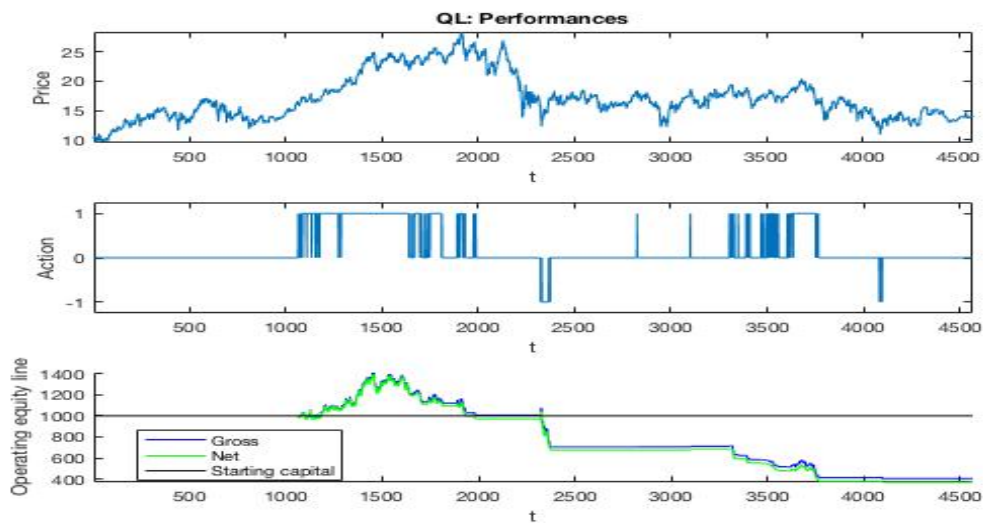


Figura 28: performance titolo ENI.MI ($N=1$, CAR , $L=22$, $\varepsilon=0.05$).

Tali evidenze si traducono nell'individuazione di un problema generale legato alla struttura dell'algoritmo: in presenza di continui break strutturali, i quali provocano continue oscillazioni di prezzi non ingenti ma comunque apprezzabili, oppure di andamenti dei prezzi non chiaramente definiti, l'algoritmo fatica ad individuare un comportamento ottimo da adottare poiché deve repentinamente eliminare il *pattern* comportamentale appreso fino a quel momento e apprenderne uno completamente nuovo, e questo processo può essere necessario un numero tale di volte da rallentare il processo di apprendimento dell'algoritmo e da inficiare il livello finale di performance, poiché vengono intraprese azioni di investimento non conformi allo schema comportamentale richiesto dall'ambiente finanziario. Una possibile soluzione a questo problema potrebbe essere quella di agire modificando il valore del *learning rate* α in maniera tale da rendere più rapido il processo di apprendimento dell'algoritmo, tenendo però in considerazione che, in base alle condizioni di convergenza espresse nel capitolo 3, un valore troppo elevato di α non permetterebbe la convergenza necessaria per ottenere l'ottimo della funzione valore. Tale problema potrebbe essere superato introducendo due *learning rate* costanti, di valore diverso, da applicare a seconda dell'ambiente di riferimento: un *learning rate* di valore sufficientemente basso da garantire la convergenza ed un valore invece più elevato per accelerare il processo di

apprendimento in presenza di ricorrenti break strutturali. Tale ipotesi può configurarsi come un possibile spunto da approfondire in successive verifiche.

Riflessione particolare merita il Calmar Ratio, funzione di reward utilizzata come alternativa allo Sharpe Ratio ed al Sortino Ratio anche per il suo costante impiego, in ambito finanziario, al fine di valutare e monitorare le performance degli investimenti. Tale indice è costruito sul massimo *drawdown* dell'*equity line* nel periodo di valutazione considerato, dunque valuta la media degli ultimi L rendimenti ottenuti rapportata al massimo *drawdown* registrato per quel periodo; se non si registra alcun *drawdown*, ovvero questo è pari a zero, il valore di *reward* per quell'istante t è semplicemente la media degli ultimi L rendimenti ottenuti. Tale struttura è stata impostata al fine di renderla simile a quella utilizzata per le altre due funzioni di *reward* e da permettere un confronto più aderente. Essendovi dunque un *drawdown* solamente quando l'*equity line* decresce nell'intervallo di valori L , l'algoritmo considera *migliore* un *drawdown* pari a zero rispetto ad un *drawdown* positivo, risultato che si ottiene se vi è un'efficace strategia di trading che comporta un crescendo di profitti oppure nel caso in cui si decida di star fuori dal mercato e non investire, decisione che rende *piatta* l'*equity line* e, di conseguenza, spinge il valore di *drawdown* a zero. Probabilmente è questa la causa di scarsa operatività derivante dall'utilizzo del Calmar Ratio come funzione di *reward*, non adeguata a cogliere segnali operativi se non in presenza di trend ben definiti, soprattutto se crescenti. Non a caso, infatti, tale funzione di reward risulta discretamente efficace, in termini di operatività e performance, quando viene utilizzata sul titolo FCA.MI, il cui andamento dei prezzi è di più facile *lettura* e vi sono opportunità operative che anche tale funzione riesce ad individuare. L'inadeguatezza di tale funzione la si coglie, come già citato, soprattutto da un punto di vista operativo: visionando la *tabella* che, oltre ad aggregare la numerosità media totale (di tutti i titoli analizzati), espressa in percentuale, di volte nelle quali $eq_{line} > C$, raggruppata per settaggi utilizzati e per funzione di reward, riporta l'operatività media totale per ogni ratio, si nota che, mentre l'operatività media per lo Sharpe Ratio ed il Sortino Ratio è simile, il Calmar Ratio presenta una diminuzione sostanziale di # non rendendo i risultati ottenuti con quest'ultimo confrontabili, sulla stessa scala di valutazione degli altri due ratio.

PERC_OVER MEDIE						
	SHARPE		SORTINO		CALMAR	
	G %	N %	G%	N%	G%	N%
N=1	52.78%	49.53%	45.54%	40.60%	59.79%	55.98%
N=5	53.01%	50.33%	43.85%	41.08%	64.88%	62.45%
L=5	46.79%	43.76%	35.09%	32.14%	60.56%	57.84%
L=22	59.01%	56.10%	54.30%	49.55%	67.62%	64.00%
$\varepsilon=0,05$	45.73%	43.79%	38.16%	33.47%	58.63%	53.94%
$\varepsilon=0,15$	52.17%	47.72%	51.58%	47.86%	66.51%	62.20%
$\varepsilon=0,25$	60.79%	58.27%	44.34%	41.20%	61.86%	61.50%
#	10.19		10.35		3.93	

Tabella 19: numerosità percentuali medie totali di volte nelle quali equity line $> C$, raggruppate per settaggio impiegato e per funzione di reward.

Analizzando i risultati in tabella, si possono trarre le seguenti conclusioni riepilogative:

- In generale, l'utilizzo del vettore s_t considerando $N = 1$ ed $N = 5$ produce gli stessi risultati in termini di efficacia del sistema di trading: i risultati sono discreti (attorno al 50%), con una preferenza per lo Sharpe rispetto al Sortino. Essendovi solo un titolo del pool considerato che presenta percentuali superiori al 90% (FCA.MI, *tabella 20*), chiaramente la media totale viene penalizzata dai restanti titoli le cui statistiche non sono sufficientemente elevate (in particolare il titolo TIT.MI, *tabella 21*). Considerando i rendimenti effettivamente realizzati, dalle tabelle dei singoli titoli presentate nelle pagine precedenti, nonché dalle due tabelle di seguito, si nota una tendenza generale a migliorare le performance con $N = 5$ per quei titoli i cui rendimenti non sono pienamente soddisfacenti (G.MI, ENI.MI, TIT.MI) rispetto a quei titoli i cui rendimenti sono mediamente buoni o molto positivi (UCG.MI, FCA.MI), sintomo questo che per titoli *difficili* da apprendere per l'algoritmo, mediamente rappresentare uno stato s_t più numeroso è di supporto maggiore per raggiungere risultati migliori, in special modo utilizzando il Sortino Ratio rispetto allo Sharpe Ratio. Questo conferma quanto asserito dalla AMH, ovvero che qualora vi siano opportunità di profitto per determinati titoli, queste vengono sfruttate dagli agenti che operano nel mercato ed aumentare la numerosità non comporta miglioramenti di performance;

qualora invece le opportunità di profitto siano più complicate da sfruttare, un maggior profitto si ottiene ampliando il range temporale di valutazione ($N = 5$).

FCA.MI												
	PERC_OVER						REND GIORN MEDIO					
	SHARPE		SORTINO		CALMAR		SHARPE		SORTINO		CALMAR	
	G %	N %	G%	N%	G%	N%	G %	N %	G%	N%	G%	N%
N=1	90.24%	89.70%	92.34%	91.87%	93.12%	92.76%	4.11%	3.65%	4.09%	3.58%	1.72%	1.49%
N=5	89.68%	88.85%	83.70%	81.32%	68.20%	68.05%	3.25%	2.81%	3.37%	2.87%	0.57%	0.35%
L=5	89.57%	88.50%	85.01%	83.82%	77.45%	77.33%	4.25%	3.78%	3.74%	3.25%	0.76%	0.58%
L=22	90.35%	90.05%	91.04%	89.37%	82.48%	82.14%	3.11%	2.68%	3.72%	3.21%	1.53%	1.26%
$\varepsilon=0,05$	79.96%	79.56%	79.09%	76.37%	79.16%	78.95%	1.39%	1.15%	1.32%	0.97%	1.36%	1.12%
$\varepsilon=0,15$	93.54%	92.16%	96.19%	96.12%	94.50%	94.08%	3.50%	3.00%	5.46%	5.00%	1.90%	1.60%
$\varepsilon=0,25$	96.38%	96.11%	88.78%	87.29%	68.33%	68.18%	6.15%	5.54%	4.41%	3.72%	0.18%	0.04%

Tabella 21: statistiche riepilogative titolo FCA.MI, aggregate per singolo settaggio e per funzione di reward.

TIT.MI												
	PERC_OVER						REND GIORN MEDIO					
	SHARPE		SORTINO		CALMAR		SHARPE		SORTINO		CALMAR	
	G %	N %	G%	N%	G%	N%	G %	N %	G%	N%	G%	N%
N=1	19.76%	16.45%	4.30%	4.15%	51.50%	51.25%	-5.49%	-6.13%	-5.99%	-6.62%	-1.01%	-1.15%
N=5	12.93%	9.99%	4.77%	4.56%	51.68%	47.26%	-4.61%	-5.18%	-6.43%	-6.94%	-1.37%	-1.56%
L=5	20.44%	14.66%	2.75%	2.40%	46.96%	46.13%	-4.43%	-5.08%	-7.06%	-7.76%	-1.37%	-1.54%
L=22	12.26%	11.78%	6.32%	6.32%	59.26%	55.36%	-5.67%	-6.23%	-5.37%	-5.80%	-1.01%	-1.17%
$\varepsilon=0,05$	25.51%	21.56%	1.08%	0.98%	49.36%	43.62%	-4.11%	-4.50%	-6.96%	-7.46%	-1.77%	-1.98%
$\varepsilon=0,15$	15.82%	10.61%	4.86%	4.74%	62.69%	62.49%	-6.16%	-6.94%	-6.80%	-7.48%	-1.08%	-1.26%
$\varepsilon=0,25$	7.71%	7.49%	7.66%	7.34%	42.71%	41.67%	-4.88%	-5.51%	-4.87%	-5.41%	-0.72%	-0.82%

Tabella 20: statistiche riepilogative titolo TIT.MI, aggregate per singolo settaggio e per funzione di reward.

- Un'altra tendenza individuata è la propensione generale, in termini di efficacia, per un settaggio $L = 22$ rispetto ad $L = 5$, poiché con il primo si realizzano percentuali consistenti (56% Sharpe, 49.6% Sortino, 64% Calmar le percentuali medie delle *equity line* nette) e maggiori rispetto al secondo. Questo avviene anche in termini di rendimento medio annuo netto. Una valutazione *risk-adjusted* delle azioni intraprese nel mercato è, quindi, più efficace se analizzata su un periodo più lungo rispetto ad un periodo più breve.
- Infine, valutando i diversi settaggi di ε , si osserva che, per lo Sharpe Ratio, vi è un miglioramento di performance all'aumentare dell'incidenza dell'azione esplorativa rispetto all'azione *greedy*, fenomeno che si osserva anche utilizzando il Sortino ed il Calmar Ratio, i quali però presentano un miglioramento di performance settando $\varepsilon = 15\%$ e non $\varepsilon = 25\%$, con il quale invece si ottengono risultati inferiori, ma comunque maggiori del caso $\varepsilon = 5\%$. Il giusto trade off tra

exploration ed *exploitation* è dunque variabile a seconda della funzione di reward utilizzata. Un'ulteriore valutazione dunque potrebbe essere effettuata diminuendo l'intervallo di variazione tra i tre valori di ε , in particolare andando a testare, per il Sortino Ratio e per il Calmar Ratio, con quale valore di questo parametro vengono raggiunte le massime performance, così come andare a verificare, per lo Sharpe Ratio, quanto è possibile aumentare il valore di ε , rispetto al 25%, per migliorare i già buoni risultati ottenuti.

Infine, merita una particolare considerazione finale anche il processo operativo di *clusterizzazione* attuato per ottenere un unico segnale operativo, per istante temporale, dalle k iterazioni effettuate. Data la natura del metodo per approssimazione lineare ed il fatto che il vettore θ venga inizialmente settato tramite un processo casuale, è necessario un numero elevato di iterazioni per ottenere dei segnali operativi consistenti e poiché, nella realtà finanziaria, difficilmente si può disporre di k capitali iniziali per rendere operativa ogni singola iterazione; in questo modo, i segnali operativi vengono ottenuti in *media*. Un limite a tale processo si evidenzia quando le capacità dell'algoritmo di apprendere l'ambiente finanziario non sono elevate e gli intervalli operativi possono non rivelarsi efficienti per quel particolare intervallo temporale o per quel particolare titolo. Una possibile soluzione potrebbe essere quella di adeguare gli intervalli di selezione al particolare andamento dei prezzi in atto, mediante uno o più indicatori di analisi tecnica che supportino l'individuazione di una particolare tendenza in atto, nonché la sua relativa forza o debolezza, in maniera tale da orientare le scelte di *trading* anche secondo questi indicatori. Questo non metterebbe in secondo piano il processo di selezione dell'algoritmo, che sarebbe comunque centrale, ma orienterebbe gli intervalli in range più consoni al particolare momento di mercato. Un primo *rudimentale* esempio, basato sull'indicazione proveniente dall'incrocio di determinate medie mobili, potrebbe essere il seguente:

- Calcolo medie mobili utilizzando tre diverse *rolling windows*, ad esempio di numerosità $mavg(5), mavg(22), mavg(45)$ (ovvero circa una settimana, un mese ed un trimestre di prezzi di mercato giornalieri).
 - se $mavg_t(5) > mavg_t(22) > mavg_t(45)$, utilizzare il sistema

- $-1 \leq \bar{a}_t < -1/3, a_t = -1$
- $-1/3 \leq \bar{a}_t \leq 1/6, a_t = 0$
- $1/6 < \bar{a}_t \leq 1, a_t = 1$
- se $mavg_t(5) < mavg_t(22) < mavg_t(45)$, utilizzare il sistema
 - $-1 \leq \bar{a}_t < -1/6, a_t = -1$
 - $-1/6 \leq \bar{a}_t \leq 1/3, a_t = 0$
 - $1/3 < \bar{a}_t \leq 1, a_t = 1$
- nei casi restanti
 - $-1 \leq \bar{a}_t < -1/3, a_t = -1$
 - $-1/3 \leq \bar{a}_t \leq 1/3, a_t = 0$
 - $1/3 < \bar{a}_t \leq 1, a_t = 1$

In questo modo, se vengono individuati determinati trend di prezzo, viene sbilanciato il sistema di selezione verso una posizione *long* o *short*, tenendo comunque come punto di riferimento la media risultate dalle k iterazioni svolte. Il sistema qui presentato è un primo possibile spunto di implementazione e sicuramente può essere ottenuta una maggior efficacia modificando il valore degli intervalli, rendendoli più o meno sbilanciati a seconda della situazione di mercato, o applicando simultaneamente diversi indicatori di analisi tecnica per comprovare eventuali situazioni di trend in atto.

CONCLUSIONI

In questo elaborato di tesi è stato sviluppato, in ambiente *Matlab*®, un algoritmo basato sul metodo Q-Learning, rientrante nel gruppo di metodi ad apprendimento per rinforzo, al fine di applicarlo in ambito finanziario come sistema di trading. Basandosi sulla struttura di alcuni elaborati di ricerca raccolti nel corso delle indagini bibliografiche di questo lavoro, il principale intento alla base di tale elaborato era fornire un possibile metodo operativo che desse valenza pratica all'emergente teoria dei mercati adattivi presentata da Andrew Lo, la quale riconcilia la classica teoria dei mercati efficienti di Eugene Fama con una moderna ottica evolucionistica secondo la quale il comportamento che i soggetti economici assumono, relazionandosi con l'ambiente nel quale agiscono, è guidato dai principi alla base della teoria dell'evoluzione, ovvero competizione, adattamento, selezione naturale. Secondo questo innovativo *framework*, l'individuo apprende dall'ambiente circostante il comportamento più adatto che gli permette di *sopravvivere* ed *evolversi*, ricevendo dallo stesso ambiente *rinforzi informativi* utili a correggerlo, nonché *ricompense*, sotto varie e possibili forme, qualora questo risulti essere efficace ed adeguato al contesto. In ambito finanziario, dunque, potrebbe essere proficuo *creare* un agente artificiale che apprenda celermente i *rinforzi* e le *ricompense* fornite dall'ambiente circostante così da sviluppare una strategia di *trading* adattiva che fornisca come output principale le indicazioni migliori sull'operatività da assumere per gestire un ammontare di capitale investito.

L'algoritmo Q-Learning, in questo elaborato, ha assunto il ruolo di agente artificiale di cui si faceva riferimento poc'anzi, ed è stato applicato ad un *pool* di cinque titoli reali del mercato italiano (UniCredit S.p.A., Assicurazioni Generali, Eni S.p.A., Fiat Chrysler Automobiles NV, Telecom Italia S.p.A.) e ne sono stati raccolti, esposti e commentati i

principali risultati. Sono stati applicati diversi settaggi, in particolare tre diverse funzioni di *reward risk-adjusted*: lo Sharpe Ratio, largamente utilizzato nella letteratura empirica di riferimento, il Sortino Ratio ed il Calmar Ratio, due misure invece raramente utilizzate, delle quali si è verificata l'efficacia o meno rispetto allo Sharpe Ratio. Sono state testate 36 diverse possibili configurazioni ed i risultati ottenuti sono vari: estremamente positivi per alcuni titoli, negativi per altri. Uno dei fattori caratterizzanti è l'utilizzo di parametri comuni per tutti i titoli selezionati, i quali possono essere accumulati da alcuni fattori ma anche presentare alcune peculiarità e diversità che possono inficiare alcune performance finali. Sono state comunque individuate delle tendenze generali di comportamento dell'algorithm e sono state ipotizzate alcune possibili miglorie (presentate al paragrafo 2 del quinto capitolo) per quelle situazioni nelle quali non si sono ottenute performance totalmente positive. Ad ogni modo, i risultati finali sono mediamente soddisfacenti.

Il confronto tra le tre diverse funzioni di *reward* ha portato ad una valutazione generalmente positiva per l'utilizzo dello Sharpe Ratio, misura già testata in numerosi articoli specialistici di settore, così come per il Sortino Ratio, il quale si configura essere, sotto determinate condizioni, una valida alternativa allo Sharpe. Non si ottiene un giudizio positivo, al contrario, riguardo il Calmar Ratio: come si è avuto modo di approfondire, tale indice non è stato valutato positivamente e, nonostante sia una misura di monitoraggio delle performance ampiamente utilizzata in ambito finanziario, in questo contesto non risulta essere efficace. Tuttavia, non si esclude a priori la possibilità di sviluppare ulteriori e diverse funzioni di *reward* basate sul *drawdown* come misura di rischio. Inoltre, una possibile modifica operativa al Sortino Ratio potrebbe essere quella di considerare il *downside risk* come volatilità tra i rendimenti realizzati nel periodo di valutazione ed un *benchmark* di riferimento, sia questo un rendimento target selezionato dall'investitore, un rendimento *risk free* di determinati titoli di stato o semplicemente il rendimento di un indice selezionato come riferimento. Questa idea conclusiva, assieme ad alcuni suggerimenti presentati nel capitolo 5, può essere uno spunto per eventuali ed ulteriori implementazioni future.

Bibliografia

- Ball, R. (1989). *What do we know about stock market "efficiency"?* Managerial Economics Research Center. Rochester: William E. Simon Graduate School of Business Administration.
- Barto, A. (2007). *Temporal Difference Learning*. Tratto da Scholarpedia: www.scholarpedia.org/article/Temporal_difference_learning.
- Barto, A., Sutton, R. (2016). *Reinforcement Learning: An Introduction II ed*. Londra: The MIT Press.
- Barucci, E. (2000). *Teoria dei mercati finanziari: equilibrio, efficienza, informazione*. Bologna: Il Mulino.
- Bertsekas, D.P., Tsitsiklis, J.N. (1996). *Neuro-Dynamic Programming*. Belmont: Athena Scientific.
- Corazza, M., Bertoluzzo, F. (2015). *Intelligent Q-Learning for financial trading system*. Working Papers - Dipartimento di Economia Ca' Foscari (Venezia).
- Corazza, M., Sangalli, A. (2015). *Q-Learning and SARSA: a comparison between two intelligent stochastic approach for financial trading*. Working Papers - Dipartimento di Economia Ca' Foscari (Venezia).
- Corazza, M., Bertoluzzo, F. (2012). *Reinforcement Learning for automatic financial trading: Introduction and some applications*. *Procedia Economics and Finance*, vol. 3, pagg. 68-77.
- Elton, E., Gruber, M., Brown, S., Goetzmann, W. (2014). *Modern portfolio theory and investments analysis*. 9ed, Wiley.
- Fama, E. (1969). *Efficient Capital Markets: A Review of Theory and Empirical Work*. *The Journal of Finance*, vol. 25 num. 2, pagg. 383-417.
- Fama, E. (1965). *Random Walks in Stock Market Prices*. *Financial Analyst Journal*, pagg. 75-80.
- Fiorini, A. (2015). *Storia del trading online*. Milano: Mediosfera.

- Gosavi, A. (2003). *Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning*. Norwell: Kluwer Academic Publishers.
- KPMG. (2017). *Digital Banking, Le sfide per il sistema bancario tra esigenze dei consumatori, rivoluzione digitale e nuovi competitors*. KPMG Advisory.
- Linciano, N., Gentile, M., Soccorso, P. (2017). *Report on financial investments of Italian households, behavioural attitudes and approaches*. Roma: Consob.
- Lo, A. (2012). *Adaptive Markets and the New World Order*. Financial Analysts Journal, Vol. 68 num. 2, pagg. 18-29.
- Lo, A. (2007). *Efficient Market Hypothesis*. In Blume, L., Durlauf, S., *The new Plagiarist: A Dictionary of Economics II ed.* New York: Palgrave MacMillan.
- Lo, A. (2005). *Reconciling Efficient Markets with Behavioral Finance: The Adaptive Market Hypothesis*. Journal of Investment Consulting, vol. 7 num. 2, pagg. 21-44.
- Lo, A. (2004). *The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective*. The Journal of Portfolio Management, pagg. 1-31.
- Moody J., Wu L., Liao Y., Saffell M. (1998). *Performance Functions and Reinforcement Learning for Trading Systems and Portfolios*. Journal of Forecasting, vol. 17, pagg. 441-470.
- Pederson, C. S., Rudholm-Alfvén, T. (2003). *Selecting a Risk-Adjusted Shareholder Performance Measure*. Journal of Asset Management, vol. 4 num. 3, pagg. 152-172.
- Samuelson, P. (1965). *Proof That Properly Anticipated Prices Fluctuate Randomly*. Industrial Management Review, vol. 6, pagg. 41-49.
- Stiglitz, J. (1981). *Pareto Optimality and Competition*. The Journal of Finance, vol. 36 num. 2, pagg. 235-238.