

**POLITECNICO DI TORINO**

**Corso di Laurea Magistrale in Ingegneria Gestionale**



**Politecnico  
di Torino**

**TESI DI LAUREA MAGISTRALE**

**Il mercato delle *commodities* applicato all'Intelligenza  
artificiale**

**Relatore: Franco Varetto**

**Candidato: Riccardo Arnese**

**ANNO ACCADEMICO 2020-2021**

# INDICE

INTRODUZIONE .....	4
1. Il mercato delle <i>commodities</i> .....	6
1.1. INVESTIMENTI NEL MERCATO DELLE MATERIE PRIME .....	6
1.2. I PRINCIPALI MERCATI DELLE <i>COMMODITIES</i> .....	7
1.3. RELAZIONE TRA SPOT E FUTURE PRICE.....	8
1.3.1. Basis .....	10
1.3.2. Theory of storage-H.Working (1933).....	11
1.3.3. Rendimento totale.....	12
1.4. FATTORI CHE INFLUENZANO IL MERCATO DELLE COMMODITY.....	12
2. Analisi di cinque <i>commodities</i> .....	14
2.1. IL COTONE.....	14
2.1.1. Fattori che guidano il prezzo del Cotone.....	18
2.1.2. Mercati di riferimento per il Cotone.....	18
2.2.1. Fattori che influenzano Il prezzo dell'oro .....	23
2.2.2. Mercati di riferimento dell'oro.....	28
2.3. IL RAME.....	29
2.3.1. Fattori che influenzano il prezzo del Rame.....	32
2.3.2. Principali mercati del rame .....	33
2.4. LA SOIA.....	34
2.4.1. Fattori che influenzano il prezzo della soia .....	35
2.4.2. Mercato di riferimento per la Soia.....	35
2.5. PETROLIO GREGGIO.....	37
2.5.1. Riserve di petrolio.....	37
2.5.2. Le diverse tipologie di Petrolio Greggio .....	38
2.5.3. Fattori che determinano il prezzo del Petrolio.....	39
3. Analisi dell'efficienza delle quotazioni delle <i>commodities</i> tramite i Random Walks .....	44
3.1. RANDOM WALK.....	46
3.2. Variance Ratio Test.....	47
3.3. Risultati variance ratio test sul mercato delle <i>commodities</i> .....	50
3.3.1. Regressione lineare e indice $R^2$ .....	51
4. Reti neurali e Deep Learning.....	62
4.1. Architettura di una rete neurale .....	63
4.2. Gradiente discendente e algoritmo di Backpropagation.....	63
4.2.1. Assunzioni e vincoli dell'algoritmo di Backpropagation.....	65

4.2.2.	<b>Algoritmo di Backpropagation</b> .....	67
4.2.3.	<b>Funzione di costo Cross-entropy</b> .....	69
4.3.	<b>OVERFITTING</b> .....	71
5.	<b>LSTM applicata al prezzo spot delle <i>commodities</i></b> .....	73
5.1.	<b>LSTM per il forecasting dei prezzi delle <i>commodities</i></b> .....	76
5.1.1.	<b>Pandas</b> .....	77
5.1.2.	<b>Numpy</b> .....	78
5.1.3.	<b>Keras e Tensorflow</b> .....	78
5.2.	<b>FUNZIONE DI ATTIVAZIONE</b> .....	78
5.3.	<b>ALGORITMO DI OTTIMIZZAZIONE</b> .....	79
5.4.	<b>LEARNING RATE</b> .....	80
6.	<b>Risultati forecasting <i>commodities</i></b> .....	81
6.1.	<b>COMMODITY ORO</b> .....	83
6.2.	<b>COMMODITY RAME</b> .....	88
6.3.	<b>COMMODITY COTONE</b> .....	93
6.4.	<b>COMMODITY SOIA</b> .....	98
6.5.	<b>COMMODITY PETROLIO</b> .....	103
	<b>CONCLUSIONE</b> .....	108
	<b>BIBLIOGRAFIA</b> .....	110
	<b>SITOGRAFIA</b> .....	111

## INTRODUZIONE

Il presente elaborato tratta lo studio e l'analisi del mercato delle materie prime applicato all'intelligenza artificiale. L'importanza crescente rivestita dalle *commodities* nel commercio, attraverso lo scambio fisico dei beni, e in ambito finanziario, attraverso la negoziazione di strumenti finanziari come i future sulle materie prime, ha portato a compiere un'analisi sia qualitativa, che quantitativa del settore.

In prima battuta si è analizzato il mercato delle commodities in generale, spiegando il meccanismo che regola la relazione tra i prezzi spot e future e fornendo una panoramica dei macro-fattori che influenzano la formazione dei prezzi.

Successivamente sono state selezionate cinque commodities appartenenti a categorie diverse:

- il **cotone**, categoria delle soft *commodities*
- l'**oro**, categoria dei metalli preziosi
- il **petrolio**, commodity di tipo energetico
- il **rame**, metallo non prezioso
- la **soia**, categoria delle materie prime alimentari

Nel secondo capitolo, per ciascuna di queste materie prime sono stati approfonditi il mercato di riferimento a livello mondiale e i vari campi di applicazione. Oltre a ciò, sulla base dei dati forniti da un report del primo trimestre del 2021 sulle *commodities* della Banca Centrale mondiale, sono stati analizzati i dati commerciali su un range temporale di 50 anni. Questo ha permesso di avere una panoramica completa sull'evolversi del mercato per ciascuna commodity.

Nel terzo capitolo è stata valutata l'efficienza del mercato in forma debole attraverso il Variance Ratio Test. Lo scopo di questa sezione è stato quello di verificare l'assenza di correlazione tra i log rendimenti lungo archi temporali differenti.

Oltre al test effettuato sulla varianza, per verificare l'efficienza del mercato e la casualità nei movimenti dei prezzi è stato valutato l'indice  $R^2$  a valle della regressione lineare delle quotazioni negli ultimi 5 anni.

La seconda parte del presente elaborato tratta il mondo delle Reti Neurali e del Deep Learning.

Nel quarto capitolo si è analizzato il comportamento matematico dei neuroni all'interno di una rete neurale e successivamente si è studiato il funzionamento dell'algoritmo di Backpropagation.

Nel capitolo quinto, dopo avere individuato il modello LSTM e approfondito il suo comportamento nello studio delle serie temporali, sono state descritte le librerie del linguaggio di programmazione Python in grado di elaborare i dati secondo le logiche *Long Short Term Memory*.

Infine, nell'ultimo capitolo sono stati riportati ed esaminati i risultati in termini di capacità predittiva derivati dall'allenamento della rete neurale, confrontando i risultati di diverse architetture in termini di errore assoluto e percentuale.

# 1. Il mercato delle *commodities*

Con il termine Commodity ci si riferisce alle materie prime, intese come beni fungibili, i quali hanno la caratteristica di poter essere acquistabili sul mercato senza differenze qualitative, indipendentemente da chi li produce.

Le *commodities* rappresentano un gruppo molto eterogeneo di beni e sono generalmente raggruppate in diversi settori, quali energia, metalli industriali, metalli preziosi e prodotti agricoli.

Le materie prime energetiche includono diversi tipi di petrolio greggio e i suoi raffinamenti, come il gasolio e l'olio combustibile.

I metalli industriali sono inglobati nella produzione di varie categorie di merci, come ad esempio il rame e i minerali di ferro.

I metalli preziosi sono rari a causa del loro valore intrinseco, ma possono essere utilizzati per scopi industriali, come ad esempio il Palladio e il Platino.

Infine, le commodity agricole includono tutti i prodotti legati all'alimentazione, come il grano e il frumento.

## 1.1. INVESTIMENTI NEL MERCATO DELLE MATERIE PRIME

La crescente importanza rivestita dalle commodity nel mercato finanziario ha portato allo sviluppo di numerosi strumenti che consentono di investire direttamente o indirettamente nelle materie prime.

Di seguito sono riportate le principali modalità d'investimento:

- Acquistare direttamente il bene reale. Questa soluzione è la meno utilizzata dagli investitori finanziari in quanto la deperibilità e le diverse modalità di stoccaggio dei beni portano a rendere lo scambio condizionato al possesso di idonee strutture di immagazzinamento.
- Acquistare azioni di società che producono beni reali. In questo caso ci si espone al rischio specifico legato ad una determinata commodity. Oltre al rischio del mercato, il prezzo delle azioni riflette talvolta la gestione societaria.

- Acquistare un portafoglio di azioni di imprese che operano nel mercato delle materie prime. A differenza dello strumento precedente, in questo caso si ricorre al servizio offerto da investitori specializzati nel mercato (ad es. i *commodities trading advisors* o **CTA**). Di conseguenza l'investimento sarà diversificato.
  
- Negoziare Future sulle materie prime. Questi strumenti finanziari sono negoziati in borsa e possono essere utilizzati ad esempio dalle imprese che producono le commodity per tutelarsi dalle incertezze sulle quotazioni. Questa categoria di investimenti è quella più utilizzata degli investitori finanziari. Nello specifico, prevede che due controparti firmino un contratto per lo scambio di un sottostante ad una scadenza pattuita precedentemente. Generalmente, solo una residua parte di questi contratti si chiude con l'effettivo scambio del sottostante, mentre la maggior parte dei contratti viene rinegoziata oppure chiusa da contratto con posizione opposta.
  
- Investire passivamente in *Exchange Traded Commodities* (ETC). Gli ETC sono strumenti finanziari emessi a fronte dell'investimento diretto dell'emittente in materie prime o derivati sulle materie prima. Questi strumenti replicano passivamente la performance di una singola commodity o di indici di commodity.

## **1.2. I PRINCIPALI MERCATI DELLE *COMMODITIES***

Di seguito sono elencati i principali mercati in cui sono scambiate le materie prime:

In Nord America i più importanti sono il Chicago Mercantile Exchange (CME), il New York Mercantile Exchange (NYMEX), il Commodity Exchange (COMEX), il Chicago Board of Trade (CBOT) e Toronto Stock Exchange (TSX). Il CME, il CBOT, il NYMEX ed il COMEX fanno parte di CME Group, fondato nel 1848 a Chicago.

Sul CME sono scambiati un ampio range di prodotti: future (e opzioni) sulle commodity, fino ai derivati su indicatori economici. Invece, sul New York Mercantile Exchange (NYMEX) sono negoziati il WTI Crude Oil, il gas naturale, il propano, l'oro, il platino, il palladio, l'argento. Fa parte del gruppo CME dal 2006;

Infine il COMEX, entrato a far parte del gruppo CME nel 2008, è la piazza dove vengono scambiati

metalli industriali e preziosi;

Sul CBOT sono negoziate le commodity agricole come il mais, l'avena, il riso, la soia, la farina di soia, il frumento, l'etanolo. Dal 2007 CBOT fa parte del gruppo CME;

Infine, il Toronto Stock Exchange (TSX) è la principale piazza di scambio canadese in termini di materie prime, dove sono negoziati prevalentemente titoli minerari ed energetici.

Nel mercato Asiatico troviamo il Dalian Commodity Exchange (DCE), in Cina, molto importante per lo scambio di future nel settore agricolo e il Multi Commodity Exchange (MCX) in cui viene scambiato il petrolio, gas naturale, semi/ granaglie e metalli industriali/ preziosi.

In Europa il mercato più importante è l'Intercontinental Exchange (ICE), con sede a Londra, dove vengono negoziate le commodity agricole come il caffè, il cotone, lo zucchero e le commodity energetiche come il Brent crude, il WTI crude, l'energia elettrica.

Sempre nel mercato londinese il London Metal Exchange (LME) è considerato il più importante mercato per i metalli tra cui il rame, il piombo, lo zinco, l'acciaio, lo stagno, l'alluminio, il nickel e il cobalto.

### **1.3. RELAZIONE TRA SPOT E FUTURE PRICE**

In questa sezione è descritto il meccanismo che regola la relazione tra il prezzo spot e future.

I rendimenti spot sono ottenuti dalla vendita istantanea di un bene fisico ad una certa data, mentre i future sulle commodity sono degli asset finanziari che affidano agli acquirenti e venditori la consegna di un certo bene ad un prezzo futuro stabilito a priori.

Con l'avvicinarsi della scadenza, il prezzo future della commodity converge al prezzo spot, portando ad ottenere una performance della posizione finanziaria.

Il meccanismo di arbitraggio tende a far convergere il prezzo indicato nel contratto future verso il prezzo spot man mano che si avvicina il mese di scadenza del titolo future. Qualora il prezzo future fosse maggiore del prezzo spot gli arbitraggisti potrebbero vendere il contratto future, acquistare sul mercato spot il sottostante e consegnarlo al possessore del future; viceversa, se il prezzo del future fosse minore di quello spot, gli arbitraggisti potrebbero acquistare il contratto future, farsi consegnare il sottostante e venderlo sul mercato al prezzo spot. Gli arbitraggi proseguirebbero fino a quando gli



effetti della domanda e dell'offerta sui prezzi dei future e/o sui prezzi spot colmano la differenza, tenuto conto dei costi di transazione per le operazioni di compravendita.

La speculazione con i future può creare situazioni paradossali in cui sono sottoscritti contratti lunghi per un ammontare molto superiore alla quantità di beni disponibili per la consegna: gli investitori in possesso di queste posizioni lunghe non accettano di chiudere i contratti alla scadenza per mettere in difficoltà chi detiene le corrispondenti posizioni corte; questi ultimi non sono in grado di far fronte alle proprie obbligazioni e per non diventare insolventi sono disposti a pagare prezzi molto elevati pur di chiudere la loro posizione; il risultato è un forte aumento delle quotazioni dei future e quelle spot.

La differenza tra prezzi spot e future definisce la pendenza della curva degli stessi future. La struttura della curva delle commodities, a differenza di quella degli *stock index futures* che presentano una struttura a termine in salita, può avere una pendenza sia negativa che positiva a seconda di come si muovono domanda e offerta della commodity in esame.

Una pendenza crescente della curva future implica che i *buyers*, ovvero gli investitori, pagheranno più per i future che per il bene fisico stesso. Qualora il prezzo spot della commodity rimanesse inalterato fino alla scadenza del future, il buyer avrà perso un rendimento proporzionale alla pendenza iniziale della curva del future. Di conseguenza, a seconda di come si muove il mercato si possono generare dei guadagni e delle perdite derivate dall'evoluzione dei prezzi future.

La maggior parte degli investitori, i quali non hanno particolare interesse a detenere fisicamente una commodity, comprano dei titoli future e li rivendono pochi giorni prima della loro scadenza rinnovando la loro posizione lunga sulla commodity acquistano future con una data di scadenza più lontana nel tempo.

Come menzionato precedentemente, i derivati finanziari obbligano il venditore a consegnare una certa quantità del sottostante ad un acquirente ad una certa data (*maturity date*), fissando in anticipo il prezzo di acquisto (*strike price*).

Una volta ottenuto il titolo, lo stesso si può scambiare con altri investitori al prezzo di mercato.

Nel caso delle opzioni, invece, a seconda che siano delle opzioni di tipo CALL o PUT, il venditore può decidere rispettivamente se comprare o vendere il sottostante.

### 1.3.1. Basis

L'equazione sotto riportata mostra il rendimento generato dall'acquisto di un future sottoscritto su una certa commodity con scadenza in  $T$ , mantenuto fino alla sua scadenza, ipotizzando che il prezzo spot della commodity rimanga invariato.

$$\frac{S(t) - F(t, T)}{F(t, T)} \times \frac{365}{T - t}$$

$S(t)$  è il prezzo spot della commodity al tempo  $t$  e  $F(t, T)$  è il prezzo future al tempo  $t$  che scade dopo  $T$  istanti di tempo. Il fattore  $\frac{365}{T-t}$  è un fattore di annualizzazione.

Il termine tecnico per questo rendimento è *basis* o *roll yield*.

Il Basis quantifica la perdita dell'investitore nel caso in cui i prezzi di mercato rimangano invariati. Se il termine è positivo, il mercato di quella commodity avrà una curva della struttura a termine con pendenza negativa. In questo caso il mercato è nello stato di Backwardation. Nel caso opposto il mercato sarà nello stato Contango, e quindi la curva avrà una pendenza positiva.

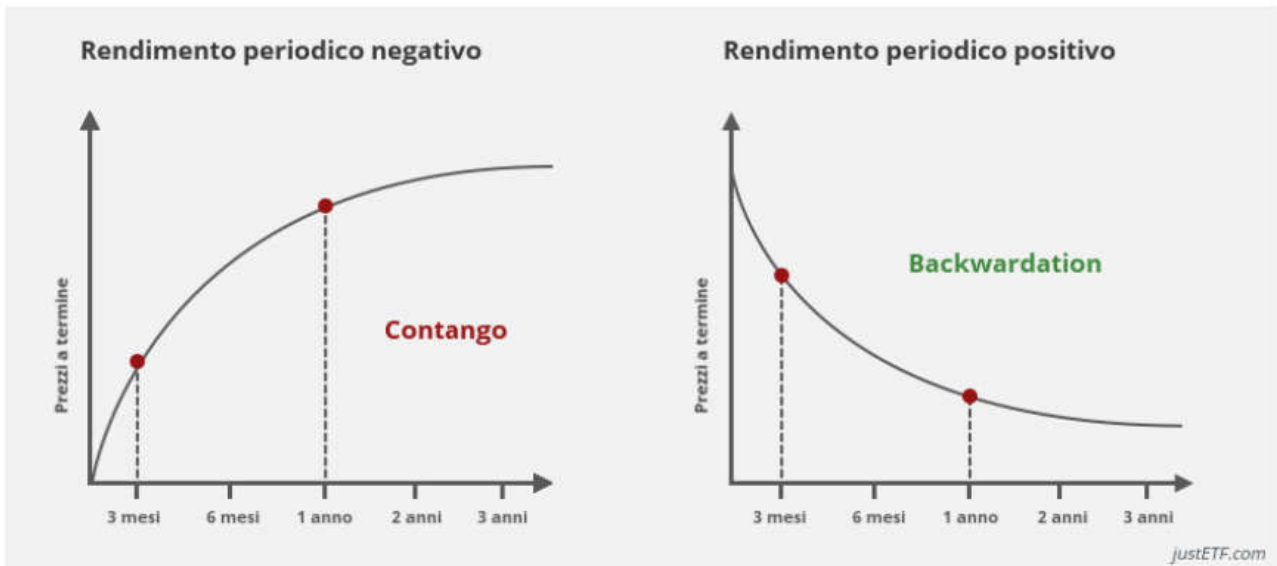


Figura 1. Contango vs Backwardation

Fonte: justetf.com

Di conseguenza l'indice è positivo in un mercato “*backwarded*” e negativo in un mercato “*contangoed*”.

I produttori delle materie prime possono godere del beneficio di fissare in anticipo i prezzi future. Fissare un prezzo future inferiore al prezzo spot, porta gli operatori finanziari che sono avversi al rischio, ad acquistare questi strumenti derivati. In questo caso la pendenza della struttura a termine dei future sarà negativa.

Quando si presenta un accumulo di una commodity da parte dei produttori, significa che si è assunta una posizione lunga su quella materia prima. Venendo parte dei contratti future, i produttori riducono la loro posizione lunga portando i prezzi dei future ad un valore inferiore a quello dei prezzi attuali.

In questo caso si parla di “*normal backwardation*” e suggerisce una struttura a termine dei prezzi future negativa. Questo fenomeno è influenzato dai livelli di stock della commodity esaminata.

Qualora i livelli di magazzino fossero bassi, il prezzo future convergerà al prezzo spot. Al contrario, livelli elevati di magazzino portano i produttori a preferire la domanda per contratti future per coprirsi dal rischio, con conseguente differenza negativa tra il prezzo future e spot.

In quest'ultimo caso, i future si prestano a costruire delle operazioni di copertura (hedge) per proteggersi da rischi di mercato: si assumono posizioni lunghe o corte in future per neutralizzare rischi latenti su attività da acquistare o da vendere.

Tuttavia, empiricamente è stato dimostrato che per la maggior parte delle commodity non si può avere uno stato di *backwardation* permanente.

### **1.3.2. Theory of storage-H.Working (1933)**

Secondo Working, mantenere un certo livello di scorte di materie prime, dà la possibilità al mercato di assorbire una domanda incerta. Il beneficio implicito generato viene chiamato come *convenience yield*, correlato positivamente con i livelli di inventario. Il beneficio marginale diminuisce progressivamente con l'aumentare delle scorte.

Questa teoria spiega che la pendenza della curva dei prezzi future dipende dalla differenza tra la capacità di assorbire la domanda e i costi di immagazzinamento.

Ricapitolando, il prezzo future di una commodity è influenzato contemporaneamente dal tasso d'interesse risk free, dai costi di stoccaggio e dal *convenience yield*. Questi ultimi due fattori sono considerati proporzionali al valore del sottostante. Queste relazioni possono essere racchiuse dalla seguente relazione:

$$F(t, T) = S(t) \times e^{(r+m-y)T}$$

dove  $r$  rappresenta il tasso d'interesse risk free,  $m$  i costi di stoccaggio e  $y$  il convenience yield.

### 1.3.3. Rendimento totale

Fino ad ora non si è considerata la variazione del prezzo spot durante l'intera vita del titolo future.

Il rendimento totale di un investitore in possesso di un future con scadenza  $T$  sottoscritto su una certa commodity, venduto pochi giorni prima di  $T$  è espresso dalla formula sottostante

$$\frac{S(T)}{S(t)} \times \frac{S(t)}{F(t, T)} - 1$$

La prima parte rappresenta la variazione temporale dello spot price, mentre la seconda parte è il rolling yield nel paragrafo precedente.

## 1.4. FATTORI CHE INFLUENZANO IL MERCATO DELLE COMMODITY

In questo paragrafo si analizzano i fattori generali che influenzano le variazioni dei prezzi delle materie prime. Tuttavia, occorre tenere a mente che ciascuna particolare commodity ha dei fattori specifici che ne influenzano i suoi rendimenti nel tempo. Ad esempio le condizioni climatiche influenzano l'offerta di grano, mentre le tecnologie di estrazione quella di petrolio.

I paesi in fase di espansione, come ad esempio la Cina, hanno un bisogno sempre più crescente di materie prime per alimentare la propria crescita in molteplici settori, dalla produzione industriale, per cui sono richieste ingenti quantità di metalli, all'alimentazione per la quale è alta la richiesta di soft *commodities*.

Un altro elemento fondamentale è la quantità di risorse disponibili, relazionata alla domanda per un certo bene. L'incontro tra la domanda e l'offerta influenza l'andamento del prezzo. Ad esempio, in periodi in cui vi è una carenza di scorte di materia prima per cui la domanda è costante, si può assistere a delle impennate del prezzo della stessa materia prima. In alcuni casi, specialmente nel mercato delle

*commodities*, si cerca di trovare dei beni capaci di sostituire la materia prima in questione. La conseguenza di questo fenomeno è la creazione di una concorrenza implicita, la quale porta alla ristabilizzazione dei prezzi sul mercato.

Prendendo in esempio la domanda di metalli industriali e oro, la prima è collegata al livello di produzione industriale, mentre la seconda dipende dalla percezione del mercato finanziario in termini di rischio e inflazione. In luce di questo, parlare generalmente della domanda del mercato come un'unica grande realtà appare riduttivo, dato che ciascun settore è influenzato da fattori idiosincratici.

Un'altra componente fondamentale nella formazione dei prezzi delle *commodities* è la forza del dollaro americano nell'economia. Dato che la maggior parte delle materie prime è quotata in dollari sui rispettivi mercati, un dollaro debole incentiva l'acquisto in quanto a parità di potere d'acquisto si può acquistare una maggior quantità di materia prima. Viceversa, un dollaro forte disincentiva lo scambio dei beni fungibili, perché i venditori “guadagnano meno dollari” in momenti di predominio del dollaro.

Nel mercato azionario e obbligazionario il rischio degli investitori viene ripagato rispettivamente attraverso l'emissione di dividendi e cedole, in base alle diverse tipologie di titoli che si possono detenere. Per contro, le *commodities* non generano nessuno ricavo diretto. Infatti, per questo mercato il premio per il rischio deriva da altre fonti, come i titoli future, i quali svincolano la valutazione delle materie prime rispetto al loro valore spot.

In conclusione, le *commodities* rivestono una rilevante importanza finanziaria nell'economia mondiale. Basti pensare alla forte correlazione tra il mercato azionario e quello delle commodity durante la crisi del 2008; questa ha portato vari ricercatori ad analizzare se questo fenomeno fosse imputabile alla speculazione attorno al mercato delle *commodities*, oppure se uno shock di domanda dovuto alla crisi. Un altro esempio a cui far riferimento è come lo shock dei prezzi del petrolio negli anni 1979 e 1981 abbia influenzato la crescita economica e l'inflazione.

## 2. Analisi di cinque *commodities*

In questo capitolo verranno analizzate le caratteristiche e gli aspetti macroeconomici di cinque commodity, scelte per la loro diversa categoria di appartenenza.

Le *commodities* prese in esame sono:

- il **cotone**, appartenente alla categoria della soft *commodities*, è ampiamente utilizzato industrialmente nel settore tessile.
- l'**oro**, considerato come la commodity per eccellenza e appartiene alla categoria dei metalli preziosi
- il **petrolio**, divenuta la principale fonte energetica, è fondamentale per il funzionamento dell'economia mondiale
- il **rame**, metallo utilizzato in molteplici applicazioni industriali ed edili
- la **soia**, una delle principali commodity del settore alimentare.

### 2.1. IL COTONE

Il cotone è da sempre largamente utilizzato come materia prima per l'industria tessile. L'invenzione della sgranatrice di cotone nel 1793, la quale separa le fibre della pianta di cotone dal resto della pianta, ha permesso una rapida crescita della produzione industriale di questo materiale su larga scala. Tuttavia, dopo la II guerra mondiale, l'avanzamento di fibre sintetiche artificiali come il poliestere e il nylon, ha sottratto al cotone una quota del mercato tessile. Come parametro di confronto, dal 1960 al 2002, il consumo di fibre sintetiche artificiali è cresciuto al tasso del 4,7%, mentre quello del cotone al tasso annuo del 1,8%.

Il prezzo del cotone, come la maggior parte delle altre *commodities*, viene fissato dall'incontro tra domanda e offerta.

La perdita di quote di mercato è stata compensata dalla crescita della popolazione, stabilizzando così la domanda di questa materia prima (Figura 3).

Una ricerca del 2004 della Banca mondiale nell'ambito dello Sviluppo e del commercio, evidenzia come le quote di mercato del cotone rispetto al complessivo consumo di fibre si sia eroso dal 1960, passando da quasi il 70% ad appena il 40% nel 2002 (Figura 2).

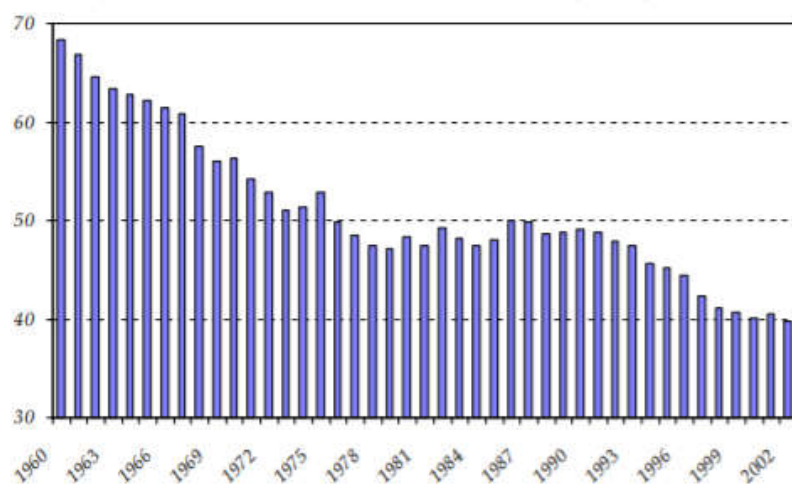


Figura 2. Quota di mercato percentuale del cotone sul totale consumo delle fibre tessili  
 Fonte: World Bank Policy Research Working Paper 3218, February 2004, John Baffes

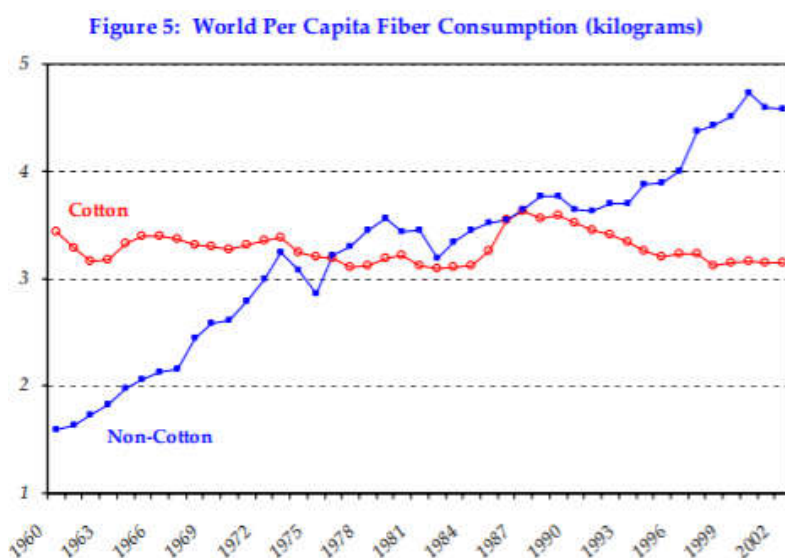


Figura 3 Consumo mondiale pro-capite di cotone in chilogrammi

Fonte: World Bank Policy Research Working Paper 3218, February 2004, John Baffes

La tabella n.1 riporta i dati, in migliaia di tonnellate relativi alla produzione, alle scorte, alle esportazioni e importazioni di cotone annue con cadenza decennale dal 1970 ad oggi, con un focus più dettagliato dagli anni dal 2017/18 sino al 2021.

Dai dati si evince che si è verificata una costante crescita nella produzione di cotone dal 1970 al 2010 (i volumi sono più che raddoppiati).

Dal 2010 al 2021 invece i volumi prodotti si sono mantenuti pressoché costanti, con qualche lieve oscillazione, complice la crisi del Coronavirus che ha rallentato il settore tessile e della moda.

Approfondendo l'analisi, si nota che nel 1970 il leader della produzione mondiale erano gli Stati Uniti, mentre la forte crescita del tessuto produttivo di Cina e India, ha portato alla creazione di una nuova gerarchia a partire dal 2010.

Le ragioni di questa inversione di leadership sono da ricercare nella forte espansione industriale che ha reso protagonista Cina e India nel settore della fabbricazione tessile.

Infatti, per la propria crescente industria tessile, la Cina è al primo posto in termini di volumi di cotone importato, seguita da Vietnam e Bangladesh; contrariamente, la Cina non si colloca nemmeno tra i principali paesi che esportano la materia prima in questione, primato detenuto dagli Stati Uniti (3,375 t/anno), seguiti dal Brasile (1,742t/anno).



Tabella 1. Produzione, Scorte, Importazioni ed Esportazione di Cotone in migliaia di tonnellate dal 1970 al 2021

Source: International Cotton Advisory Committee (April 15, 2021 update)

	1970/71	1980/81	1990/91	2000/01	2010/11	2017/18	2018/19	2019/20	2020/21
<b>Production</b>									
India	909	1,322	1,989	2,380	5,865	6,350	5,661	6,205	6,307
China	1,995	2,707	4,508	4,505	6,600	5,890	6,040	5,800	5,910
United States	2,219	2,422	3,376	3,742	3,942	4,555	3,999	4,336	3,201
Brazil	594	623	717	939	1,960	2,006	2,779	3,002	2,523
Pakistan	543	714	1,638	1,816	1,948	1,795	1,670	1,320	890
Turkey	400	500	655	880	817	882	977	815	656
Australia	19	99	433	804	926	1,058	485	134	562
Uzbekistan	n/a	1,671	1,593	975	910	800	641	716	552
Greece	110	115	213	421	180	258	277	355	321
Benin	14	6	59	141	60	257	295	311	294
Turkmenistan	n/a	n/a	437	187	385	304	283	283	289
<i>Others</i>	4,936	3,652	3,334	2,738	2,276	2,643	2,865	3,067	2,605
<b>World</b>	<b>11,740</b>	<b>13,831</b>	<b>18,951</b>	<b>19,527</b>	<b>25,869</b>	<b>26,798</b>	<b>25,972</b>	<b>26,344</b>	<b>24,110</b>
<b>Stocks</b>									
China	412	476	1,589	3,755	2,167	9,033	8,885	8,938	8,938
India	376	491	539	922	1,886	1,989	1,878	3,430	3,430
Brazil	321	391	231	755	1,400	1,598	2,340	2,787	2,959
Turkey	24	112	150	292	542	1,064	1,115	1,373	1,373
United States	915	581	510	1,306	566	819	826	1,314	640
Bangladesh	n/a	14	37	50	176	422	422	458	475
Argentina	65	61	123	114	232	347	320	437	437
<i>Others</i>	2,492	3,024	3,581	3,429	3,534	3,510	2,778	2,624	2,687
<b>World</b>	<b>4,605</b>	<b>5,151</b>	<b>6,761</b>	<b>10,622</b>	<b>10,503</b>	<b>18,782</b>	<b>18,564</b>	<b>21,361</b>	<b>20,939</b>
<b>Exports</b>									
United States	848	1,290	1,697	1,467	3,130	3,639	3,365	3,381	3,375
Brazil	220	21	167	68	435	909	1,310	1,946	1,742
India	34	140	255	24	1,086	1,132	765	696	1,144
Greece	72	13	86	270	147	234	295	319	352
Benin	14	8	58	140	64	196	292	224	304
Australia	4	53	329	849	545	852	791	295	260
Burkina Faso	9	22	73	112	136	236	200	154	256
<i>Others</i>	2,675	2,866	2,403	2,866	2,090	2,057	2,238	2,008	2,053
<b>World</b>	<b>3,875</b>	<b>4,414</b>	<b>5,069</b>	<b>5,797</b>	<b>7,634</b>	<b>9,256</b>	<b>9,256</b>	<b>9,023</b>	<b>9,486</b>

<b>Imports</b>									
China	108	773	480	52	2,609	1,320	2,100	1,554	2,220
Vietnam	33	40	31	84	350	1,521	1,510	1,459	1,482
Bangladesh	n/a	45	80	248	896	1,671	1,544	1,374	1,383
Pakistan	1	1	0	101	314	608	406	555	1,103
Turkey	1	n/a	46	381	729	956	786	1,017	961
Indonesia	36	106	324	570	471	762	685	547	602
India	155	n/a	n/a	350	87	365	392	496	287
<i>Others</i>	<i>3,862</i>	<i>4,363</i>	<i>4,739</i>	<i>4,031</i>	<i>4,957</i>	<i>3,118</i>	<i>3,731</i>	<i>2,810</i>	<i>3,668</i>
<b>World</b>	<b>4,086</b>	<b>4,555</b>	<b>5,220</b>	<b>5,764</b>	<b>7,804</b>	<b>9,001</b>	<b>9,054</b>	<b>8,258</b>	<b>9,486</b>

### **2.1.1. Fattori che guidano il prezzo del Cotone**

Il fattore principale che guida la formazione del prezzo del cotone è la domanda dell'industria tessile per l'abbigliamento a livello mondiale, a sua volta influenzato dai 'trend' e dalla moda.

Come precedentemente menzionato, una componente che influenza la domanda per il cotone, è l'apparizione e la popolarità di nuove materie prime tessili (nylon, cotone).

La produzione di cotone e di conseguenza l'offerta, è stanziata in paesi che alternano periodi di siccità e umidità, condizioni favorevoli alla crescita della pianta.

È facile pensare, quindi, che le condizioni metereologiche influenzano la variazione del prezzo del cotone, in quanto hanno un impatto considerevole sulla qualità dei raccolti e di conseguenza sull'utilizzo per scopi specifici di questa materia prima, con successiva ripercussione sul mercato di impiego.

Come ultimo fattore, si può considerare l'attenzione del governo statunitense a favorire la produzione di cotone attraverso aiuto e sovvenzioni, grazie ad associazioni come il Cotton Council International (CCI), ramo del National Cotton Council of America (NCC), che promuovono l'esportazione e collaborano con cotonifici, produttori di tessuti e con il dipartimento dell'agricoltura degli Stati Uniti.

### **2.1.2. Mercati di riferimento per il Cotone**

Esistono due parametri che garantiscono l'attendibilità del prezzo del cotone di cotone: l'indice 'A' del Cotton Outlook pubblicato a Liverpool e quello dei 'futures' di New York.

L 'Indice 'A' di Liverpool viene pubblicato dall'organismo britannico Cotton Outlook ed è basato sulla media dei prezzi di cinque dei tredici cotoni più economici di diversi paesi, dimensionata in centesimi di dollaro la libbra (0.45 kg).

I mercati per lo scambio spot della materia prima si trovano presso i grandi porti di importazione (Liverpool, Brema, Le Havre, Tokio, Shanghai) e il prezzo viene stabilito prendendo come riferimento l'Indice 'A' di Liverpool.

Un percorso diverso è seguito dal cotone statunitense (contrattato alla borsa di New York) e quello egiziano di fibra lunga, scambiato ad Alessandria d'Egitto.

In questi mercati la maggior parte del cotone si negozia tramite contratti future, mentre l'acquisto della materia prima spot viene utilizzato come strumento di copertura.

## 2.2. ORO

L'oro è il metallo prezioso identificato dal simbolo della tavola periodica 'Au' (numero atomico 79) e rientra tra i metalli nobili data la sua scarsa reattività chimica. Proprio per questo le sue proprietà chimico-fisiche rimangono inalterate se entra a contatto con l'umidità, l'ossigeno e calore. Gli unici composti chimici con cui reagisce sono l'acqua regia e lo ione di cianuro.

L'oro è presente in natura in forma di scagliette, granuli oppure in agglomerati più grandi, le pepite. I giacimenti auriferi sono classificabili in primari e secondari.

I giacimenti primari sono di natura idrotermale e sono contenuti all'interno di rocce ignee denominate "vene" o "filoni". I principali giacimenti primari sono quelli del Canada (Ontario), del Brasile (Minas Geraes), dell'Australia occidentale e degli Stati Uniti (Colorado e California). Anche sulle Alpi Italiane sono presenti alcuni esigui giacimenti.

I depositi secondari sono di natura alluvionale e derivano dall'erosione di quelli primari. I più importanti giacimenti alluvionali sono quelli sudafricani del Witwatersrand; altri sono presenti in America Settentrionale (Alaska, California), Russia (Urali e Siberia), Brasile, Australia, Messico e Colombia. Tuttavia, tra il 2005 e il 2010, la Cina ha scalato la classifica dei maggiori produttori d'oro piazzandosi al primo posto.

Le doti di essere un metallo duttile e malleabile, unite alla sua bassa reattività chimica, hanno consentito a questa materia prima di essere impiegata in diverse applicazioni industriali.

Oltre ad essere utilizzato comunemente a scopo ornamentale per l'industria dei gioielli, l'oro svolge un'importante funzione nella componentistica elettronica nelle parti critiche dei computer e degli apparati di telecomunicazione, proprio per la sua capacità di non corrodersi.

Inoltre, l'oro viene impiegato nell'industria aeronautica per la realizzazione dei visori delle tute spaziali e nell'ambito medico e odontoiatrico con svariati utilizzi.

A livello finanziario, storicamente il sistema economico era basato sul gold standard, sistema in cui il valore di una valuta corrisponde ad una certa quantità d'oro mantenuta come riserva da ciascun Stato.

Dopo le guerre mondiali e dopo un periodo di transizione governato dal sistema gold exchange standard, in cui si era stabilita a livello internazionale la convertibilità in oro del solo dollaro americano, nel 1971 cessò la convertibilità dollaro-oro e il prezzo dell'oro fu lasciato ai meccanismi regolamentati dal mercato.

Oggi giorno le Banche Centrali e i governi mondiali detengono ancora molte riserve d'oro a garanzia del valore delle proprie valute, che corrispondono soltanto ad una piccola percentuale del valore economico circolante, risultato ottenuto a valle del radicale cambiamento avvenuto nel sistema monetario.

La tabella n.6 riporta i dati aggiornati a Marzo 2021, relativi alle riserve di oro detenute dai Governi e dalla Banche Centrali.

I dati, consultabili e scaricabili dal sito del World Gold Council (WGC), associazione industriale delle principali aziende minerarie aurifere, sono estratti dall'ente International Financial Statistic (IFS) del fondo monetario internazionale (IFM). Un altro dato analizzato dal WGC è la percentuale di riserve auree rispetto al totale delle riserve di ciascun governo, parametrizzata al prezzo trimestrale fine mese pubblicato dall'IntercontinentalExchange (ICE) giornalmente. Il prezzo impiegato è quello stabilito dal London Bullion Market Association (LBMA), mercato di riferimento per la quotazione dell'oro.

Tabella 2 Riserve Auree detenute dai governi mondiali – Fonte World Gold Council

<b>WORLD OFFICIAL GOLD HOLDINGS</b>							
International Financial Statistics, April 2021*							
	Country	Tonnes	%of reserves**		Country	Tonnes	%of reserves**
1	United States	8,133.5	77.6%	51	Finland	49.1	20.7%
2	Germany	3,362.4	74.2%	52	Cambodia	45.5	11.8%
3	IMF	2,814.0	<sup>1)</sup>	53	Jordan	43.5	15.4%
4	Italy	2,451.8	69.0%	54	Bolivia	42.5	51.2%
5	France	2,436.3	64.3%	55	Bulgaria	40.8	6.6%
6	Russian Federation	2,295.4	21.8%	56	Malaysia	38.9	1.9%
7	China,P.R.: Mainland	1,948.3	3.2%	57	WAEMU <sup>3)</sup>	36.5	9.3%
8	Switzerland	1,040.0	5.3%	58	Serbia, Rep. of	35.9	12.0%
9	Japan	846.0	3.4%	59	Slovak Rep.	31.7	18.9%
10	India	695.3	6.6%	60	Ukraine	26.1	5.3%
11	Netherlands, The	612.5	66.8%	61	Syrian Arab Republic	25.8	7.8%
12	Turkey <sup>5)</sup>	512.6	37.7%	62	Morocco	22.1	3.5%
13	ECB	504.8	35.1%	63	Ecuador	21.9	20.2%
14	Taiwan Province of China	423.6	4.1%	64	Afghanistan, Islamic Rep. of	21.9	12.7%
15	Kazakhstan, Rep. of	395.8	64.0%	65	Nigeria	21.5	3.1%
16	Portugal	382.6	77.8%	66	Kyrgyz Rep.	16.8	34.7%
17	Uzbekistan, Rep. of	355.8	59.1%	67	Bangladesh	14.0	1.7%
18	Saudi Arabia	323.1	3.8%	68	Cyprus	13.9	66.8%
19	United Kingdom	310.3	9.9%	69	Curaçao and Sint Maarten	13.1	30.5%
20	Lebanon	286.8	38.4%	70	Mauritius	12.4	9.2%
21	Spain	281.6	19.6%	71	Czech Rep.	9.8	0.3%
22	Austria	280.0	54.9%	72	Ghana	8.7	6.3%
23	Poland, Rep. of	228.7	7.6%	73	Mongolia	8.2	9.9%
24	Belgium	227.4	39.9%	74	Paraguay	8.2	4.6%
25	Algeria	173.6	16.6%	75	Tajikistan, Rep. of	8.0	23.7%
26	Philippines	163.9	8.5%	76	Myanmar	7.3	5.2%
27	Venezuela, Republica Bolivariana de	161.2	81.9%	77	Guatemala	6.9	2.0%
28	Thailand	154.0	3.4%	78	North Macedonia, Republic of	6.9	8.1%

29	Singapore	127.4	1.8%	79	Tunisia	6.8	4.1%
30	Sweden	125.7	12.6%	80	Sri Lanka	6.7	7.6%
31	South Africa	125.3	12.9%	81	Latvia	6.7	6.8%
32	Mexico	119.9	3.3%	82	Nepal	6.4	3.0%
33	Libya	116.6	7.9%	83	Ireland	6.0	4.2%
34	Greece	113.9	58.2%	84	Lithuania	5.8	6.1%
35	Korea, Rep. of	104.4	1.3%	85	Colombia	4.7	0.4%
36	Romania	103.6	11.9%	86	Bahrain, Kingdom of	4.7	10.2%
37	BIS <sup>2)</sup>	102.0	<sup>1)</sup>	87	Brunei Darussalam	4.5	6.8%
38	Iraq	96.4	8.1%	88	Mozambique, Rep. of	3.9	5.3%
39	Hungary	94.5	13.7%	89	Slovenia, Rep. of	3.2	13.3%
40	Egypt, Arab Rep. of	80.4	11.2%	90	Aruba, Kingdom of the Netherlands	3.1	14.3%
41	Australia	79.8	8.9%	91	Bosnia and Herzegovina	3.0	1.9%
42	Kuwait	79.0	8.4%	92	Albania	2.8	3.3%
43	Indonesia	78.6	3.1%	93	Luxembourg	2.2	10.3%
44	Brazil	67.4	1.1%	94	Hong Kong SAR	2.1	0.0%
45	Denmark	66.5	5.1%	95	Iceland	2.0	1.6%
46	Pakistan	64.6	18.9%	96	Papua New Guinea	2.0	4.6%
47	Argentina	61.7	8.5%	97	Trinidad and Tobago	1.9	1.5%
48	Qatar	56.7	7.5%	98	Haiti	1.8	4.0%
49	United Arab Emirates	50.9	2.6%	99	Yemen, Republic of	1.6	1.7%
50	Belarus, Rep. of <sup>4)</sup>	50.0	39.2%	100	Suriname	1.5	13.0%

### 2.2.1. Fattori che influenzano Il prezzo dell'oro

Nell'analisi della quotazione dell'oro è opportuno fare una distinzione dei fattori che possono influenzare i prezzi di mercato a seconda dell'arco temporale della nostra analisi.

I driver principali della quotazione dell'oro, analizzati dal World Gold Council sono quattro:

- Espansione economica: un buono stato di salute dell'economia è correlato positivamente con la domanda di oro nel settori dei gioielli, della tecnologia e degli investimenti a lungo termine nella commodity.
- Rischio e incertezza del mercato: l'oro è da sempre considerato come un bene rifugio, quindi in momenti di incertezza nel mercato aumenta la domanda di investimenti per questa commodity.

- Costo Opportunità: i movimenti dei tassi di interesse e dei tassi di cambio delle valute influenzano sicuramente la domanda di investimenti nella commodity.

Per quanto concerne i tassi di interesse, si può notare che una loro diminuzione, porta gli investitori a spostare le proprie risorse finanziarie su asset con tassi di rivalutazione maggiore, tra cui l'oro, il quale assume un tasso di rivalutazione più alto rispetto al mercato obbligazionario a seconda dell'andamento dei tassi di interesse.

Inoltre, si può affermare che in momenti di debolezza delle valute dei mercati in cui è quotato l'oro, gli investimenti diventeranno più appetibili agli investitori, che acquistando la commodity, contribuiranno a far aumentare il suo prezzo.

- Entusiasmo del mercato: i trend di mercato nell'acquisto dei titoli forward, future ETF possono diminuire o aumentare le performance della quotazione dell'oro. Ad esempio, in fasi di tensione geo-politica il prezzo dell'oro può salire, in quanto gli investitori lo utilizzano come bene rifugio per diversificare il rischio del proprio portafoglio.

Oltre a questi elementi specifici del mercato, come tutte le materie prime, la determinazione dei prezzi è guidata dall'incontro tra la domanda e l'offerta.

Per quanto riguarda l'offerta, la tabella n.3, riporta i dati della World Central Bank, in tonnellate, relativi alla produzione mondiale annuale di oro dal 1970 al 2020, con un focus più dettagliato negli ultimi cinque anni.

Come riportato nella tabella, si può notare come i volumi produttivi siano rimasti pressochè inalterati negli ultimi 5 anni.

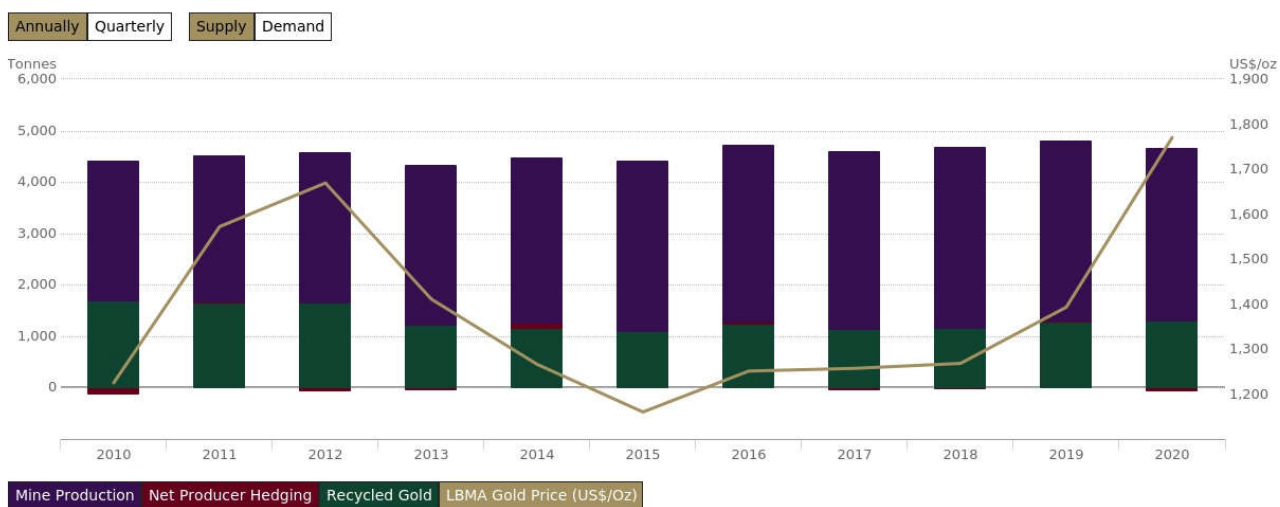
La Cina è attualmente il primo produttore mondiale di oro (11,3%), seguita da Australia (10%) e Russia (9,4%). Un dato rilevante della statistica è la produzione della materia prima del Sud Africa. Nel 1990 il paese ha garantito una produzione di 605 tonnellate d'oro (circa il 26,4% del totale), contro una produzione di appena 97 tonnellate (3%) nell'anno 2020.



Tabella 3 Produzione mondiale di Oro in tonnellate

	1990	2000	2005	2010	2016	2017	2018	2019	2020
	(metric tons)								
<b>Production</b>									
China	100	175	209	341	453	426	401	380	365
Australia	242	296	263	260	291	292	313	326	326
Russia	n/a	144	163	201	253	270	280	305	305
United States	294	353	256	231	222	237	226	200	188
Canada	169	156	121	91	164	171	194	183	170
Ghana	17	72	67	93	129	130	149	142	150
Kazakhstan	n/a	27	18	30	75	85	97	103	118
Uzbekistan	n/a	88	84	90	100	102	102	102	102
Mexico	9	24	30	79	132	127	118	109	102
South Africa	605	428	297	191	142	137	117	105	97
Sudan	0	6	5	2	93	107	94	94	94
Peru	9	134	206	164	153	152	140	128	87
Indonesia	11	125	158	106	81	99	112	109	86
Brazil	102	61	38	62	80	80	85	78	78
Argentina	1	26	28	64	57	61	58	59	59
Burkina Faso	3	1	1	23	39	46	53	51	55
Papua New Guinea	34	73	67	67	62	65	68	74	54
Mali	2	29	44	39	47	48	61	48	49
Colombia	n/a	37	36	54	63	43	36	37	48
Others	n/a	305	413	583	585	572	555	698	704
<b>World</b>	<b>2,133</b>	<b>2,560</b>	<b>2,504</b>	<b>2,771</b>	<b>3,222</b>	<b>3,252</b>	<b>3,259</b>	<b>3,332</b>	<b>3,237</b>

## Supply and demand statistics



Data as of 31 March, 2021

Sources: ICE Benchmark Administration, Metals Focus, Refinitiv GFMS, World Gold Council;

See disclaimer on [gold.org](http://gold.org)

Figura 4 Offerta annuale di oro in funzione del prezzo (US\$/oz)

Dal lato della domanda, emergono i volumi consumati da parte di Cina e India, che detengono il primato.

I due paesi, avente un'economia in forte espansione, utilizzano l'oro nei settori industriali chimici ed elettronici. Tuttavia, la percentuale più alta del consumo deriva da ragioni socio-culturali. I metalli preziosi vengono considerati come riserve di valore e tramandati alle generazioni successive.

È da sottolineare che i consumi di questa materia prima si sono ridimensionati nel 2020. La Cina ha registrato una diminuzione rispetto al 2019 del 29% dei consumi, mentre l'India del 35%.

Tabella 4. Volumi di consumo d'oro mondiale in tonnellate

Source: Metals Focus, Refinitiv GFMS, ICE - World Gold Council

**Consumer demand in selected countries  
(Tonnes)**

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
India	1,001.7	974.0	914.2	958.6	833.5	857.2	666.1	771.2	760.4	690.4	446.4
Pakistan	33.8	39.9	39.4	44.2	35.8	37.6	42.2	44.9	40.2	38.0	28.8
Sri Lanka	0.0	0.0	0.0	0.0	9.1	10.7	10.5	11.2	9.6	7.9	4.3
<b>Greater China</b>	676.3	873.0	918.3	1,449.8	1,083.7	1,062.2	985.0	1,029.7	1,058.4	899.5	641.5
China, P.R.:											
Mainland	645.7	816.3	856.3	1,345.5	1,005.3	995.5	929.4	971.6	994.3	849.1	612.9
Hong Kong SAR	24.2	43.8	49.7	85.5	61.4	52.8	42.9	46.0	52.1	39.6	17.8
Taiwan Province of China	6.3	13.0	12.4	18.8	17.0	13.8	12.8	12.2	11.9	10.8	10.8
Japan	-19.1	-36.4	5.1	20.9	13.7	32.7	34.0	13.3	29.0	-3.1	4.5
Indonesia	50.6	60.8	61.7	88.1	63.4	59.0	59.5	58.8	64.0	54.5	37.6
Malaysia	17.0	20.4	20.6	28.6	26.9	22.1	19.4	19.1	18.9	17.6	13.1
Singapore	9.9	13.1	14.8	20.9	20.5	18.1	17.1	16.6	16.4	15.0	9.4
Korea, Republic of	19.2	26.6	28.0	40.8	39.7	45.5	39.2	41.2	41.0	38.9	35.4
Thailand	71.3	113.6	110.0	153.8	108.7	90.2	81.5	75.4	80.6	46.5	-81.5
Vietnam	82.3	103.4	84.3	99.6	66.7	63.4	58.3	53.9	59.5	56.4	39.8
<b>Middle East</b>	327.2	301.4	300.2	388.9	328.8	302.9	229.0	241.7	258.9	231.0	170.4
Saudi Arabia	84.5	72.5	65.9	84.7	84.0	84.4	60.2	54.4	49.6	46.0	31.1
UAE	76.9	68.2	60.0	78.5	66.0	60.2	51.2	52.2	41.9	39.1	26.7
Kuwait	10.3	9.7	9.3	15.5	17.8	16.2	15.0	16.5	17.0	15.9	13.0
Egypt	55.3	36.4	44.2	52.8	51.1	43.2	28.2	24.5	27.2	29.0	21.9
Islamic Republic of Iran	85.0	98.2	102.1	127.7	75.4	67.3	45.1	64.5	91.2	69.6	56.3
Other Middle East	15.2	16.6	18.6	29.7	34.6	31.6	29.2	29.6	32.0	31.5	21.5
Turkey	108.8	145.5	114.0	184.1	116.7	72.1	70.1	93.6	74.2	89.6	147.1
Russian Federation	60.3	64.4	67.7	79.7	75.5	47.9	42.0	42.3	45.6	48.1	36.0
<b>Americas</b>	303.8	266.8	224.2	251.4	224.6	248.5	266.5	215.4	209.9	205.9	236.0
United States	226.7	198.5	160.2	187.8	164.7	190.6	210.1	158.9	154.4	150.8	186.0
Canada	20.4	20.6	17.6	19.6	18.4	17.5	17.9	17.0	16.5	17.4	20.7
Mexico	27.5	21.9	18.4	18.4	18.0	19.0	18.2	19.3	18.9	18.1	13.3
Brazil	29.2	25.9	28.1	25.6	23.5	21.4	20.2	20.2	20.1	19.6	16.1
<b>Europe ex CIS</b>	394.5	421.6	317.3	336.5	274.4	297.6	278.0	266.5	245.2	221.0	304.8
France	22.0	25.7	18.4	16.8	15.1	13.1	9.2	13.2	11.8	12.9	14.5
Germany	136.6	154.4	118.9	143.4	111.3	126.1	121.1	117.0	106.9	96.2	166.0
Italy	34.2	26.9	22.6	21.1	19.9	19.2	19.0	18.8	18.5	18.2	13.7
Spain	11.6	8.2	8.4	7.8	8.3	8.5	8.3	8.4	8.7	8.8	6.9
United Kingdom	40.8	38.8	35.0	32.7	33.3	34.4	37.1	33.6	34.6	32.2	29.6
Switzerland	86.9	96.8	62.2	61.7	47.7	50.3	45.6	42.5	36.6	29.8	42.9
Austria	13.7	15.1	10.4	13.0	10.2	12.2	10.8	9.7	8.2	6.2	11.6
Other Europe	48.6	55.8	41.5	40.0	28.6	33.7	27.0	23.4	19.9	16.5	19.6
<b>Total above</b>	3,137.5	3,388.2	3,219.7	4,145.9	3,321.6	3,267.7	2,898.5	2,994.7	3,011.7	2,657.2	2,073.5
Other & stock change	123.8	218.2	259.5	310.9	278.1	283.6	278.5	290.1	327.1	332.5	223.9
<b>World total</b>	3,261.4	3,606.5	3,479.2	4,456.8	3,599.7	3,551.3	3,176.9	3,284.9	3,338.7	2,989.8	2,297.4

### **2.2.2. Mercati di riferimento dell'oro**

Il prezzo dell'oro spot viene fissato sul mercato Over The Counter (OTC) di Londra, precisamente sul London Bullion Market, dedicato ai metalli preziosi Oro e Argento.

La determinazione del prezzo dell'oro, quotato in dollari, è denominata London Gold Fixing. Questo processo avviene due volte ogni "business day", precisamente alle 10:30 e alle 15:00 ora Inglese ed è gestito dalla London Bullion Market Association (LBMA), che ha come scopo quello di gestire l'integrità e la trasparenza per l'industria globale dei metalli preziosi, garantendone lo scambio e la qualità nel Loco London Market, il più grande mercato mondiale per i metalli preziosi.

Il prezzo dell'oro LBMA viene utilizzato come benchmark di riferimento per tutti gli scambi della materia prima. La sua determinazione avviene tramite delle aste. Le aste sono organizzate in round da 30 secondi e gestite dall' ICE Benchmark Administration (IBA), che inizialmente pubblica il prezzo per un round.

I partecipanti decidono quindi di inserire, modificare o annullare i propri ordini per quel prezzo. Se alla fine del round lo sbilanciamento tra acquisto e vendita è troppo elevato, allora il prezzo verrà definito in un nuovo round.

Se lo sbilancio rientra nella soglia l'asta è conclusa e il prezzo è fissato. L'eventuale sbilanciamento è ripartito equamente tra tutti i partecipanti diretti (anche se non hanno effettuato ordini o non hanno effettuato il login) e il volume netto per ogni partecipante operato al prezzo finale.

Il prezzo finale viene quindi pubblicato come LBMA Gold Price o LBMA Silver Price in dollari USA e anche convertito nei benchmark in sterline inglesi ed euro utilizzando i tassi di cambio di quando si è concluso il round finale.

I partecipanti hanno anche 30 minuti immediatamente prima dell'inizio dell'asta per mettere in coda i loro ordini. Questo è noto come "Round zero".

Per quanto riguarda i tre contratti future più utilizzati dalla comunità finanziaria sono quelli del NYMEX, New York Mercantile Exchange, del CBOT, Chicago Board of Trade e quello del TOCOM, il Tokyo Commodity Exchange. Sia al NYMEX che al CBOT, il valore del contratto, ovvero il quantitativo reale di oro sottostante ad un contratto, è di 100 once Troy.

### 2.3. IL RAME

Il rame è un metallo contraddistinto dal simbolo atomico Cu (numero atomico 29) e rientra tra gli elementi di transizione della tavola periodica, con molte affinità chimico-fisiche con metalli come argento e oro.

La sua struttura atomica reticolare cubica a facce centrate, nella quale gli elettroni non sono legati strettamente al nucleo dei rispettivi atomi e dunque sono abbastanza liberi di muoversi e di creare un flusso di corrente, permette al materiale di essere un ottimo conduttore di calore ed elettricità.

Di conseguenza, il rame è molto utilizzato nel settore dei trasporti, nell'elettronica e nelle telecomunicazioni; inoltre è impiegato nella realizzazione di leghe come bronzi e ottoni.

Le principali miniere di rame sono situate lungo la Cordigliera delle Ande e le Montagne Rocciose: i principali Paesi estrattori della materia prima sono il Cile, il Perù. Nello specifico, come riportato in tabella n.5, attualmente il Cile è di gran lunga il paese avente la maggior riserva di rame al mondo, con il 27,8% dell'estrazione mineraria mondiale nel 2020.

Altre importanti riserve di rame sono collate in Cina, Stati Uniti e Australia.

La roccia estratta contiene in media meno dell'1% di rame. Quindi prima di essere venduto sul mercato come prodotto il rame deve subire delle lavorazioni chimico-fisiche.

Il primo passo importante per la raffinazione è la "concentrazione", che consente di separare il minerale dalla roccia di scarto. Generalmente questo processo viene realizzato vicino alle miniere per abbattere i costi di trasporto della materia prima.

Successivamente, il minerale viene lavorato con specifici reagenti chimici fino a raggiungere livelli di concentrazione del rame fino al 40%.

Dopo il processo di concentrazione, il minerale viene trasportato all'interno delle raffinerie dove subisce le lavorazioni necessarie a renderlo vendibile sul mercato.

Attraverso la raffinazione, il Rame raggiunge livelli di concentrazione del 99,99%.

Esistono molteplici processi di raffinazione del minerale che dipendono dal risultato ottenuto dal processo di contrazione. A seconda che il minerale sia ricco di solfuri oppure ossidi vengono applicati rispettivamente i processi pirometallurgici ed idrometallurgici.



Figura 5. Evoluzione dei prezzi del Rame dal 1970 al 2020

La tabella n.5 riporta i dati, in migliaia tonnellate, aggiornati ad Aprile 2021 della World Central Bank relativi alla produzione mineraria, alla produzione di rame raffinato e il suo consumo a livello mondiale dal 1970 ad oggi, con un focus più dettagliato dal 2017 fino a fine 2020.

I dati della Banca Centrale Mondiale riportano che nel 2020 la Cina è stato il primo produttore (circa il 42% della produzione mondiale) e consumatore (58,3% della domanda complessiva) di rame raffinato. I dati confermano la forte espansione della domanda della materia prima impiegata nelle infrastrutture e nell'industria elettronica; tanto è vero che quattro delle cinque maggiori raffinerie e 10 delle prime 20 si trovano nella Cina continentale. Le cinque più grandi raffinerie hanno una capacità combinata di oltre 7 milioni di tonnellate, circa il 33% della capacità globale.

Un' ulteriore evidenza che dimostra l'importanza assunta dal rame negli ultimi 50 anni è da ricercare nei volumi prodotti e consumati, i quali sono più che triplicati dal 1970.

Infine, una componente importante da considerare è il riciclo dei rottami di rame, riconosciuta come produzione di rame secondario, il quale è in grado di mantenere le medesime proprietà fisiche e la qualità del rame primario. Il recupero e il riciclo dei prodotti a fine vita è un esempio della sostenibilità del rame.

Uno studio pubblicato 2012, realizzato dal ICSG (International Copper Study Group), conferma che il 44,8% del rame utilizzato in Europa proviene dal riciclo.

Tabella 5 Produzione mineraria e raffinata e consumo di Rame in migliaia di tonnellate dal 1970 al 2021  
 Fonte: Sources: British Geological Survey, Metallgesellschaft, U.S. Geological Survey, World Bureau of Metals Statistics, World Bank.

	1970	1980	1990	2000	2010	2017	2018	2019	2020
<b>Mine Production</b>									
Chile	686	1068	1588	4,602	5,419	5,504	5,832	5,787	5,732
Peru	220	367	323	553	1,247	2,446	2,437	2,455	2,149
China	n/a	165	300	549	1,180	1,656	1,507	1,601	1,855
Congo, Dem. Rep.	386	n/a	n/a	33	378	1,095	1,225	1,433	1,400
United States	1560	1181	1588	1,440	1,129	1,258	1,216	1,257	1,196
Australia	158	244	330	832	870	849	911	934	896
Zambia	684	610	546	249	732	794	854	790	861
Russia	n/a	n/a	n/a	580	703	722	773	791	791
Mexico	61	175	299	365	270	742	751	770	750
Kazakhstan	n/a	n/a	n/a	433	404	745	621	711	721
Canada	610	716	794	634	522	597	548	561	585
Indonesia	0	59	162	1,006	871	666	651	400	492
Poland	83	343	370	454	425	419	401	449	442
<i>Others</i>	1,755	2,811	3,027	1,486	1,988	2,700	2,678	2,841	2,767
<b>World</b>	<b>6,202</b>	<b>7,739</b>	<b>9,327</b>	<b>13,217</b>	<b>16,139</b>	<b>20,193</b>	<b>20,404</b>	<b>20,781</b>	<b>20,638</b>
<b>Refined Production</b>									
China	120	295	558	1,312	4,540	8,889	8,949	9,447	10,021
Chile	647	811	1,192	2,669	3,244	2,430	2,461	2,269	2,329
Japan	603	1,014	1,008	1,437	1,549	1,488	1,595	1,495	1,583
Russia	n/a	n/a	n/a	824	900	949	1,020	1,028	1,028
United States	1,489	1,730	2,017	1,802	1,093	1,079	1,111	1,030	905
Congo, Dem. Rep.	683	n/a	n/a	29	254	699	821	842	850
Germany	134	425	532	709	704	695	670	600	656
Korea, Rep.	5	88	192	471	556	664	665	638	638
Poland	69	357	346	486	547	522	502	566	561
Kazakhstan	n/a	n/a	n/a	395	323	429	443	477	483
Mexico	n/a	n/a	n/a	399	247	411	422	448	412
Spain	n/a	n/a	n/a	316	347	420	429	383	387
Australia	n/a	n/a	n/a	484	424	386	377	426	381
<i>Others</i>	2,978	4,755	4,829	3,429	4,484	4,420	4,188	3,828	3,583
<b>World Total</b>	<b>6,729</b>	<b>9,475</b>	<b>10,675</b>	<b>14,761</b>	<b>19,214</b>	<b>23,479</b>	<b>23,652</b>	<b>23,477</b>	<b>23,818</b>
<b>Refined Consumption</b>									
China	180	286	512	1,869	7,385	11,790	12,482	12,800	14,528
United States	1,860	1,868	2,150	2,979	1,760	1,771	1,814	1,829	1,758
Germany	788	870	1,028	1,309	1,312	1,180	1,200	1,017	1,057
Japan	821	1,158	1,577	1,351	1,060	998	1,039	1,011	891
Korea, Rep.	10	85	324	862	856	767	717	633	619
Italy	274	388	475	674	619	635	552	556	470

Turkey	14	33	103	248	369	445	482	464	438
India	55	77	135	246	514	486	512	527	432
Mexico	54	117	127	464	274	361	407	442	405
Others	3,236	4,502	4,349	5,094	5,197	4,902	4,719	4,583	4,309
<b>World</b>	<b>7,291</b>	<b>9,385</b>	<b>10,780</b>	<b>15,096</b>	<b>19,347</b>	<b>23,335</b>	<b>23,926</b>	<b>23,860</b>	<b>24,907</b>

*Note: n/a implies data not available. Refined production and consumption include significant recycled material.*

### 2.3.1. Fattori che influenzano il prezzo del Rame

L'ampio raggio di impiego del Rame in svariati settori, rende il suo prezzo influenzabile da numerosi fattori.

Una delle componenti principali che influenza il prezzo del rame è il tasso di crescita dei mercati emergenti come Cina e India. La costruzione di infrastrutture abitative e di trasporto ha come elemento chiave l'utilizzo del rame e di conseguenza ne guida la domanda.

L'industria edilizia raccoglie una buona parte della domanda di questa commodity, ad esempio nella realizzazione di impianti idraulici, cavi elettrici e nella costruzione di tetti.

Pertanto, indicatori economici come i tassi sui mutui e la demografia influenzano il settore dell'edilizia e conseguentemente anche la domanda di rame.

Altri elementi che contribuiscono alla formazione e alla fluttuazione del prezzo del Rame sono di carattere sociale, come guerre e scioperi. Quest'ultimi fattori possono bloccare o rallentare la produzione della materia prima per interi periodi temporali, contribuendo così alla riduzione dell'offerta del bene e all'impennamento dei prezzi.

Un altro fattore di carattere esogeno che condiziona l'ascesa dei prezzi può essere riconosciuto nei disastri naturali e nel deperimento delle miniere da cui si estrae il rame.

L'innalzamento dei prezzi e la scarsità dell'offerta di questo metallo possono portare il mercato a ricercare soluzioni alternative a seconda del campo d'applicazione, influenzando conseguentemente la domanda.

Ad esempio l'alluminio è impiegato come alternativa al rame negli ingranaggi elettrici e nei cavi.

Anche il nichel, il piombo e il ferro possono essere considerati dei sostituti del rame in svariati settori.

Questi ultimi aspetti contribuiscono a far crescere il prezzo della commodity, creando instabilità nei mercati in cui viene scambiata.

I maggiori rischi sul prezzo del rame sono legati alle condizioni di crescita economica di paesi importanti per il mercato, come la Cina, ma anche dalla presenza di nuovi prodotti alternativi, come il PVC utilizzato nelle tubature.





Source: See World Bank Commodities Price Data.

*Figura 6. Prezzo spot del Rame dal 01/2005 al 03/2021*

### **2.3.2. Principali mercati del rame**

Il rame come materia prima si scambia principalmente sul London Metal Exchange (LME), sul quale è quotato quotidianamente. Su questo mercato si possono scambiare i future e le opzioni sul rame così come dei prodotti derivati che hanno il LME Index come sottostante.

Altri contratti a termine e opzioni sul rame vengono scambiati sul New York Mercantile Exchange (NYMEX) negli Stati Uniti e sullo Shanghai Future Exchange in Cina.

## 2.4. LA SOIA

La Soia (*Glycine Max*) è una pianta dall'elevato contenuto proteico appartenente alla famiglia delle leguminose ed è uno dei prodotti alimentari più coltivati al mondo.

Oltre ad essere consumata nella sua forma originaria, dalla soia vengono ricavati diversi beni alimentari come la farina, i germogli, il latte, l'olio e il tofu, il cui commercio è in forte espansione al livello mondiale.

Inoltre, questa materia prima è impiegata nella produzione di prodotti farmaceutici e materiali eco-sostenibili, tra cui i bio-fuels.

Il tasso di domanda è in crescita grazie alla molteplicità dei campi di applicazione di questa commodity. Infatti, la soia è utilizzata in modo sempre maggiore grazie alla crescita di prodotti alimentari e prodotti per la salute da essa derivati e per l'atteggiamento dei policy-makers sempre più orientati verso l'impiego di carburanti come bio-diesel.

La tabella n.6 riporta i dati in migliaia di tonnellate relativi alla produzione, alla macinazione, alle esportazioni e importazioni di soia annue con cadenza decennale dal 1970 ad oggi, con un focus più dettagliato dagli anni dal 2017 al 2021.

Come si può notare, la Cina nell'ultimo anno ha importato circa 100 milioni di tonnellate di soia (circa il 60% del totale delle importazioni), confermando il paese come principale attore nella crescita della domanda della materia prima.

L'Unione Europea è il secondo mercato a livello mondiale per quanto riguarda le importazioni di soia, occupando quasi il 10% del totale. Oltre ad importare l'olio di soia per il bio-diesel e prodotti alimentari dedicati alla nutrizione animale, è da sottolineare una contro-tendenza nella dieta alimentare europea, la quale promuove sempre di più prodotti di origine vegetale in grado di sostituire la carne. Il trend è favorito dalla crescente preoccupazione dei governi in termini di impatto sull'ambiente e sulla salute derivato dalla produzione e il consumo di carne animale. Questo fattore beneficia sicuramente il mercato della soia, che grazie alle sue proprietà proteiche, riesce a proporre diverse soluzioni alternative, come il tofu.

Tuttavia, il settore della soia può essere influenzato negativamente dall'impatto ambientale dei suoi raccolti. Un'articolo del 2021 pubblicato dall'International Institute for Sustainable Development riporta che circa l'80% delle varietà di soia sono geneticamente modificate. Nonostante la capacità di generare dei raccolti più voluminosi, la soia GM può portare a rendere il suolo infertile, con impatti disastrosi nel lungo termine per l'ambiente e la biodiversità.

In termini di produzione, si può notare che nel 2020 il Brasile ha superato in termini di volumi la produzione di soia statunitense.

Brasile, Stati Uniti e Argentina insieme hanno garantito nell'ultimo anno solare circa l'82% del fabbisogno mondiale di Soia.

#### **2.4.1. Fattori che influenzano il prezzo della soia**

Uno dei fattori principali che guidano il prezzo di questa commodity sono le condizioni meteorologiche. In, particolare, le abbondanti piogge nelle aree nord e centro-occidentali del Brasile, dove viene raccolta la maggior parte della produzione mondiale, possono influire negativamente sui raccolti in termini di ritmi di lavoro e quindi volumi generati. Anche condizioni climatiche opposte, come ad esempio la siccità, può ridurre l'offerta di materia prima con conseguente aumento dei prezzi spot.

Come rivelato dai dati sulle importazioni, la domanda cinese per questa commodity influisce sulla formazione del prezzo di scambio sui mercati, dal momento che l'85% del suo fabbisogno deriva principalmente da Stati Uniti e Brasile.

Dati i grandi volumi di soia scambiati tra Cina e Stati Uniti, sono da considerare come componenti determinante per il mercato gli accordi commerciali e dazi doganali, protagonisti della guerra commerciale tra i due paesi iniziata nel 2018.

Un altro elemento di notevole importanza, menzionato nel paragrafo precedente, è il trend nei consumi alimentari di carne della popolazione animale.

I semi di soia sono largamente utilizzati nell'alimentazione animale per le loro elevate proprietà proteiche e quindi il loro consumo è positivamente correlato con il consumo della carne.

#### **2.4.2. Mercato di riferimento per la Soia**

Il Chicago Board of Trade è il mercato che attualmente quota in dollari il corso della soia. Su questo mercato vengono scambiati la maggior parte dei titoli future e opzioni relative alla commodity.

Tabella 6. Produzione, frantumazione, importazioni ed esportazioni di Soia in migliaia di tonnellate dal 1970 al 2021

Fonte: U.S. Department of Agriculture (April 9, 2021 update).

	1970/1971	1980/1981	1990/1991	2000/2001	2010/2011	2017/2018	2018/2019	2019/2020	2020/2021
<b>Production</b>									
Brazil	n/a	15.2	15.8	39.5	75.3	123.4	119.7	128.5	136.0
United States	30.7	48.9	52.4	75.1	90.7	120.1	120.5	96.7	112.5
Argentina	n/a	3.5	11.5	27.8	49.0	37.8	55.3	48.8	47.5
China	8.7	7.9	11.0	15.4	15.4	15.3	16.0	18.1	19.6
India	0.0	0.4	2.6	5.3	10.1	8.4	10.9	9.3	10.7
Paraguay	0.1	0.6	1.3	3.5	7.1	11.0	8.5	9.9	9.7
Canada	0.3	0.7	1.3	2.7	4.4	7.7	7.4	6.1	6.4
Russia	n/a	n/a	0.7	0.3	1.1	3.6	4.0	4.4	4.3
Ukraine	n/a	n/a	0.1	0.1	1.7	4.0	4.8	4.5	3.1
Bolivia	0.0	0.0	0.4	1.1	2.3	2.9	3.0	2.8	2.9
European Union	n/a	n/a	n/a	1.3	1.2	2.5	2.7	2.6	2.6
Others	2.4	3.6	7.3	3.7	6.3	7.1	8.2	7.3	8.0
World	42.1	80.9	104.3	175.8	264.7	343.8	361.0	339.0	363.2
<b>Crushings</b>									
China	1.5	1.5	3.9	18.9	55.0	90.0	85.0	91.5	96.0
United States	20.7	27.8	32.3	44.6	44.9	55.9	56.9	58.9	59.6
Brazil	n/a	13.8	14.2	22.7	36.3	44.2	42.5	46.0	46.8
Argentina	n/a	0.9	7.0	17.3	37.6	36.9	40.6	38.8	40.5
European Union	n/a	n/a	n/a	16.8	12.3	15.0	15.6	16.4	17.0
India	0.0	0.4	2.4	4.5	9.4	7.7	9.6	8.4	9.7
Mexico	0.3	1.5	1.9	4.5	3.6	5.3	6.2	6.0	6.2
Egypt	n/a	0.1	0.1	0.3	1.6	3.2	3.5	4.7	4.7
Russia	n/a	n/a	0.4	0.4	2.1	4.6	4.7	4.7	4.5
Others	n/a	n/a	n/a	16.5	19.4	32.0	34.0	36.2	37.5
World	35.3	69.8	86.8	146.5	222.2	294.8	298.5	311.4	322.5
<b>Exports</b>									
Brazil	n/a	1.8	2.5	15.5	30.0	76.1	74.9	92.1	86.0
United States	11.8	19.7	15.2	27.1	41.0	58.1	47.7	45.8	62.1
Argentina	n/a	2.7	4.5	7.3	9.2	2.1	9.1	10.0	6.9
Paraguay	0.0	0.6	1.0	2.4	5.1	6.0	4.9	6.6	6.0
Canada	0.0	0.1	0.2	0.7	2.9	4.9	5.3	3.9	4.2
Others	0.5	0.4	2.1	0.7	3.4	5.9	7.0	6.6	5.8
World	12.3	25.3	25.4	53.7	91.6	153.2	148.8	165.1	170.9
<b>Imports</b>									
China	n/a	0.5	0.0	13.2	52.3	94.1	82.5	98.5	100.0
European Union	n/a	n/a	n/a	17.7	12.5	14.6	15.0	15.7	15.4
Mexico	0.1	1.4	1.4	4.4	3.5	4.9	5.9	5.7	6.0
Argentina	n/a	n/a	n/a	0.3	0.0	4.7	6.4	4.9	4.7
Egypt	n/a	0.0	0.0	0.3	1.6	3.3	3.7	4.9	4.6
Thailand	n/a	0.0	n/a	1.3	2.1	2.5	3.2	3.8	4.0
Japan	3.2	4.2	4.4	4.8	2.9	3.3	3.3	3.3	3.4
Others	9.3	20.1	19.8	11.1	14.7	26.4	26.0	28.1	29.7
World	12.6	26.2	25.5	53.1	89.7	153.7	145.9	165.0	167.8

## **2.5. PETROLIO GREGGIO**

Il petrolio greggio estratto dal sottosuolo è una miscela di idrocarburi densa presente soltanto in alcune regioni della Terra, ragione per cui ha assunto un alto valore economico nell'ultimo secolo.

Caratterizzato da una colorazione che va dal giallo chiaro al marrone scuro o verdastro, il greggio può presentare una composizione più o meno densa a seconda della regione di estrazione.

Una densità minore del greggio estratto corrisponde ad un maggior valore economico della materia prima, in quanto gli idrocarburi meno densi sono i principali componenti della benzina.

L' American Petroleum Institute ha sviluppato un'unità di misura, il grado API, che rispecchia il valore commerciale del petrolio greggio, ed è utilizzata per indicare il peso specifico di una miscela idrocarburica liquida. Miscele di greggio meno dense hanno un grado API più alto rispetto alle miscele più dense, aventi un API più basso.

### **2.5.1. Riserve di petrolio**

Il CIA World Factbook, pubblicazione annuale della Central Intelligence Agency (CIA), ente che pubblica annualmente i dati statistici fondamentali a livello mondiale da un'analisi geologica e ingegneristica, ha stimato con un elevato grado di accuratezza le più grandi riserve di petrolio che possono essere sfruttate a livello commerciale per ciascun paese.

Nella tabella n.7 è riportata la classifica, datata 01/01/2018, dei primi 20 paesi in termini di numero di barili disponibili come riserve secondo il report della Central Intelligence Agency.

Nella classifica pubblicata nel 2018 dalla CIA non sono state riportate le riserve di petrolio degli Stati Uniti. Questo dato è stato recuperato dal sito dell'istituzione Statunitense EIA (Energy Information Administration). L' EIA è un'agenzia statistica che fa parte del Dipartimento di Energia Statunitense che si occupa di raccogliere, analizzare e pubblicare i dati relativi al mercato energetico per favorire la comprensione dell'impatto dell'energia sull'economia del paese.

Un report di Gennaio 2021, riporta che le riserve ufficiali di petrolio statunitensi ammontavano a circa 43,8 miliardi di barili al 31 Dicembre 2021, collocando così il paese al decimo posto in termini di volumi di barili di greggio, dopo la Libia e prima della Nigeria.

È da sottolineare che i paesi con un elevato numero di riserve non sono collocati tra i maggiori produttori di petrolio a causa agli elevati costi di estrazione, i quali dipendono a loro volta dalla geologia del territorio.

Tabella 7. Classifica mondiale dei paesi con le maggiori riserve di Petrolio Greggio  
 Fonte: CIA World Factbook,

Stato	N° di Barili	Area Geografica
Venezuela	302,300,000,000	Sud America
Saudi Arabia	266,200,000,000	Medio Oriente
Canada	170,500,000,000	Nord America
Iran	157,200,000,000	Medio Oriente
Iraq	148,800,000,000	Medio Oriente
Kuwait	101,500,000,000	Medio Oriente
United Arab Emirates	97,800,000,000	Medio Oriente
Russia	80,000,000,000	Asia Centrale
Libya	48,360,000,000	Africa
Nigeria	37,450,000,000	Africa
Kazakhstan	30,000,000,000	Asia Centrale
China	25,630,000,000	Asia Orientale
Qatar	25,240,000,000	Medio Oriente
Brazil	12,630,000,000	Sud America
Algeria	12,200,000,000	Africa
Angola	9,523,000,000	Africa
Ecuador	8,273,000,000	Sud America
Azerbaijan	7,000,000,000	Medio Oriente
Mexico	6,630,000,000	Nord America
Norway	6,376,000,000	Europa

### 2.5.2. Le diverse tipologie di Petrolio Greggio

Data la diversa composizione della miscela di idrocarburi del petrolio Greggio, esistono diversi indici che ne misurano il valore. Gli indici più importanti considerati a livello mondiale sono tre: il WTI, il Brent e il Dubai Fateh, legati alle aree geografiche dove sono situati i diversi giacimenti petroliferi.

#### WTI

Il WTI, West Texas Intermediate, è utilizzato come benchmark sul petrolio statunitense.

Il WTI viene estratto in America, in particolare negli stati interni del Texas, la Louisiana e il North Dakota. Vista la collocazione nell'entroterra, il collegamento ai porti marittimi è limitato, con un conseguente aumento dei costi di trasporto. Gli elevati costi di trasporto sono compensati dalla sua bassa viscosità e bassa concentrazione di zolfo (inferiore al 0,5%), che lo rendono meno costoso da raffinare, confermandone la sua elevata qualità. Il suo API si aggira intorno a 40, considerato come un valore alto dell'indice.

Il suo prezzo spot viene definito a Cushing (Oklahoma) e viene preso come parametro di riferimento per lo scambio di titoli future scambiato sul New York Mercatile Exchange (NYME).

## **BRENT**

Il Brent è il benchmark per il petrolio di riferimento del mercato europeo e prende il nome da un giacimento di petrolio nel Mare del Nord, precisamente in Scozia. Il petrolio Brent tuttavia comprende la produzione di 19 di giacimenti petroliferi, tra cui i più importanti sono il Brent Blend, Forties Blend, Oseberg ed Ekofisk e il Troll.

Come il WTI, anche il Brent presenta una bassa concentrazione di zolfo (inferiore al 0,5%), con un indice API che si aggira intorno a 38. Perciò dal punto di vista chimico-fisico è più pesante rispetto al WTI. Tuttavia, la vicinanza al mare rende i costi di trasporto nettamente inferiori rispetto al petrolio statunitense.

L'indice Brent ricopre un importante ruolo nella definizione dei prezzi, in quanto viene utilizzato per prezzare più del 60% del totale del petrolio e impiegato principalmente nella realizzazione di benzine e distillati

Il punto di raccolta più grande per il petrolio Brent è l'isola di Sullom Voe, a nord della Scozia. Questo terminale è la più grande piazza per lo scambio spot di questa tipologia di greggio.

In passato era possibile scambiare i titoli sul petrolio Brent solo sul London International Petroleum Exchange.

Adesso invece il mercato di riferimento per il Brent è l'ICE, Intercontinentale Exchange di Londra, dove il petrolio non viene scambiato nella sua forma fisica, ma attraverso lo scambio di titoli future.

## **DUBAI FATEH**

Il FATEH, conosciuto anche come DUBAI/OMAN viene utilizzato come indice di riferimento per il petrolio proveniente dagli Emirati Arabi Uniti.

Questa tipologia di greggio ha una concentrazione di zolfo del 2%, circa 6 sei volta superiore rispetto al BRENT e WTI, per questo motivo è denso e pesante e più complesso da raffinare. Il suo indice API è intorno a 32.

La borsa di riferimento per il Dubai Fateh è la Borsa di Tokyo e viene prezzato in yen a differenza del WTI e del Brent, quotati in dollari.

### **2.5.3. Fattori che determinano il prezzo del Petrolio**

Prima di analizzare nel dettaglio le determinanti del prezzo del petrolio, occorre conoscere l'Organizzazione dei Paesi esportatori di petrolio, meglio conosciuta come OPEC. In principio, la sua

missione è quella di coordinare la stabilizzazione dei mercati petroliferi, unificando le politiche di mercato per i paesi che ne fanno parte.

L'OPEC fu fondato nel 1960 a Baghdad, Iraq. Inizialmente comprendeva solamente i cinque membri fondatori, ovvero Iran, Iraq, Kuwait, Arabia Saudita e Venezuela.

Successivamente entrarono a far parte dell'organizzazione anche Qatar (1961), Indonesia (1962), Libia (1962), Emirati Arabi Uniti (1967), Algeria (1969), Nigeria (1971), Ecuador (1975), Gabon (1975), Angola (2007), Guinea Equatoriale (2017) e Congo (2018).

Indonesia, Ecuador e Qatar non fanno più parte dell'organizzazione.

Quindi ad oggi gli stati membri dell'OPEC sono 13.

L'OPEC è responsabile per il coordinamento della produzione di petrolio per i paesi membri. Per questo motivo l'OPEC svolge una funzione molto importante nella definizione dell'offerta di greggio a livello mondiale, dato che i paesi membri garantiscono quasi il 40% del fabbisogno.

Non fanno parte dell'organizzazione grandi produttori come gli Stati Uniti, Russia, Canada e Cina, che hanno una grande influenza nella formazione dei prezzi del petrolio e nelle relazioni geo-politiche con l'OPEC.

Sul lato dell'offerta, generalmente livelli di stock elevati portano ad un abbassamento dei prezzi e viceversa.

A titolo d'esempio nel 2014, i paesi dell'OPEC, non hanno trovato un accordo per la definizione dei volumi di produzione. In particolare, il braccio di ferro dell'Arabia Saudita nel non voler rallentare la produzione di petrolio ha portato il mercato ad avere un'eccedenza di materia prima, assorbita successivamente da un crollo dei prezzi del greggio.

Un'ulteriore evidenza dell'importanza politica dell'OPEC nel mercato petrolifero è attribuita al ruolo giocato da questa organizzazione nella crisi energetica del 1973. I paesi arabi membri dell'OPEC, decisero di sostenere l'azione bellica di Egitto e Siria contro Israele, aumentando il prezzo del barile e interrompendo gli approvvigionamenti verso i paesi filo-israeliani. Le conseguenze sul sistema industriale furono disastrose e da allora si iniziarono a ricercare nuove fonti energetiche e nuovi giacimenti petroliferi, come quelli Norvegesi.

È da notare che il mercato del petrolio può subire degli shock di offerta. In particolare, condizioni meteo avverse, come uragani, possono portare a conseguenze disastrose per l'approvvigionamento del greggio qualora si verificassero malfunzionamenti negli oleodotti. Inoltre, la movimentazione delle petroliere, riveste un elevato rischio di generare incidenti, con disastrose conseguenze per l'ambiente.

Tra i più noti disastri possiamo citare il naufragio della petroliera Haven, considerato come il peggior



disastro petrolifero del Mediterraneo.

Infine, i conflitti militari e il terrorismo, protagonisti della situazione politica del Medio Oriente negli ultimi vent'anni, hanno avuto un forte impatto sulla produzione del greggio. Ad esempio l'Iraq, al centro del conflitto in Medio Oriente, ha diminuito la propria produzione di petrolio nel decennio dal 2000 al 2010, in controtendenza agli altri stati vicini aventi una situazione politica più stabile.

Sul lato della domanda è da considerare un consumo crescente del petrolio da parte della Cina, avvenuto contemporaneamente alla sua rapida crescita economica a partire dagli anni 2000.

Questo è confermato dai volumi di barili giornalieri mediamente consumati dal paese; nel decennio 2000-2010 infatti il Paese ha raddoppiato il proprio consumo, passando da un impiego medio di 4967 barili a 9390 barili al giorno.

Questo ha contribuito ad una ripresa della crescita dei prezzi successivamente alla crisi di Lehman Brothers dell'anno 2008, anno in cui la commodity ha raggiunto il suo prezzo massimo prima di un crollo verticale.

Un esempio lampante di come la domanda per la commodity influenzi il suo prezzo è quello che si è verificato il 20 Aprile 2020, in piena crisi pandemica da Coronavirus. In corrispondenza di tale data, il prezzo del petrolio è andato in negativo per la prima volta nella storia; precisamente il costo del barile ha raggiunto il picco, negativo, di -37,00 \$ dollari al barile. Il fattore principale di questo fenomeno è attribuito al congelamento della domanda a causa del blocco delle attività causate dal coronavirus. Complice del crollo del prezzo è stata una mancata corrispondente contrazione della produzione; come conseguenza, il costo di stoccaggio ha superato i prezzi di mercato. La mancanza di aree per lo stoccaggio quindi ha portato i produttori a "pagare" per smaltire i barili in eccesso di petrolio.

Nonostante la rapida crescita del tasso di consumo di greggio della Cina, gli Stati Uniti rimangono il primo paese consumatore della commodity. Al netto di qualche oscillazione, il consumo medio giornaliero è leggermente diminuito rispetto all'inizio del millennio, passando da un consumo medio giornaliero di 19.594 (anno 2000) a 19.400 (anno 2019) barili.

Nell'ambito della formazione dei prezzi ricoprono grande importanza i report nell'ambito delle riserve di petrolio nel paese rilasciati dall' Energy Information Administration (EIA), entità governativa statunitense. Le analisi rilasciate settimanalmente sono un indicatore affidabile per determinare la domanda dinamica di petrolio del primo produttore e consumatore mondiale di crude oil.

Un ulteriore fattore che riveste grande importanza in un'economia sempre più "green" è lo sviluppo di energie rinnovabili in grado di sostituire in diversi impieghi il petrolio. L'innovazione tecnologica

e lo sviluppo di tecnologie alternative possono portare ad un calo della domanda del petrolio e dei suoi derivati.

Da ultimo, come per la maggior parte delle *commodities*, la forza del dollaro è un indicatore capace di condizionare lo scambio di titoli che hanno come sottostante il petrolio. Quando il dollaro perde terreno rispetto alle altre valute, gli investimenti petroliferi sono più convenienti a parità di potere d'acquisto.

Questo è dovuto al fatto che gli indicatori più importanti, il WTI e il Brent, sono quotati in dollari.

La figura 7 riporta l'andamento prezzo del petrolio dal 2000 al 2020 e gli eventi principali legati al mercato che hanno influenzato il suo prezzo lungo l'ultimo ventennio.

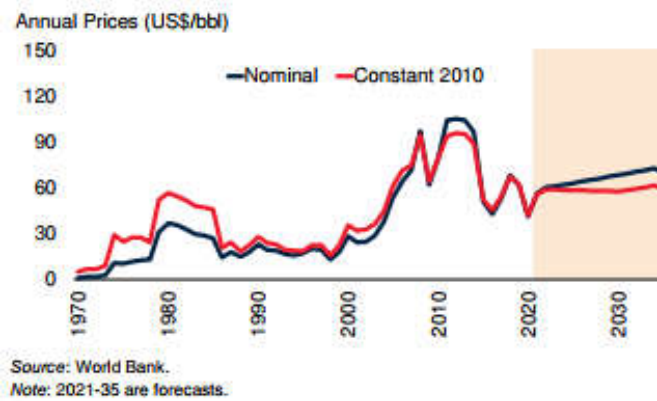


Figura 7. Andamento dei prezzi per petrolio dal 1970 a Marzo 2021

La tabella n.8 riporta i dati, in migliaia di barili, relativi alla produzione e al consumo di greggio giornaliero medio con cadenza decennale dal 1970 ad 2010, con un focus più dettagliato dagli anni dal 2016 al 2019.

Tabella 8. Produzione e consumo di Petrolio in migliaia di barili giornalieri dal 1970 al 2021  
Fonte: BP Statistical Review (June 2020 update).

	1970	1980	1990	2000	2010	2016	2017	2018	2019
<b>Production</b>									
United States	11,297	10,170	8,914	7,733	7,558	12,349	13,135	15,360	17,045
Saudi Arabia	3,851	10,270	7,106	9,121	9,865	12,406	11,892	12,261	11,832
Russia	n/a	n/a	10,342	6,583	10,379	11,269	11,255	11,438	11,540
Canada	1,473	1,764	1,968	2,884	3,457	4,655	5,037	5,501	5,651
Iraq	1,549	2,658	2,149	2,613	2,469	4,423	4,538	4,632	4,779
United Arab Emirates	780	1,735	1,985	2,599	2,937	4,038	3,910	3,912	3,998
China	616	2,122	2,778	3,257	4,077	3,999	3,846	3,798	3,836
Iran	3,848	1,479	3,270	3,850	4,421	4,578	5,007	4,801	3,535
Kuwait	3,036	1,757	964	2,244	2,564	3,150	3,009	3,050	2,996
Brazil	167	188	651	1,276	2,125	2,591	2,721	2,679	2,877
Nigeria	1,083	2,058	1,787	2,174	2,533	1,900	1,969	2,007	2,109
Kazakhstan	n/a	n/a	571	740	1,676	1,655	1,838	1,927	1,931
Mexico	487	2,129	2,941	3,456	2,959	2,456	2,224	2,068	1,918
Qatar	363	476	434	851	1,630	1,938	1,882	1,900	1,883
Norway	n/a	528	1,716	3,326	2,133	1,991	1,965	1,845	1,731
Algeria	1,054	1,134	1,367	1,549	1,689	1,577	1,540	1,511	1,486
Angola	103	150	475	746	1,812	1,745	1,671	1,519	1,417
Libya	3,357	1,862	1,424	1,475	1,799	412	929	1,165	1,227
United Kingdom	4	1,676	1,933	2,710	1,358	1,015	1,005	1,092	1,118
Oman	332	285	695	955	865	1,004	971	978	971
Venezuela	3,754	2,228	2,244	3,112	2,842	2,347	2,096	1,475	918
Colombia	226	131	446	687	786	886	854	865	886
India	140	193	715	726	901	874	885	869	826
Others	n/a	n/a	8,150	10,052	10,575	8,815	8,618	8,601	8,683
<b>World</b>	<b>48,075</b>	<b>62,942</b>	<b>65,022</b>	<b>74,718</b>	<b>83,409</b>	<b>92,072</b>	<b>92,798</b>	<b>95,254</b>	<b>95,192</b>
<b>Consumption</b>									
United States	14,710	17,062	16,939	19,594	18,324	18,618	18,883	19,428	19,400
China	554	1,707	2,297	4,697	9,390	12,248	12,842	13,375	14,056
India	390	643	1,210	2,258	3,378	4,632	4,860	5,112	5,271
Japan	3,876	4,989	5,240	5,542	4,434	4,006	3,971	3,855	3,812
Saudi Arabia	435	592	1,136	1,627	3,206	3,875	3,838	3,769	3,788
Russia	n/a	n/a	5,042	2,540	2,878	3,219	3,195	3,282	3,317
Korea, Rep.	162	476	1,041	2,260	2,370	2,771	2,801	2,781	2,760
Canada	1,472	1,898	1,747	2,043	2,333	2,393	2,393	2,443	2,403
Brazil	513	1,080	1,229	1,843	2,271	2,436	2,481	2,377	2,398
Germany	2,765	3,014	2,685	2,741	2,373	2,307	2,374	2,260	2,281
Iran	224	570	1,004	1,404	1,788	1,764	1,808	1,835	2,018
Mexico	441	1,072	1,611	1,952	2,040	1,950	1,883	1,821	1,733
Indonesia	138	386	652	1,148	1,411	1,572	1,660	1,724	1,732
United Kingdom	2,031	1,649	1,751	1,713	1,604	1,597	1,610	1,584	1,545
France	1,860	2,220	1,895	1,986	1,703	1,529	1,539	1,538	1,530
Others	n/a	n/a	20,886	23,138	27,352	29,485	29,872	30,166	30,230
<b>World</b>	<b>45,313</b>	<b>61,408</b>	<b>66,364</b>	<b>76,485</b>	<b>86,856</b>	<b>94,404</b>	<b>96,013</b>	<b>97,348</b>	<b>98,272</b>

### 3. Analisi dell'efficienza delle quotazioni delle *commodities* tramite i Random Walks

L'analisi dell'efficienza del mercato è stata oggetto di studio da parte di numerosi studiosi.

La più famosa definizione di mercato efficiente è stata formalizzata nel 1970 da Eugene Fama:

*“Un Mercato in cui i prezzi riflettono sempre pienamente l'informazione disponibile è chiamato efficiente”<sup>1</sup>.*

Più recentemente, l'economista Malkiel (1992) ha fornito una definizione più dettagliata del concetto di mercato efficiente:

*“Il mercato dei capitali è efficiente se riflette pienamente e correttamente tutte le informazioni rilevanti per determinare i prezzi dei titoli.*

*Formalmente, si dice che il mercato sia efficiente rispetto a una serie di informazioni se il prezzo dei titoli rimanesse inalterato se tali informazioni venissero rivelate a tutte le persone.*

*Inoltre, l'efficienza rispetto a un set di informazioni implica che è impossibile realizzare profitti economici negoziando la base [di quel set di informazioni]”<sup>2</sup>*

La prima frase di questa definizione riprende l'affermazione di Eugene Fama.

Nella seconda frase viene riportato che l'efficienza del mercato può essere testata valutando l'impatto sulle quotazioni dei titoli, una volta rivelate le informazioni ai partecipanti del mercato, mentre nella terza viene proposta un'alternativa per giudicare l'efficienza del mercato, ossia quella di misurare il profitto generato dallo scambio di informazioni già note al mercato.

Nella letteratura, sono state individuate tre forme di efficienza del mercato (Roberts, 1967):

- Forma debole (*Weak form*): l'informazione è limitata alle quotazioni e ai rendimenti storici dei titoli.
- Forma semi-forte (*Semi-strong form*): i prezzi riflettono l'insieme dei dati pubblicamente disponibili (*publicly available information*).

---

1 Fama, Eugene F., 1970, "Efficient Capital Markets: A Review of the Theory and Empirical Work," *Journal of Finance*  
2 Malkiel, B., 1992, "Efficient Market Hypothesis," in Newnan, P., M. Milgate, and J. Eatwell (eds.), *New Palgrave Dictionary of Money and Finance*, Macmillan, London.

- Forma forte (*Strong form*): l'informazione comprende anche i dati non pubblicati (*private information*).

Sulla base delle diverse classificazioni, occorre evidenziare che l'efficienza del mercato non implica la capacità degli operatori di riflettere il reale valore delle quotazioni, ma che nei prezzi dei titoli non sono incorporate distorsioni sistematiche. Le distorsioni verranno riassorbite in seguito all'uscita di nuove informazioni.

Come conseguenza, gli operatori del mercato non sono in grado di prevedere i prezzi futuri dei titoli, ma di fornire la miglior previsione possibile sulla base delle informazioni disponibili in un determinato istante di tempo.

Per formalizzare il concetto di informazione incorporata nei prezzi, sono stati impiegati tre modelli:

- Modello Fair-Game: le informazioni in possesso degli operatori finanziari in un certo istante di tempo non sono utili per impostare una strategia di investimento che preveda sistematicamente di generare rendimenti in eccesso positivi.

Il rendimento in eccesso  $\varepsilon_{t+1}$  è definito come:

$$\varepsilon_{t+1} = \frac{P_{t+1} - P_t}{P_t} = \frac{E[P_{t+1}|\theta_t] - P_t}{P_t} = \frac{P_{t+1} - E[P_{t+1}|\theta_t]}{P_t}$$

dove  $P_t$  e  $\theta_t$  indicano rispettivamente il prezzo e le informazioni disponibili all'istante di tempo  $t$ ,

$$\varepsilon_{t+1} = r_{t+1} - E[r_{t+1}|\theta_t]$$

dove  $r_{t+1}$  è il rendimento atteso all'istante  $t + 1$

- Modello Martingala: un martingala è definito come un processo stocastico che soddisfa la seguente condizione:  $E[Y_{t+1} | Y_0, \dots, Y_t] = Y_t$ , riprendendo così il modello Fair-Game. Secondo questo approccio, in un mercato efficiente il prezzo futuro atteso è uguale al prezzo corrente moltiplicato per  $(1 + \text{il rendimento atteso})$ :

$$E[P_{t+1}|\theta_t] = P_t * (1 + E[r_{t+1}|\theta_t])$$

Il prezzo atteso scontato per il tasso di rendimento atteso è un martingala, mentre il prezzo atteso è un Submartingala, il quale gode della seguente proprietà:

$$E[Y_{t+1} | Y_0, \dots, Y_t] > Y_t$$

- Modello Random Walk: secondo quest'ultimo approccio i prezzi sono identicamente e indipendentemente distribuiti. Sono chiamati processi senza memoria perché le quotazioni in un determinato istante di tempo non sono utili per conoscere i rendimenti futuri e il valore

atteso dei rendimenti in eccesso è nullo. I Random Walks hanno la proprietà di essere dei Martingala.

### 3.1. RANDOM WALK

Il manuale di econometria “The Econometrics of Financial Markets” (Campbell, Lo, MacKinlay) classifica tre modelli di processi randomici: Random Walk 1, Random Walk 2, Random Walk 3.

#### RANDOM WALK 1

L'ipotesi di questo modello, denominato RW1, è la distribuzione indipendente e identica dei log-rendimenti dei prezzi, generalizzati dalla seguente equazione:

$$p_t = \mu + p_{t-1} + \epsilon_t$$
$$\epsilon_t \sim IID(0, \sigma^2)$$

Dove  $\mu$  è il rendimento atteso (drift) ed  $\epsilon_t$  una variabile casuale i.i.d. con media 0 e varianza  $\sigma^2$ .

La peculiarità di RW1 è la correlazione nulla dei rendimenti attesi e di ogni loro funzione lineare.

Il Random Walk è un processo non stazionario con media e varianza lineari nel tempo.

#### RANDOM WALK 2

Rispetto a RW1, con Random Walk 2 decade l'assunzione che i log-rendimenti siano identicamente distribuiti. Rimuovere questo vincolo ha permesso di valutare la randomicità dei prezzi lungo lag-temporali più ampi; in linea di principio, i cambiamenti tecnologici, sociali e istituzionali lungo periodi temporali estesi rendono difficile pensare che la distribuzione statistica dei rendimenti rimanga costantemente inalterata.

Per questo motivo RW2 è considerato una generalizzazione di RW1, perché è ammessa l'eteroschedasticità della variabile  $\epsilon_t$ .

#### RANDOM WALK 3

Random Walk 3 (RW3), rilassa il vincolo di indipendenza nella distribuzione della variabile casuale  $\epsilon_t$ . La particolarità di questo modello risiede nella correlazione nulla tra i log-rendimenti lungo lag-temporali diversi.

Per valutare l'efficienza nel mercato delle commodity è stato utilizzato il modello RW3. In

particolare, è stato applicato il Variance Ratio Test, adattato al caso dei rendimenti con distribuzione eteroschedastica.

### 3.2. Variance Ratio Test

L'ipotesi principale dei tre modelli di Random Walks illustrati nel paragrafo precedente, è che la varianza dei profitti del cammino casuale sia lineare nel suo intervallo di dati. Questo significa che la varianza  $X_t - X_{t-q}$  è q-volte la varianza di  $X_t - X_{t-1}$ , dove  $X_t$  è considerato il logaritmo naturale del prezzo al tempo t. Quindi il Variance Ratio valuta il rapporto delle varianze dei Log-rendimenti lungo lag temporali diversi.

Al fine di costruire un test statistico per valutare le ipotesi del modello RW1, vengono presi come riferimento i parametri esposti nello studio di Lo and MacKinlay (1988, 1989).

Il Variance Ratio è definito come:

$$VR(q) = \frac{\sigma^2(q)}{\sigma^2(1)}$$

dove  $\sigma^2(1)$  è la varianza di  $X_t - X_{t-1}$ , mentre  $\sigma^2(q)$  è 1/q-volte la varianza di  $X_t - X_{t-q}$ :

$$\sigma^2(1) = \frac{1}{nq - 1} \sum_{t=1}^{nq} (X_t - X_{t-1} - \hat{\mu})^2$$

$$\sigma^2(q) = \frac{1}{m} \sum_{t=1}^{nq} (X_t - X_{t-q} - q\hat{\mu})^2$$

I parametri  $\hat{\mu}$  e m sono definiti come

$$\hat{\mu} = \frac{1}{nq} \sum_{t=1}^{nq} (X_t - X_{t-1}) = \frac{1}{nq} (X_{nq} - X_0)$$

$$m = q * (nq - q + 1) \left(1 - \frac{q}{nq}\right)$$

Sotto l'ipotesi nulla  $H_0$  che i rendimenti siano non correlati sull'asse temporale il Variance Ratio segue la seguente distribuzione normale:

$$\sqrt{nq}(VR(q) - 1) \sim N\left(0, \frac{2(2q - 1)(q - 1)}{3q}\right)$$

con media 0 e varianza  $\frac{2(2q-1)(q-1)}{3q}$

La normale standard per questa distribuzione è calcolata come:

$$Z(q) = \frac{(VR(q) - 1)}{\sqrt{\frac{2(2q - 1)(q - 1)}{3q}}}$$

La variabile casuale standardizzata segue una distribuzione normale con media 0 e varianza 1.

Qualora i log-rendimenti fossero distribuiti eteroschedasticamente, l'ipotesi nulla verrebbe rifiutata. Per questo motivo si è cercato di costruire un test per verificare correlazione nulla dei log-rendimenti anche in presenza di eteroschedasticità, dato che la varianza delle variabili casuali cambia nel tempo. Pertanto il precedente test era valido per valutare l'efficienza del mercato per il modello RW1, in cui le variabili casuali erano identicamente e indipendentemente distribuite.

Per quanto riguarda il modello RW3, il quale considera solamente la correlazione tra i log-rendimenti, è stato costruito test ad hoc su quattro ipotesi nulle:

$$H(1): \forall t, E[\epsilon_t] = 0 \text{ e } E[\epsilon_t \epsilon_{t-q}] = 0 \quad \forall q$$

$H(2): \forall t \text{ e } \forall t \forall q \geq 0$  esiste un  $\delta > 0$  per cui  $E\left[|\epsilon_t \epsilon_{t-q}|^{2(r+\delta)}\right] < A < \infty$  dove  $r$  è un coefficiente della v.c.  $\epsilon_t$

$$H(3): \lim_{nq \rightarrow \infty} \frac{1}{nq} \sum_{t=1}^{nq} [\epsilon_t^2] = \sigma^2 < \infty$$

$$H(4): \forall t, E[\epsilon_t \epsilon_{t-j} \epsilon_t \epsilon_{t-k}] = 0 \quad \forall j, k \neq 0 \text{ e } j \neq k$$

L'ipotesi  $H(1)$  è relativa ai rendimenti non correlati che si desidera valutare per il cammino casuale.



Le condizioni  $H(2)$  e  $H(3)$  forniscono un limite massimo alla dipendenza ed eterogeneità che permettono di rispettare la Legge dei grandi Numeri e il Teorema del limite Centrale.

L'ultima condizione  $H(4)$  richiede che i log-rendimenti abbiano delle autocorrelazioni campionarie asintoticamente non correlate.

Le suddette ipotesi nulle permettono variazioni nella varianza della distribuzione delle v.c. e i parametri calcolati da Lo and Mackinlay in caso di distribuzione eteroschedastica sono i seguenti:

$$VR(q) = 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) * \hat{\rho}(k)$$

In cui  $\hat{\rho}(k)$  è una stima del coefficiente di autocorrelazione.

Sotto le ipotesi nulle Lo e Mackinlay mostrano che la distribuzione di  $VR(q) - 1$  converge a 0 per ogni  $q$  e numero di log-prezzi crescenti. Inoltre, è stato calcolato uno stimatore della varianza robusto all'eteroschedasticità  $\theta(q)$ , definito come:

$$\hat{\theta}(q) = 4 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right)^2 * \widehat{\delta}_k$$

Dove  $\widehat{\delta}_k$  è la varianza asintotica relativa ad ogni coefficiente  $\hat{\rho}(k)$ , ed è definita come:

$$\widehat{\delta}_k = \frac{\sum_{j=k+1}^{nq} (p_j - p_{j-1} - \hat{\mu})^2 (p_{j-k} - p_{j-k-1} - \hat{\mu})^2}{[\sum_{j=1}^{nq} (p_j - p_{j-1} - \hat{\mu})^2]^2}$$

Per cui la normale standard per questa statistica è calcolata come:

$$Z(q) = \frac{\sqrt{nq}(VR(q) - 1)}{\sqrt{\hat{\theta}}}$$

### 3.3. Risultati variance ratio test sul mercato delle *commodities*

In questo paragrafo verranno esposti i risultati generati dal test statistico Variance Ratio robusto all'eteroschedasticità.

Al fine di valutare la correlazione dei log-rendimenti lungo lag-temporali diversi, si è calcolata la statistica  $Z$  esposta nel paragrafo precedente sia nel caso omoschedastico che nel caso eteroschedastico.

Nello specifico, il test è stato effettuato utilizzando le quotazioni giornaliere lungo un arco temporale di 5 anni, relative alle cinque *commodities* descritte nel secondo capitolo: cotone, oro, petrolio, rame e soia.

Il test, utilizzato per verificare l'efficienza del mercato in termini di randomicità dei rendimenti giornalieri, verifica che la correlazione delle variazioni del prezzo non siano correlate su periodi temporali diversi, rispettando così le ipotesi del modello RW3.

Come metrica per il test si è calcolato il *Variance Ratio* da 1 a 64 giorni; successivamente si è evidenziato il valore più significativo della variabile standardizzata, capace di indicare l'inefficienza nel mercato lungo un determinato arco temporale. In termini pratici, questo valore corrisponde alla  $Z(q)$  più lontana dal valore zero, corrispondente ad un mercato efficiente con riferimento alle correlazioni dei rendimenti.

Dai risultati, emerge che la variabile normale standardizzata non è sistematicamente discostata nei due casi omoschedastico ed eteroschedastico, tranne che per la materia prima oro, la cui statistica omoschedastica è sistematicamente più alta del modello eteroschedastico. La spiegazione va ricercata nel cambio subito dalla distribuzione della varianza del minerale prezioso negli ultimi 5 anni.

Tuttavia, come previsto dal modello RW3, il quale non prevede che i rendimenti siano identicamente distribuiti nel tempo, la variabile casuale standardizzata è sistematicamente più piccola rispetto al caso omoschedastico.

### 3.3.1. Regressione lineare e indice $R^2$

Al fine di valutare l'efficienza del mercato per il mercato delle *commodities*, come alternativa al *Variance Ratio test*, è stato applicato il modello di regressione lineare semplice:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \text{ con } i = 1, \dots, n$$

dove  $n$  corrisponde al numero di osservazioni,  $\beta_0$  e  $\beta_1$  sono rispettivamente l'intercetta e la pendenza della retta di regressione. La variabile indipendente  $X_i$  applicata al modello è l'osservazione  $i$ -esima, coincidente con la data del log-rendimento giornaliero, mentre la variabile dipendente  $Y_i$  è il log-rendimento delle quotazioni giornaliere. Infine,  $u_i$  rappresenta l'errore di regressione, costituito da fattori omessi dalla retta di regressione.

Idealmente, in un mercato efficiente le due variabili non dovrebbero essere correlate. Una delle statistiche in grado di fornire una misura di bontà del modello è l' $R^2$ . Questo indicatore della regressione valuta la frazione della varianza di  $Y$  spiegata da  $X$ ; il suo valore può variare tra 0 (nessun adattamento) e 1 (perfetto adattamento) ed è privo di unità. Da un mercato efficiente in linea teorica ci si aspetterebbe un valore di  $R^2$  tendente a 0.

Dal punto di vista matematico l' $R^2$  è definito come:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

ESS, *Explained Sum of Squares*, è la devianza spiegata dal modello. In termini pratici, viene calcolata come la sommatoria delle deviazioni dei valori stimati e la loro media, al quadrato.

Invece TSS, *Total Sum of Squares*, è la devianza totale, che considera anche i fattori omessi dal modello di regressione. Come evidenziato dalla formula corrisponde alla sommatoria della devianza dei valori osservati rispetto alla media, al quadrato.

Nella regressione semplice con una sola  $X$ , l' $R^2$  è equivalente al quadrato del coefficiente di correlazione tra  $X$  e  $Y$ .

## COTONE

Dal Variance Ratio Test effettuato sulla commodity cotone emerge che la variabile standardizzata segue un percorso quasi sovrapposto nei casi omoschedastico ed eteroschedastico lungo lag-temporali differenti. Il Test evidenzia che non si può rifiutare l'ipotesi nulla che la correlazione tra i log-rendimenti sia correlata lungo diversi lag-temporali (p-value=23,77%)

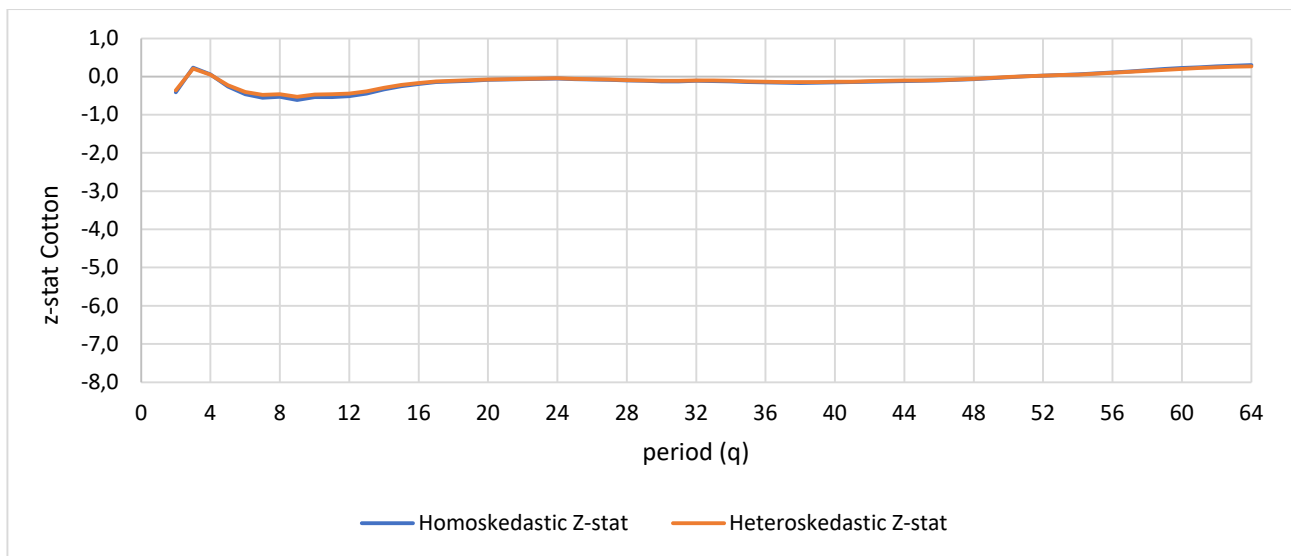


Figura 8. Variabile standardizzata "Z" del Variance Ratio relativa ai log-rendimenti del Cotone nei casi Omoschedastico ed Eteroschedastico

Tabella 9. Valori più significativi del Test - Commodity Cotone

COTTON	Homoscedastic	Heteroskedastic
Most significant Z-stat	-1,3443	-1,1808
p-value	17,89%	23,77%

Per quanto riguarda il modello di regressione, il valore dell' $R^2$ , prossimo allo zero, conferma quanto riportato del Variance Ratio Test in termini di efficienza del mercato. Come è evidenziato dal grafico, la retta presenta una pendenza e un'intercetta quasi nulle, rispettivamente  $-2E-08$  e  $0,0003$ .

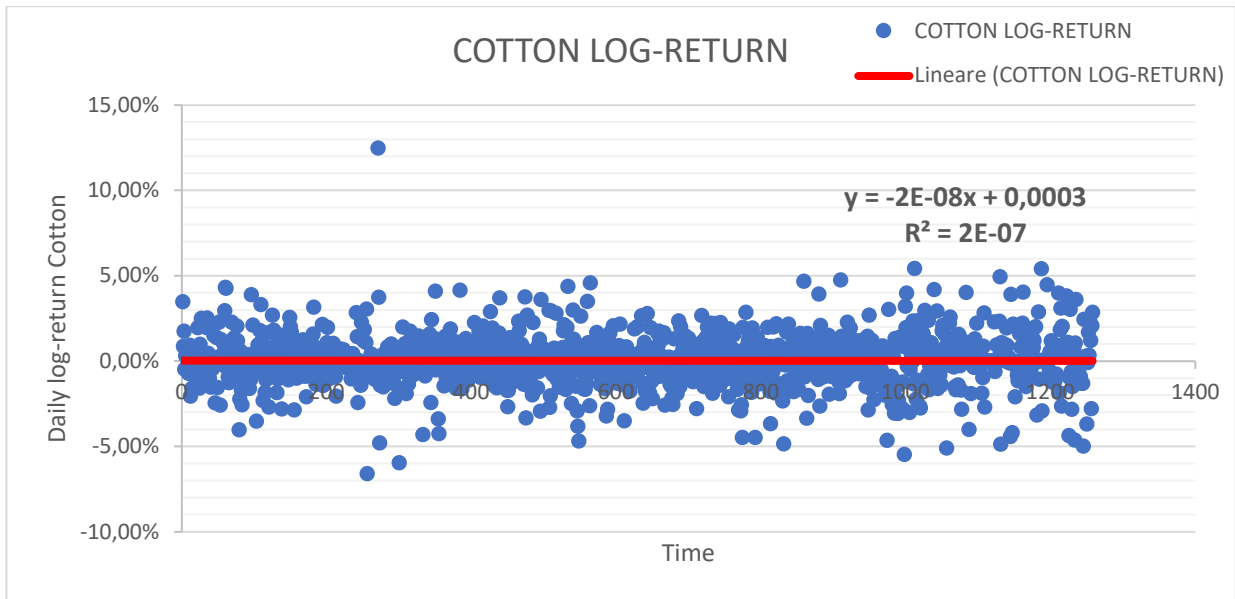


Figura 9. Regressione lineare log-rendimenti cotone

## ORO

Il Variance Ratio Test effettuato sulla commodity Oro evidenzia che non si può rifiutare l'ipotesi nulla che la correlazione tra i log-rendimenti sia correlata lungo diversi lag-temporali con un elevato livello di fiducia (p-value= 40,46%). A differenza delle altre *commodities*, per l'oro è empiricamente dimostrato che il p-value aumenti sensibilmente nel caso in cui il test realizzato sia robusto all'eteroschedasticità. Questo può suggerire con un'alta probabilità che la distribuzione dei log-rendimenti sia cambiata negli ultimi 5 anni.

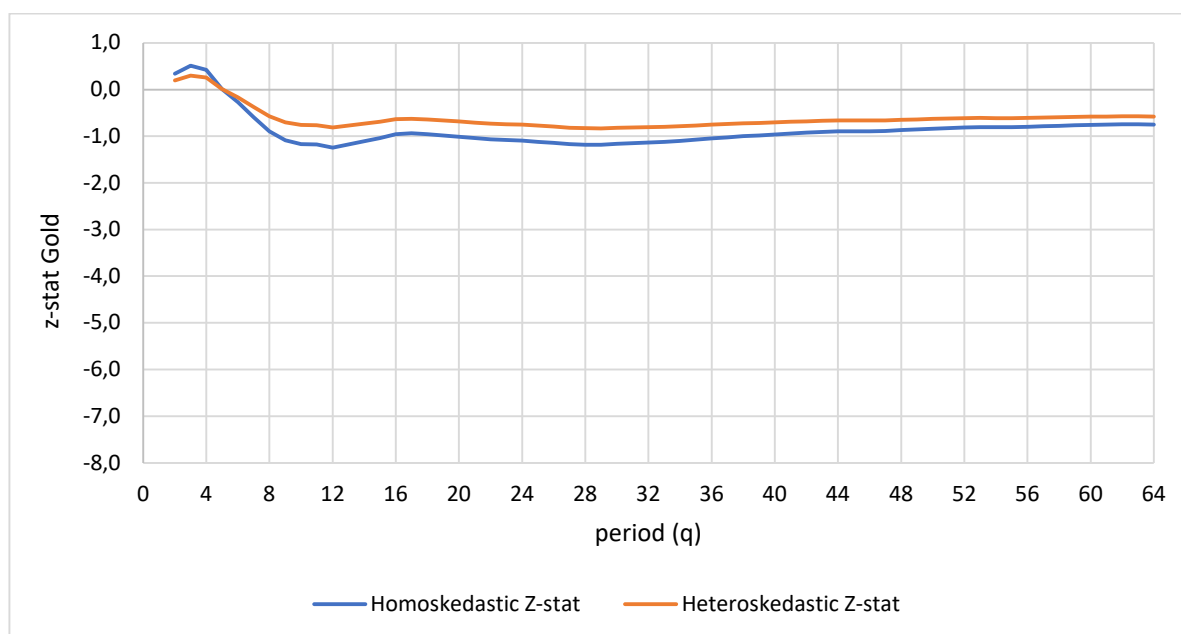


Figura 10. Variabile standardizzata "Z" del Variance Ratio relativa ai log-rendimenti dell'Oro nei casi Omoschedastico ed Eteroschedastico

Tabella 10. Valori più significativi del Test - Commodity Oro

GOLD	Homoscedastic	Heteroskedastic
Most significant Z-stat	-1,2460	-0,8334
p-value	21,28%	40,46%

In termini di regressione lineare, il valore dell' $R^2$  è prossimo allo zero ( $4E-05$ ). Questo risultato, unito al valore della variabile standardizzata per il Variance Ratio eteroschedastico, è sufficiente per confermare l'efficienza delle quotazioni dell'oro negli ultimi 5 anni.

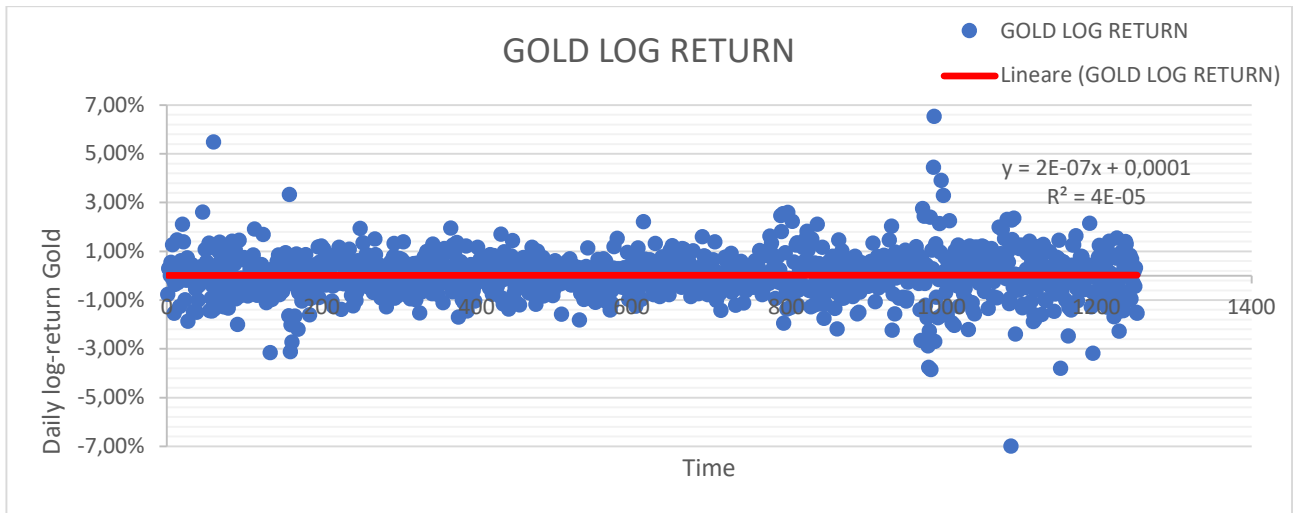


Figura 11. Regressione lineare log-rendimenti oro

## PETROLIO

In corrispondenza di quotazioni negative, non è possibile dal punto matematico calcolare il log-rendimento. Per applicare il Variance Ratio Test alla commodity Petrolio, si è reso necessario eliminare dalla serie storica i log rendimenti relativi al 20/04/2020 e 21/04/2020. Come esposto nel secondo capitolo, in corrispondenza di questi giorni il prezzo del petrolio è diventato negativo a causa della crisi generata dal Covid-19.

Come verificatosi per l'oro, i valori più significativi della variabile casuale standardizzata sono ampiamente diversi nei due casi omoschedastico ed eteroschedastico, con p-value che portano a formulare conclusioni completamente diverse. In particolare, se ci soffermassimo al Test che considera le variabili casuali identicamente distribuite lungo la serie temporale, rifiuteremmo l'ipotesi nulla che i log-rendimenti non siano correlati. Viceversa, il Ratio robusto all'eteroschedasticità, permette di non rifiutare l'ipotesi nulla con un elevato livello di fiducia (p-value 31%).

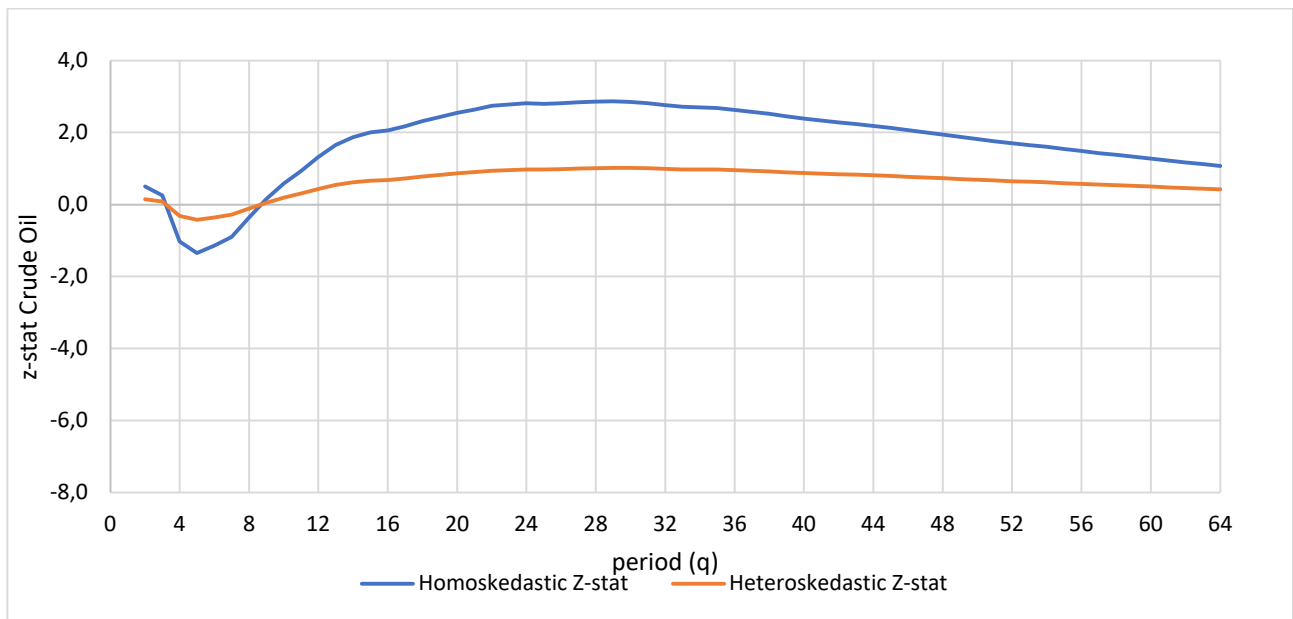


Figura 12. Variabile standardizzata "Z" del Variance Ratio relativa ai log-rendimenti del Petrolio nei casi Omoschedastico ed Eteroschedastico

Tabella 11. Valori più significativi del Test - Commodity Petrolio

CRUDE OIL	Homoscedastic	Heteroskedastic
Most significant Z-stat	2,8704	1,0151
p-value	0,41%	31,00%



In termini di regressione lineare, l' $R^2$  ha un valore equivalente a 0,0003. Si può confermare quindi che il mercato del petrolio sia efficiente in termini di correlazione tra i log-rendimenti.

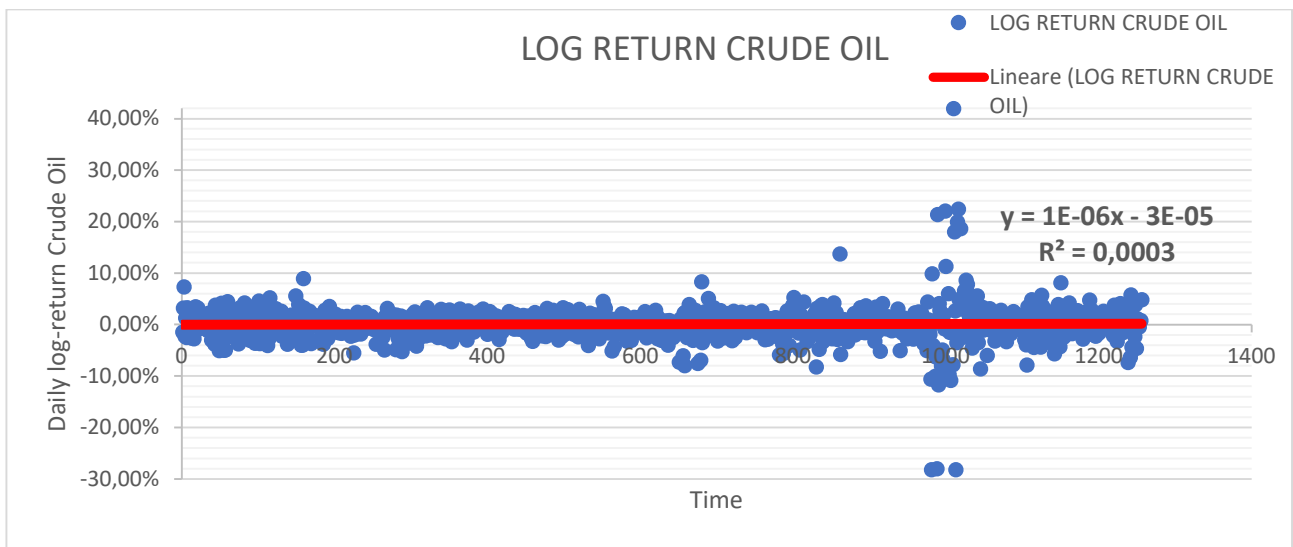


Figura 13. Regressione lineare log-rendimenti Petrolio

## RAME

L'andamento grafico della statistica Z per la commodity Rame suggerisce una differenza trascurabile tra i casi omoschedastico ed eteroschedastico. Questo è confermato dai valori più significativi delle variabili standardizzate. In entrambi i casi non si può rifiutare l'ipotesi nulla che i log-rendimenti siano correlati lungo diversi lag-temporali con un intervallo di fiducia del 95% ( $\alpha=5\%$ ), nonostante il p-value assuma un valore più alto nel caso eteroschedastico (12,85%).

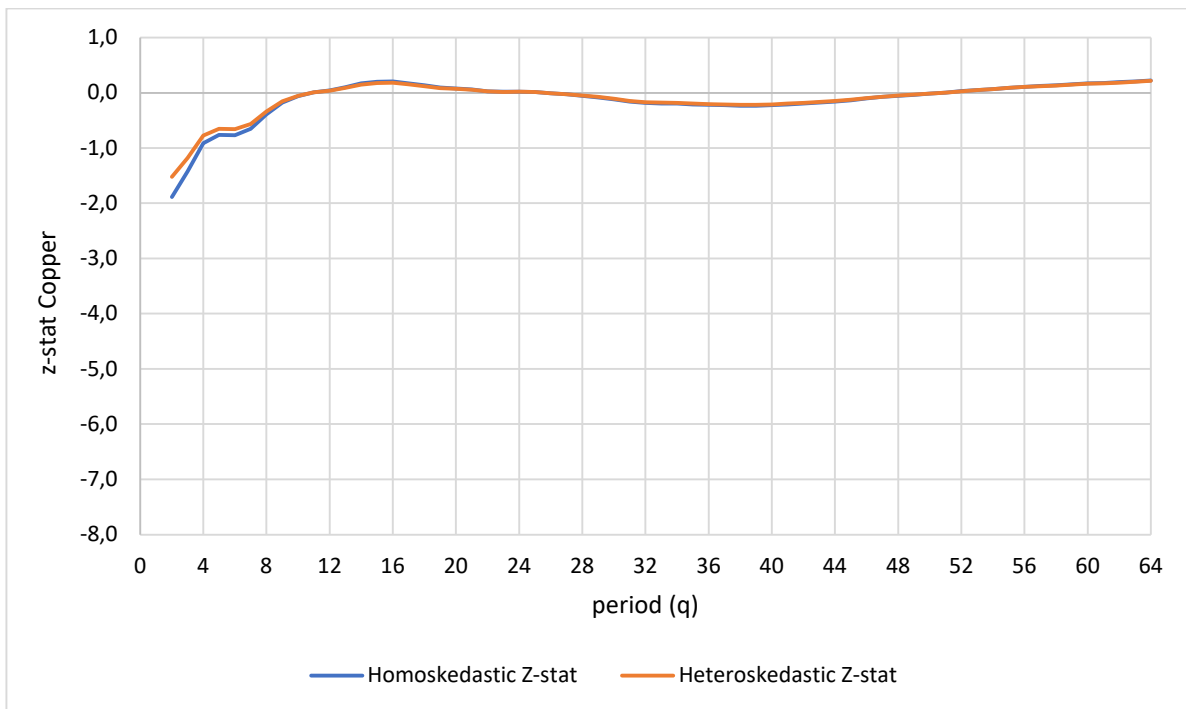


Figura 14. Variabile standardizzata "Z" del Variance Ratio relativa ai log-rendimenti del Rame nei casi Omoschedastico ed Eteroschedastico

Tabella 12. Valori più significativi del Test - Commodity Rame

COPPER	Homoscedastic	Heteroskedastic
Most significant Z-stat	-1,8850	-1,5200
p-value	5,94%	12,85%

Applicando il modello di regressione lineare sui log-rendimenti delle quotazioni giornaliere del rame, è emerso un valore dell' $R^2$  pari al 0,03%. Questo valore permette di considerare questo mercato efficiente in termini di correlazione tra i log-rendimenti. Quanto riportato è confermato dalla pendenza prossima allo zero della retta di regressione ( $6E-07$ ).

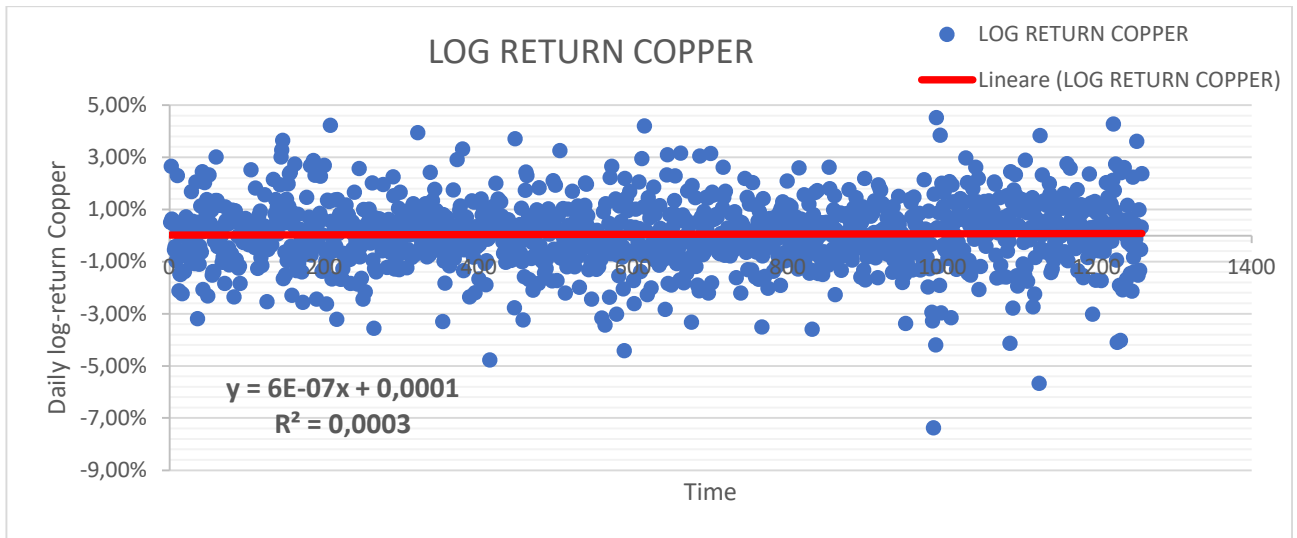


Figura 15. Regressione lineare log-rendimenti Rame

## SOIA

Dal Variance Ratio Test effettuato sulla commodity Soia, emerge che la variabile standardizzata segue quasi lo stesso percorso nei due casi homoschedastico ed eteroschedastico lungo lag-temporali differenti. Questo può portare a pensare con un'elevata probabilità che la distribuzione delle variabili non sia mutata lungo i cinque anni di osservazione. Inoltre, occorre sottolineare l'elevato p-value, che conferma l'efficienza del mercato per questa materia prima; come conseguenza, non si può rifiutare l'ipotesi nulla che la correlazione tra i log-rendimenti sia diversa da zero.

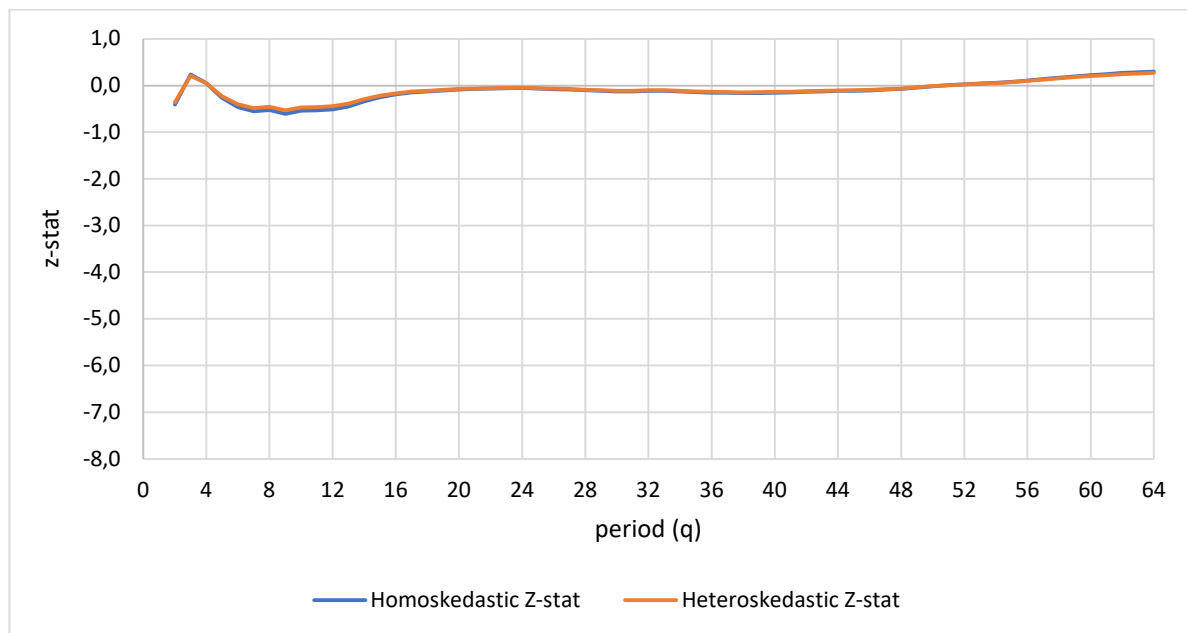


Figura 16. Variabile standardizzata "Z" del Variance Ratio relativa ai log-rendimenti della Soia nei casi Omoschedastico ed Eteroschedastico

Tabella 13. Valori più significativi del Test - Commodity Rame

SOYBEAN	Homoscedastic	Heteroskedastic
Most significant Z-stat	-0,6090	-0,5283
p-value	54,25%	59,73%

La regressione lineare evidenzia che l' $R^2$  della soia sia il più alto tra le *commodities* analizzate. Il suo valore è 0,18%. Trattandosi di un valore intorno allo zero, si può affermare comunque che il mercato si comporti in maniera efficiente in termini di correlazione tra i log-rendimenti.

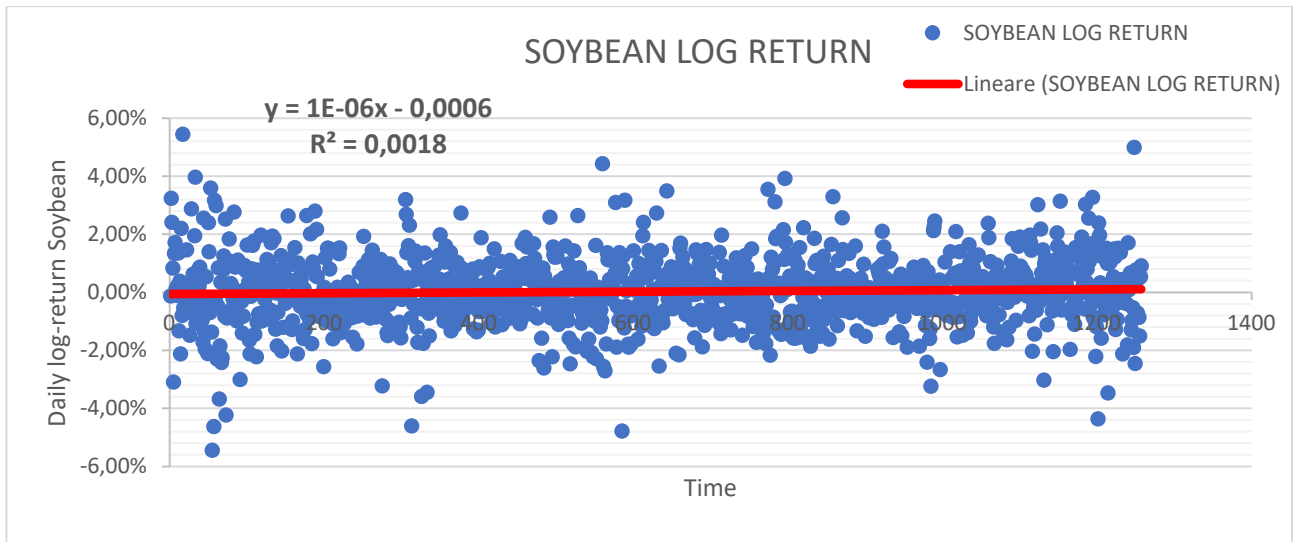


Figura 17. Regressione lineare log-rendimenti Soia

## 4. Reti neurali e Deep Learning

Prima di approfondire il funzionamento matematico delle reti neurali è opportuno introdurre il concetto di Percettrone.

Un Percettrone è la più semplice struttura di rete neurale; riceve in input diverse variabili binarie  $x_i$ , ciascuna caratterizzata da un peso  $w_i$  e genera un output binario, 0 o 1. Il risultato dell'output dipende dalla somma pesata di ciascuna variabile e da un valore soglia (*threshold value*).

$$output = \begin{cases} 0 & \text{if } \sum_i w_i x_i \leq threshold \\ 1 & \text{if } \sum_i w_i x_i > threshold \end{cases}$$

Apportando alcune opportune semplificazioni l'output si può scrivere come:

$$output = \begin{cases} 0 & \text{if } w \cdot x + b \leq 0 \\ 1 & \text{if } w \cdot x + b > 0 \end{cases}$$

dove  $b = -threshold$  e  $\sum_i w_i x_i$  è sostituito da  $w \cdot x$

Questa tipologia di neurone artificiale racchiude al suo interno dei limiti in termini di sensibilità dell'output al variare dei pesi delle variabili binarie.

Per superare questo problema e quindi permettere alle reti neurali di trovare un campo di applicazione più ampio, la teoria sulle reti neurali si è evoluta attorno ad un altro tipo di neurone artificiale, il *Sigmoid Neuron*.

A differenza dei percettroni, nel neurone sigmoide le variabili in input sono comprese tra 0 e 1 e l'output è rappresentato dalla funzione sigmoide, compresa tra 0 e 1:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

dove  $z = w \cdot x + b$ .

Per valori molto grandi oppure molto piccoli di  $z$  la funzione presenta lo stesso comportamento del percettrone. Tuttavia il vantaggio che deriva dall'utilizzo di questa tipologia di neurone viene riscontrato nella sensibilità dell'output al variare dei pesi  $w_i$  e della soglia  $b$ .

## 4.1. Architettura di una rete neurale

I neuroni sono i mattoni che costituiscono l'architettura di una rete neurale.

Generalmente, le architetture sono costituite da molteplici livelli: *Input layer*, *Hidden layers* e *Output layer*. A seconda di come i neuroni si scambiano le informazioni, si possono distinguere due categorie di reti neurali: feedforward e recurrent. Nelle prime l'output di un livello è utilizzato come input nel livello successivo e non sono consentiti loop comunicativi. Contrariamente, nelle reti neurali recurrent non è esclusa la possibilità di loop e sono caratterizzate da neuroni che possono attivarsi per periodi temporali limitati prima di diventare inattivi, stimolando in cascata l'attivazione di altri neuroni.

## 4.2. Gradiente discendente e algoritmo di Backpropagation

Lo scopo di allenare la rete neurale è quello di trovare un valore dei pesi  $w$  e bias  $b$  che minimizzino la funzione di costo quadratica  $C$ , definita come:

$$C(w, b) = \frac{1}{2n} \|y(x) - a\|^2$$

dove  $w$  rappresenta i pesi nella rete,  $b$  i bias,  $n$  gli input utilizzati come training della rete e  $a$  il vettore output della rete neurale; infine  $y(x)$  rappresenta le manifestazioni reali che vengono confrontate con l'output della rete neurale.

Uno degli algoritmi che consente la minimizzazione della funzione di costo è l'algoritmo del Gradiente Discendente.

La funzione di costo permette di capire come le variazioni nei pesi e nei bias siano collegate ad una diminuzione del costo totale. Generalizzando le variabili della funzione di costo con il vettore  $v = v_1, v_2.. v_n$ , e applicando l'algoritmo al caso in cui siano presenti solamente due variabili  $v_1$  e  $v_2$ , si può scrivere la variazione della funzione di costo  $\Delta C$  come:

$$\Delta C \approx \frac{\partial C}{\partial v_1} \Delta v_1 + \frac{\partial C}{\partial v_2} \Delta v_2$$

Successivamente, viene definito il gradiente della funzione come:

$$\nabla C \equiv \left( \frac{\partial C}{\partial v_1}, \frac{\partial C}{\partial v_2} \right)^T$$

Dove  $T$  indica la trasposta del vettore delle derivate parziali.

Successivamente possiamo scrivere che  $\Delta C \approx \nabla C * \Delta v$ ; da questa notazione si può dedurre che il gradiente permette di capire come scegliere  $\Delta v$  in modo da rendere  $\Delta C$  negativo.

Ponendo  $\Delta v = -\eta \nabla C$ , dove  $\eta$  è un parametro positivo chiamato *learning rate*, si ha che:

$$\Delta C \approx -\nabla C^T \eta.$$

Infine, aggiornando il parametro  $v$  con la regola  $v \rightarrow v' = v - \eta \nabla C$  si raggiunge iterativamente un minimo globale.

Le regole applicate a due variabili si possono estendere ad  $n$  variabili.

Applicando la medesima regola alla rete neurale precedentemente esposta, si può definire la regola di aggiornamento dei parametri  $w$  e  $b$ :

$$w_k \rightarrow w'_k = w_k - \eta \frac{\partial C}{\partial w_k}$$

$$b_l \rightarrow b'_l = b_l - \eta \frac{\partial C}{\partial b_l}$$

L'algoritmo può essere riadattato qualora il numero degli input sia molto elevato. In tal caso esiste una variante dell'algoritmo Gradiente Discendente, il Gradiente Discendente Stocastico, capace di ridurre la complessità computazionale dell'algoritmo con molteplici variabili in input.

Il Gradiente discendente stocastico estrae casualmente  $m$  tra le variabili in input etichettandole  $X_1, X_2, \dots, X_m$ . Per ciascun mini-batch ne calcola il gradiente. Al crescere di  $m$  si ottiene il valore del gradiente di  $C$  si avvicina a quello reale:

$$\frac{\sum_{j=1}^m \nabla C_{Xj}}{m} \approx \frac{\sum_x \nabla C_x}{n} = \nabla C$$

Conseguentemente i parametri  $w_k$  e  $b_l$  vengono aggiornati ed allenati analogamente:

$$w_k \rightarrow w'_k = w_k - \frac{\eta \sum_j C_{Xj}}{m \frac{\partial C}{\partial w_k}}$$

$$b_l \rightarrow b'_l = b_l - \frac{\eta \sum_j C_{Xj}}{m \frac{\partial C}{\partial b_l}}$$



Come passo successivo, viene scelto casualmente un altro mini-batch si ripete l'aggiornamento su un altro ciclo (può essere chiamata anche epoca) fino ad esaurire tutti le variabili training in input.

#### 4.2.1. Assunzioni e vincoli dell'algoritmo di Backpropagation

L'algoritmo di Backpropagation permette di calcolare il gradiente, quindi le derivate parziali  $\partial C / \partial w$  e  $\partial C / \partial b$ , con lo scopo finale di minimizzare la funzione di costo.

Per spiegare il funzionamento dell'algoritmo, occorre fare alcune assunzioni in termini di notazione. Si considera il termine  $w_{jk}^l$  il peso del collegamento dal k-esimo neurone del  $(l-1)$ -esimo livello verso il j-esimo neurone del  $l$ -esimo livello. Analogamente  $b_j^l$  rappresenta il bias del j-esimo nodo situato nel livello  $l$ . Infine  $a_j^l$  rappresenta la funzione di attivazione del j-esimo neurone nel livello  $l$ , la quale è influenzata dal risultato della medesima funzione al livello precedente:

$$a_j^l = \sigma \left( \sum_k w_{jk}^l a_k^{l-1} + b_j^l \right)$$

ponendo  $\sum_k w_{jk}^l a_k^{l-1} + b_j^l = z_j^l$ , si ottiene:

$$a_j^l = \sigma(z_j^l)$$

Riscrivendo in forma vettoriale la funzione di attivazione per un singolo layer:

$$a^l = \sigma(w^l a^{l-1} + b^l)$$

Ponendo  $w^l a^{l-1} + b^l = z^l$ , si trova:

$$a^l = \sigma(z^l)$$

Per applicare l'algoritmo è opportuno porre due assunzioni sulla funzione di costo riportata all'inizio del paragrafo:

1. La funzione di costo deve poter essere scritta come media di ciascuna componente  $C_x = \frac{1}{2} \|y - a^L\|^2$ , dove  $C_x$  rappresenta il costo di un singolo training input. Di conseguenza il costo totale può essere riscritto come  $C = \frac{1}{n} \sum_x C_x$
2. Il costo è funzione dell'output della rete neurale:  $C = C(a^L)$

Successivamente, è necessario introdurre la quantità  $\delta_j^l$ , rappresentante l'errore del j-esimo neurone al livello  $l$ , definita come:

$$\delta_j^l = \frac{\partial C}{\partial z_j^l}$$

La forma vettoriale dell'errore in singolo layer è  $\delta^l$ .

Dopo aver posto le opportune assunzioni, si può scrivere che l'errore è responsabile della variazione del costo totale di una quantità pari a  $\frac{\partial C}{\partial z_j^l} \Delta z_j^l$

## LE QUATTRO EQUAZIONI FONDAMENTALI DELL'ALGORITMO DI BACKPROPAGATION

L'algoritmo di Backpropagation è basato su quattro equazioni fondamentali che ne guidano il funzionamento logico e matematico:

1.  $\delta_j^l = \frac{\partial C}{\partial a_j^l} \sigma'(z_j^l)$ . La prima equazione evidenzia ciascuna componente dell'errore  $\delta^l$  in output.

Il primo termine,  $\frac{\partial C}{\partial a_j^l}$ , quantifica la velocità con cui varia il costo, in funzione dell'output del neurone j-esimo, mentre il secondo termine misura l'impatto di una modifica del termine  $z_j^l$  sulla funzione di attivazione.

Utilizzando l'operatore matematico  $\odot$ , il prodotto Hadamard<sup>3</sup>, l'equazione può essere riscritta come  $\delta^L = \nabla_a C \odot \sigma'(z^L)$

$\nabla_a C$  è il vettore delle derivate parziali della funzione di costo rispetto agli output di attivazione. Derivando la funzione di costo di un singolo layer per  $a_j^l$  si trova che

$$\frac{\partial C}{\partial a_j^l} = (a_j^l - y_j).$$

Generalizzando, si ottiene:

$$\delta^L = (a^L - y) \odot \sigma'(z^L)$$

2.  $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$ . La seconda equazione mette in relazione l'errore al livello  $l$ -esimo con l'errore al livello successivo. In questo caso  $(w^{l+1})^T$  rappresenta il vettore trasposto

---

<sup>3</sup> Il prodotto Hadamard, indicato con l'operatore  $\odot$ , compie l'operazione di moltiplicazione di due vettori o matrici aventi lo stesso numero di righe e colonne. Ad esempio, supponendo di avere due vettori  $a = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$  e  $b = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ , il risultato dell'operazione  $a \odot b$  è equivalente a:  $\begin{bmatrix} 3*1 \\ 2*4 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \end{bmatrix}$

dei pesi del livello successivo.

Intuitivamente, unendo le prime due equazioni siamo in grado di calcolare l'errore  $\delta^l$  per tutti i layers della rete neurale.

3.  $\frac{\partial C}{\partial b_j^l} = \delta_j^l$ . La terza equazione fornisce un ordine di grandezza del tasso di variazione del costo

totale rispetto a ciascun bias nella rete. L'equazione si può riscrivere come:  $\frac{\partial C}{\partial b} = \delta$

4.  $\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l$ . Infine la quarta equazione rappresenta il tasso di variazione del costo rispetto ai

pesi all'interno della rete. Ponendo  $a_k^{l-1} = a_{in}$  e  $\delta_j^l = \delta_{out}$  l'espressione può essere riscritta

come  $\frac{\partial C}{\partial w} = a_{in} \delta_{out}$ .

Dalle equazioni sopra riportate, si può dedurre che la rete neurale seguirà un processo di learning più lento al presentarsi di bassi valori di attivazione in input oppure per valori estremi, alti o bassi, della funzione di attivazione in output.

#### 4.2.2. Algoritmo di Backpropagation

All'interno della rete neurale, una variazione  $\Delta w_{jk}^l$  di uno dei pesi, modificherà il valore della

funzione di attivazione del neurone associato di una quantità  $\Delta a_j^l = \frac{\partial a_j^l}{\partial w_{jk}^l} \Delta w_{jk}^l$ . In cascata, i valori

delle funzioni di attivazione per i successivi layer cambieranno di  $\Delta a_q^{l+1} = \frac{\partial a_q^{l+1}}{\partial a_j^l} \Delta a_j^l$ .

Facendo le opportune sostituzioni, si può scrivere che la variazione della funzione di attivazione al layer successivo corrisponde a:

$$\Delta a_q^{l+1} = \frac{\partial a_q^{l+1}}{\partial a_j^l} \frac{\partial a_j^l}{\partial w_{jk}^l} \Delta w_{jk}^l$$

L'impatto della modifica di un peso avrà un impatto sulla funzione di costo totale per una quantità equivalente a:

$$\Delta C = \frac{\partial C}{\partial w_{jk}^l} \Delta w_{jk}^l$$

Considerando una rete neurale con L livelli, si può riscrivere la variazione della funzione di costo su un percorso di attivazione  $a_j^l, a_q^{l+1}, a_n^{l-1}, a_m^l$ :

$$\Delta C = \frac{\partial C}{\partial a_m^L} \frac{\partial a_m^L}{\partial a_n^{L-1}} \frac{\partial a_n^{L-1}}{\partial a_p^{L-2}} \dots \frac{\partial a_q^{L+1}}{\partial a_j^L} \frac{\partial a_{qj}^L}{\partial w_{jk}^L} \Delta w_{jk}^L$$

Considerando tutti i percorsi di attivazione, si può scrivere la variazione totale della funzione di costo come sommatoria della variazione lungo ciascun percorso:

$$\Delta C = \sum_{mnp..q} \frac{\partial C}{\partial a_m^L} \frac{\partial a_m^L}{\partial a_n^{L-1}} \frac{\partial a_n^{L-1}}{\partial a_p^{L-2}} \dots \frac{\partial a_q^{L+1}}{\partial a_j^L} \frac{\partial a_{qj}^L}{\partial w_{jk}^L} \Delta w_{jk}^L$$

Quest'ultima equazione ci consente di calcolare la derivata parziale della funzione di costo per il peso che unisce il k-esimo neurone del livello  $l-1$  e il j-esimo neurone del layer  $l$ :

$$\frac{\partial C}{\partial w_{jk}^L} = \sum_{mnp..q} \frac{\partial C}{\partial a_m^L} \frac{\partial a_m^L}{\partial a_n^{L-1}} \frac{\partial a_n^{L-1}}{\partial a_p^{L-2}} \dots \frac{\partial a_q^{L+1}}{\partial a_j^L} \frac{\partial a_{qj}^L}{\partial w_{jk}^L}$$

Dopo aver esposto le equazioni fondamentali e spiegato il modello da un punto di vista intuitivo, di seguito sono riportati i passi dell'algoritmo che minimizza la funzione di costo:

1. Inizialmente viene calcolato il valore della funzione di attivazione  $a^1$  per il primo layer in input.
2. Per tutti gli L strati della rete, si calcolano  $z^l = w^l a^{l-1} + b^l$  e  $a^l = \sigma(z^l)$ . Questo step prende il nome di *Feedforward*
3. Si calcola l'errore  $\delta^L$  utilizzando la prima equazione fondamentale  $\delta^L = \nabla_a C \odot \sigma'(z^L)$ .
4. Andando a ritroso nella rete si calcola l'errore  $\delta^l$  (*Backpropagation*) per ciascun livello utilizzando la seconda equazione fondamentale  $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$ .
5. Come ultimo passo, si calcola il gradiente della funzione di costo impiegando la terza e la quarta equazione fondamentale, rispettivamente  $\frac{\partial C}{\partial b_j^l} = \delta_j^l$  e  $\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l$ .

L'algoritmo di Backpropagation può essere generalizzato al caso in cui siano presenti molteplici dati in input. Ad esempio combinando l'algoritmo di Backpropagation al Gradiente Discendente Stocastico, siamo in grado di calcolare il gradiente per un mini-batch di  $m$  training input:

1. Disporre del set di training input
2. Per ciascun input, inizializzare la funzione di attivazione in input  $a^{x,1}$  e seguire i seguenti step:

- a. Per ciascun layer, calcolare  $z^{x,l} = w^l a^{x,l-1} + b^l$  e  $a^{x,l} = \sigma(z^{x,l})$ . Come esposto per l'algoritmo per un singolo training input, questo passo prende il nome di *Feedforward*
  - b. Si calcola l'errore  $\delta^{x,L}$  utilizzando la prima equazione fondamentale  $\delta^{x,L} = \nabla_a C_x \odot \sigma'(z^{x,L})$ .
3. Andando a ritroso nella rete si aggiornano i pesi della rete seguendo la regola  $w_k \rightarrow w'_k = w_k - \frac{\eta}{m} \frac{\sum_j C_{xj}}{\partial w_k}$  e i bias utilizzando l'analoga regola  $b_l \rightarrow b'_l = b_l - \frac{\eta}{m} \frac{\sum_j C_{xj}}{\partial b_l}$

Per completare l'algoritmo del gradiente discendente stocastico, le iterazioni dovranno avvenire per più epoche e su più mini-batch.

### 4.2.3. Funzione di costo Cross-entropy

La funzione di costo quadratica introdotta nel paragrafo precedente presenta un apprendimento lento delle informazioni a causa della forma piatta della funzione sigmoide per valori prossimi ad 1. La ragione per cui l'apprendimento risulta lento è suggerita dalla forma matematica della derivata parziale del costo rispettivamente per i pesi e i bias, che guida l'aggiornamento dei parametri della rete neurale:

$$\frac{\partial C}{\partial w} = (a - y)\sigma'(z)x$$

$$\frac{\partial C}{\partial b} = (a - y)\sigma'(z)$$

Come evidenziato dalla forma della funzione sigmoide (figura 18), per valori di output prossimi ad 1, la derivata della funzione è prossima a zero; di conseguenza anche la derivata parziale rispetto ai pesi e bias sarà bassa e quindi il tasso di apprendimento basso.

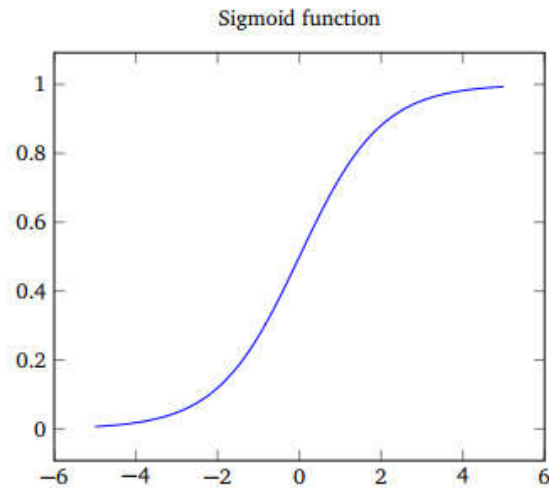


Figura 18. Funzione sigmoide

Per risolvere i problemi nella velocità dell'apprendimento della rete neurale è stata introdotta una funzione di costo cross-entropy, che per un singolo neurone è definita come:

$$C = \frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)]$$

in cui  $n$  è il numero totale delle variabili training,  $x$  è ciascuna variabile in input e  $y$  l'output desiderato.

La funzione di costo cross-entropy, analogamente alla funzione di costo quadratica, non è mai negativa e tende a zero per valori di output generati tendenti a quelli desiderati.

I valori delle derivate parziali rispetto ai pesi e ai bias sono rispettivamente:

$$\frac{\partial C}{\partial w_j} = \frac{1}{n} \sum_x x_j (\sigma(z) - y)$$

$$\frac{\partial C}{\partial b} = \frac{1}{n} \sum_x (\sigma(z) - y)$$

Le due espressioni suggeriscono che per errori di output più grandi il tasso di apprendimento della rete aumenta. Il vantaggio rispetto alla funzione di costo quadratica è intuitivamente collegato anche dall'eliminazione del termine  $\sigma'(z)$  dal calcolo della derivata parziale e di conseguenza suoi valori piccoli non influenzano negativamente l'apprendimento della rete.

Generalizzando i calcoli per più neuroni, organizzati su  $L$  livelli, la funzione di costo cross-entropy diventa:

$$\frac{1}{n} \sum_j [y \ln a_j^L + (1 - y) \ln(1 - a_j^L)]$$

### 4.3. OVERFITTING

L'Overfitting è un fenomeno che si verifica nei modelli statistici e informatici quando il modello a disposizione si adatta eccessivamente ai dati di apprendimento su cui è stato costruito il modello. Questo non ci consente di impiegare il nostro algoritmo su dati nuovi, in quanto il modello ha imparato le informazioni e il "rumore di fondo" dei dati training. Il fenomeno si intensifica al crescere dei parametri liberi.

Al fine di prevenire l'Overfitting esistono diversi metodi:

- Early stopping: questo metodo punta a fermare l'apprendimento della rete prima che la stessa impari i rumori di fondo che avrebbero un impatto negativo sulla valutazione di dati al di fuori del data set di apprendimento. Ad esempio, utilizzando un set di dati per la validazione del modello per testare l'*accuracy*, si interrompe l'apprendimento della rete una volta che l'*accuracy* del set di validazione si satura.
- Disporre di un data set contenente più dati per l'apprendimento per captare le relazioni fondamentali tra i neuroni della rete.
- Eliminare le caratteristiche ridondanti
- Regolarizzazione: permette di ridurre l'Overfitting mantenendo inalterata la dimensione della rete neurale.
  - o Regolarizzazione L2 (*Weight Decay*)

La regolarizzazione Weight Decay aggiunge il parametro  $\frac{\lambda}{2n} \sum_w w^2$  alla funzione di costo, dove  $\lambda$  è il parametro di regolarizzazione,  $n$  la dimensione del training set e  $\sum_w w^2$  la somma dei pesi della rete al quadrato.

La funzione di costo regolarizzata assume perciò la seguente forma matematica:

$$C = C_0 + \frac{\lambda}{2n} \sum_w w^2$$

Dove  $C_0$  può essere sostituito sia dalla funzione di costo quadratica o cross-entropy

- **Regolarizzazione L1**

Questo approccio aggiunge alla funzione di costo il termine  $\frac{\lambda}{n} \sum_w |w|$ , perciò si ottiene:

$$C = C_0 + \frac{\lambda}{n} \sum_w |w|$$

Entrambi i metodi di regolarizzazione L1 e L2 tendono a restringere i pesi della rete. Quando il parametro  $\lambda$  presenta un valore basso, è preferibile minimizzare la funzione di costo, mentre con  $\lambda$  alto sarà preferibile aggiornare la rete con pesi bassi.

- **REGOLARIZZAZIONE DROPOUT**

La tecnica di regolarizzazione dropout modifica la struttura della stessa eliminando casualmente alcune connessioni neurali dei livelli nascosti e valutando le diverse reti neurali generate in maniera separata. Successivamente le informazioni raccolte dalle reti vengono aggregate e valutate nel loro insieme, rendendo la rete di partenza robusta alla perdita di diversi neuroni.



## 5. LSTM applicata al prezzo spot delle *commodities*

In questo capitolo è esposto il funzionamento delle architetture *Long Short Term Memory* (LSTM). Le reti neurali LSTM sono una sottocategoria delle RNN (Recurrent Neural Network) e sono state introdotte da Hochreiter e Schmidhuber nel 1997. La caratteristica che contraddistingue questo tipo di architetture rispetto alle RNN è la capacità di evitare le dipendenze a lungo termine. Conservando e selezionando le informazioni lungo una serie attraverso delle connessioni di feedback, oltre alle classiche connessioni feedforward delle RNN, la rete LSTM elabora i dati sequenzialmente memorizzando le proprietà importanti degli stati precedenti, che saranno impiegate come input dello stadio successivo.

Una rete LSTM presenta uno stato della cella, che ha il ruolo di memorizzare le informazioni all'interno dell'architettura. La cella ha come input lo stato della cella al momento precedente e i dati in input dello stadio corrente.

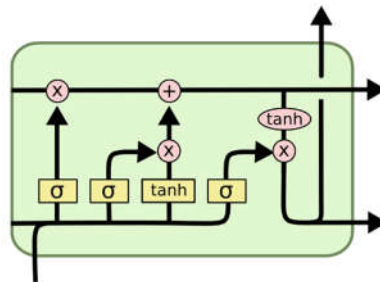


Figura 19. Architettura LSTM

La capacità della rete di aggiungere o rimuovere informazioni allo stato della cella è regolata dalle porte (gates).

Come primo passo, lo stato precedente e i dati di input dello stato corrente entrano nel “*forget gate layer*”, dove vengono separate le informazioni importanti da quelle che non lo sono. Questa decisione viene elaborata da una funzione di attivazione sigmoide, che genera valori tra 0 e 1:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) + b_f$$

Un valore più vicino a 1 viene associato ad un elemento di rilevante importanza, mentre un valore prossimo allo 0 è associato ad un elemento irrilevante per la cella.

Successivamente, i valori generati dalla funzione sigmoide vengono moltiplicati puntualmente per lo stato della cella precedente, quindi le componenti della cella aventi un valore della funzione  $f_t$  basso avranno un minor peso sui passaggi successivi.

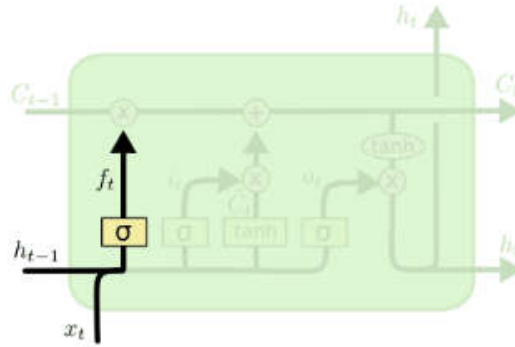


Figura 20. LSTM forget gate

Dopo questo passaggio è opportuno trovare le informazioni rilevanti per il nuovo stato della cella. Nello specifico, i dati in input nella cella vengono doppiamente elaborati da due operatori:

- il nuovo vettore di memoria  $\tilde{C}_t$ : è una rete neurale ottenuta applicando come funzione la tangente iperbolica ai dati input. L'output dell'operatore  $\tanh$  è compresi tra -1 e 1:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t]) + b_C$$

- l'input gate identifica quali dei valori del nuovo vettore di memoria debbano essere salvati applicando la funzione di attivazione sigmoide:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t]) + b_i$$

Il risultato dell'input gate e del nuovo vettore di memoria vengono moltiplicati puntualmente e lo stato della cella  $C_t$  aggiornato:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

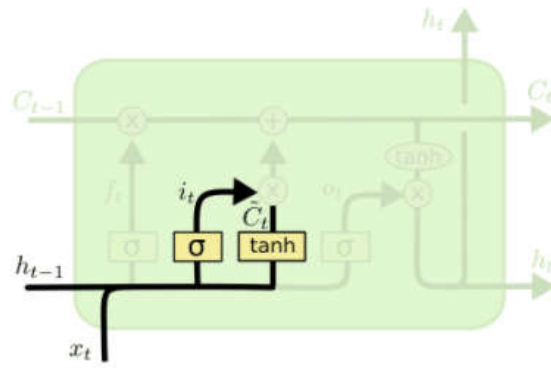


Figura 21 Nuovo vettore di memoria e Input Gate

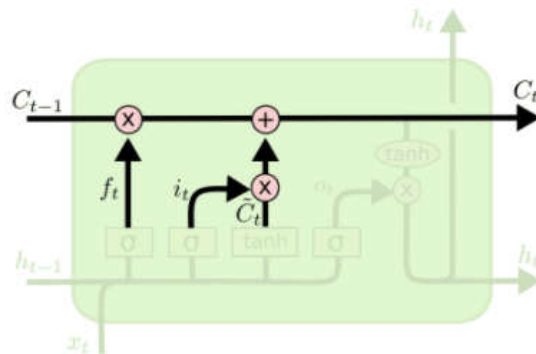


Figura 22. Aggiornamento stato della cella

Infine, l'ultima iterazione si occupa di generare l'output del nuovo stato nascosto, elaborando i dati in input dello stato precedente con una funzione di attivazione sigmoide:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t]) + b_o$$

Successivamente viene applicato l'operatore tanh allo stato aggiornato della cella  $C_t$  in modo da normalizzare i dati tra -1 e 1.

Moltiplicando l'output della funzione di attivazione sigmoide per  $\tanh(C_t)$ , si trova l'output dello stato nascosto  $h_t$ :

$$h_t = o_t * \tanh(C_t)$$

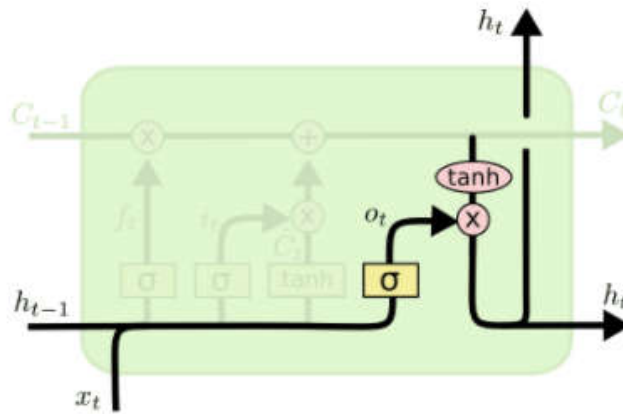


Figura 23. Output gate

## 5.1. LSTM per il forecasting dei prezzi delle *commodities*

Le proprietà menzionate nel paragrafo precedente permettono all'architettura LSTM di trovare applicazione nell'analisi di lunghe serie di dati, come l'andamento del prezzo di un titolo.

Per ciascuna delle materie prime analizzate nei capitoli precedenti (cotone, oro, petrolio greggio, rame e soia), si è costruita un'architettura LSTM per il forecasting delle quotazioni.

L'input per la costruzione del modello è il dataset dei prezzi di chiusura lungo un arco temporale di 5 anni; questi dati sono stati recuperati dal sito del Nasdaq, al netto dei prezzi della commodity oro, disponibili su [www.borsainside.com](http://www.borsainside.com):

- COTONE: <https://www.nasdaq.com/market-activity/commodities/tt:nmx/historical>
- ORO: <https://www.borsainside.com/quotazioni/quotazione-oro/>
- PETROLIO GREGGIO: <https://www.nasdaq.com/market-activity/commodities/cl%3Anmx/historical>
- RAME: <https://www.nasdaq.com/market-activity/commodities/hg%3Acmx/historical>
- SOIA: <https://www.nasdaq.com/market-activity/commodities/zs/historical>

Come passo successivo, dopo aver opportunamente normalizzato i dati tra 0 e 1, il dataset è stato suddiviso in *train data*, *validation data* e *test data*. Il 70% dei dati è stato impiegato per l'allenamento della rete neurale, il 20% per validare l'architettura e il restante 10% dei dati è stato impiegato per valutare il modello in termini di capacità predittiva.

Le metriche impiegate per stimare l'accuratezza degli output sono lo scarto quadratico medio, l'errore assoluto medio e l'errore assoluto medio percentuale.

Queste metriche sono implementate nel modello tramite la libreria open source Scikit-learn, contenente al suo interno algoritmi di apprendimento automatico per il linguaggio di programmazione Python

Il calcolo dello scarto quadratico medio (MSE) sulle previsioni della rete neurale è calcolato dal metodo `mean_squared_error` come la sommatoria delle differenze al quadrato tra i valori predetti e quelli reali, divisa per la dimensione del vettore delle osservazioni:

$$MSE = \frac{\sum_{i=0}^n (x_{reale_i} - x_{prev_i})^2}{n}$$

Analogamente l'errore assoluto medio (MAE) viene calcolato con il metodo `mean_absolute_error` ed è equivalente alla sommatoria del valore assoluto delle differenze tra i valori predetti e quelli reali, divisa per la dimensione del vettore delle osservazioni:

$$MAE = \frac{\sum_{i=0}^n |x_{reale_i} - x_{prev_i}|}{n}$$

Infine, per avere un ordine di grandezza in termini percentuali, si è calcolato il Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{\sum_{i=0}^n \left| \frac{x_{reale_i} - x_{prev_i}}{x_{reale}} \right|}{n}$$

### 5.1.1. Pandas

Pandas è una libreria open source sviluppata a partire dal 2008 da Wes McKinney, allora impiegato presso AQR Capital Management; Pandas è scritta nel linguaggio di programmazione Python e viene utilizzata come strumento nelle analisi quantitative e nella gestione dei dati. I metodi messi a disposizione da questo modulo nell'analisi dei dati finanziari permettono di recuperare i dati da fonti molteplici come SQL, CSV, Excel e di compiere le opportune azioni di trasformazione, indicizzazione e filtraggio su tabelle e serie temporali. Nel modello implementato per l'andamento del prezzo delle *commodities*, la libreria è stata utilizzata per importare il dataset e prendere come riferimento il prezzo di chiusura e la data dell'osservazione *i*-esima.

### **5.1.2. Numpy**

Per eseguire calcoli e operazioni su vettori e matrici dei dati finanziari è stato utilizzato il modulo open source NumPy, creato nel 2005 da Travis Oliphant e distribuito sotto licenza BSD. L'interfaccia Array, messa a disposizione dalla libreria, permette di svolgere svariate operazioni su strutture vettoriali e matriciali con un grande vantaggio in termini di complessità computazionale rispetto alle liste di Python.

### **5.1.3. Keras e Tensorflow**

Al fine di implementare l'architettura LSTM nel linguaggio Python, è stato impiegato il software open source Keras.

Keras è stata inventata nel 2015 da François Chollet ed è un'interfaccia di programmazione delle applicazioni (API) per il deep-learning scritta nel linguaggio di programmazione Python per l'apprendimento automatico delle reti neurali e lo sviluppo di reti neurali profondo e supporta come librerie di back-end TensorFlow, Theano e CNTK. Nella creazione della rete LSTM è stata utilizzata la libreria TensorFlow, creata dal Google Brain Team nel 2015 per la risoluzione dei problemi comuni di machine learning e deep learning.

## **5.2. FUNZIONE DI ATTIVAZIONE**

Uno dei principali parametri impiegati per confrontare i risultati generati dalle diverse architetture neurali è stato la funzione di attivazione.

Nel capitolo 4 era stata introdotta la funzione di attivazione sigmoide. Tuttavia la libreria Keras mette a disposizione molteplici funzioni di attivazione dei neuroni come la tangente iperbolica (TANH) e la Rectified Linear Unit (RELU).

La TANH è una funzione differenziabile, monotona e presenta un dominio di valori compresi tra -1 e +1. La sua derivata non è differenziabile.

La ReLU è la funzione di attivazione più utilizzata nel deep learning e nelle RNN e il suo dominio varia tra 0 e  $+\infty$ . È una funzione monotona e differenziabile ed efficiente dal punto di vista computazionale. Il suo più grande vantaggio è quello di superare il problema del gradiente di fuga.

La differenza tra queste due funzioni risiede nella capacità di leggere gli input negativi della rete. La TANH mapperà gli input negativi con dei valori negativi, mentre la ReLU assegnerà ai medesimi input un valore nullo.

Il problema appena esposto è superato implementando una variante dell'architettura ReLU, la Leaky ReLU, in grado di allargare il dominio dei valori possibili tra  $-\infty$  e  $+\infty$ .

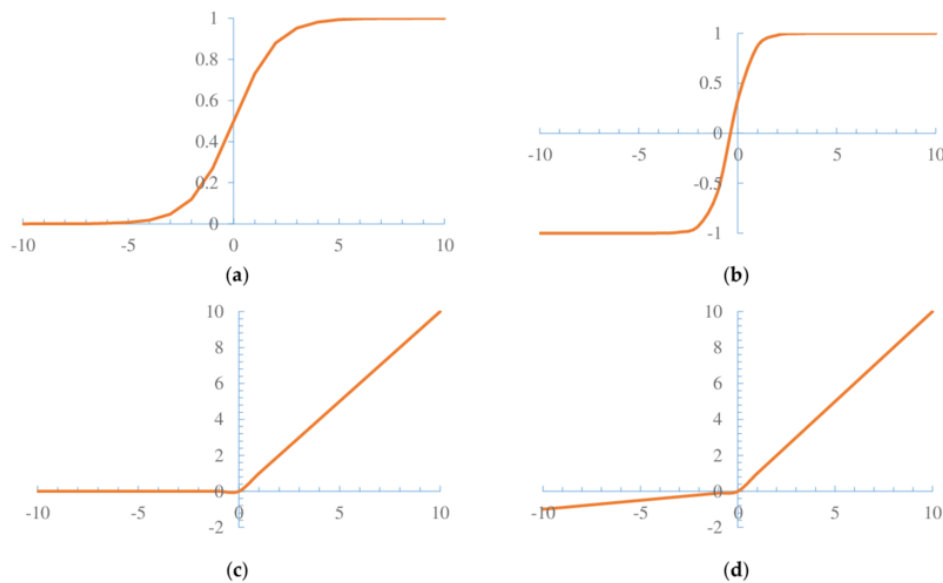


Figura 24. Sigmoid (a), TANH (b), ReLU (c) e Leaky ReLU (d) a confronto

Source: Researchgat.net

### 5.3. ALGORITMO DI OTTIMIZZAZIONE

La scelta dell'algoritmo di ottimizzazione è un aspetto fondamentale nell'addestramento di una rete neurale.

La libreria open source Keras propone molteplici algoritmi per l'addestramento della rete:

- SGD
- RMSprop
- Adam
- Adadelta
- Adagrad
- Adamax
- Nadam
- Ftrl

Nello studio del prezzo delle commodity ci si è soffermati sulla differenza in termini di efficienza predittiva tra gli algoritmi SGD, già affrontato nel capitolo 4, e Adam (Adaptive Moment Estimation).

L'algoritmo Adam è un'estensione dell'algoritmo SGD ed è il più utilizzato per la sua bassa complessità computazionale ed efficienza nell'aggiornamento dei pesi della rete.

Tale risultato è raggiunto combinando i vantaggi in termini di aggiornamento dei learning rate di due altre estensioni dell'algoritmo SDG: AdaGrad e RMSprop.

## 5.4. LEARNING RATE

Il learning rate è un iperparametro delle reti neurali che controlla la velocità di apprendimento del modello. Utilizzando come parametro la quantità di errore stimato derivata dall'aggiornamento dei pesi aggiornati, il learning rate guida l'aggiornamento dei pesi e di conseguenza l'apprendimento dell'architettura neurale.

Il suo valore è compreso nell'intervallo [0:1]; un valore di learning rate alto permette al modello di apprendere i dati in maniera più veloce, con il rischio di trovare delle soluzioni sub-ottimali. Contrariamente, un basso learning rate permettere al modello di trovare delle soluzioni ottimali, ma con tempi di elaborazione più elevati.

Le funzioni di ottimizzazione SGD e Adam implementano di default in Python un learning rate pari a 0,001.

Per testare l'algoritmo in termini di capacità predittiva si è deciso di variare il Learning rate di un ordine di grandezza  $\pm 10^1$ . Quindi sono stati testati i seguenti valori di LR: 0,01, 0,001, 0,0001



## 6. Risultati forecasting *commodities*

In questo paragrafo verranno esposti i risultati dei test compiuti sulle quotazioni nel mercato delle commodity.

Per ciascuna materia prima, partendo da un'architettura LSTM di base (Tabella n. 14), in prima battuta si è studiato l'output in termine di errore al variare della funzione di attivazione in input. Come riferimento per i test sono state utilizzate le funzioni di attivazione: Sigmoide, Tangente Iperbolica, e ReLU.

L'architettura è stata allenata alternando i due algoritmi di ottimizzazione SGD e Adam, modificando iterativamente il learning rate, come riportato nel paragrafo 6.3.

Successivamente, sono state effettuate delle prove modificando il numero di nodi all'interno della rete e eliminando uno dei livelli nascosti.

Come parametri di riferimento sono stati calcolati il RMSE, il MAE e il MAPE per ciascun test effettuato.

*Tabella 14 Architettura LSTM di base*

Model	Layer Type	Output Shape	Parametri
Livello 1	LSTM	(None, 100, 50)	10400
Livello 2	LSTM	(None, 100, 50)	20200
Livello 3	LSTM	(None,50)	20200
Livello 4	Dense	(None,50)	2550
Livello 5	Dense	(None,1)	51

*Tabella 15. Architettura LSTM 25 nodi per livello*

Model	Layer Type	Output Shape	Parametri
Livello 1	LSTM	(None, 100, 25)	2700
Livello 2	LSTM	(None, 100, 25)	5100
Livello 3	LSTM	(None,25)	5100
Livello 4	Dense	(None,25)	650
Livello 5	Dense	(None,1)	26

*Tabella 16. Architettura LSTM 4 livelli*

Model	Layer Type	Output Shape	Parametri
Livello 1	LSTM	(None, 100, 50)	10400
Livello 2	LSTM	(None, 100, 50)	20200
Livello 3	LSTM	(None,50)	20200
Livello 4	Dense	(None,1)	51

## 6.1. COMMODITY ORO

I risultati del test effettuati sulla materia prima ORO hanno rivelato che la funzione di attivazione ottimale in termini di errore medio percentuale è la tangente iperbolica (MAPE = 0,84%). Tale risultato è stato raggiunto applicando il learning rate di default dell' algoritmo (Lr=0,001).

L' algoritmo di ottimizzazione più efficiente per l' allenamento della rete neurale è Adam, mentre il gradiente stocastico discendente mostra errori sistematicamente maggiori rispetto al primo.

Per questa ragione la struttura neurale è stata modificata e successivamente testata impiegando l' algoritmo Adam. In prima battuta sono stati dimezzati i nodi nella rete per ciascun livello, con un errore medio percentuale in output oscillante tra 0,9% e 1,1%.

Infine è stata valutata la robustezza del modello eliminando uno degli hidden layer; quest' ultima struttura a quattro livelli ha generato un ottimo risultato in termini di capacità predittiva (MAPE 0,92%).

La tabella n. 17 riporta la previsione dei prezzi della commodity negli ultimi 26 giorni del range temporale analizzato, mentre la figura n. 26 mostra graficamente le oscillazioni rispetto ai valori reali dei valori output generati in output dal modello.



Figura 25. Prezzo Oro in \$/oz- 03/2016 - 03/2021

Tabella 17. Forecasting Commodity Oro

<b>Indice</b>	<b>Valori Reali</b>	<b>OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001</b>	<b>OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001</b>	<b>OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Nodi per livello=25</b>	<b>OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Rete neurale 4 livelli</b>
1	1807,30	1799,00	1786,57	1788,35	1789,22
2	1798,80	1789,85	1780,81	1776,26	1781,19
3	1758,30	1784,79	1773,68	1773,64	1778,59
4	1733,00	1745,39	1738,19	1734,01	1741,45
5	1714,40	1728,46	1717,88	1725,82	1728,30
6	1731,40	1710,33	1701,74	1707,94	1709,70
7	1717,10	1728,91	1717,31	1730,96	1728,05
8	1692,70	1708,23	1701,54	1704,04	1706,47
9	1698,20	1688,22	1678,93	1689,55	1690,28
10	1686,70	1695,83	1685,14	1698,51	1697,74
11	1710,60	1680,03	1672,96	1679,64	1681,38
12	1733,00	1705,74	1694,62	1709,83	1707,31
13	1715,00	1720,50	1712,71	1717,84	1719,12
14	1725,80	1700,04	1694,23	1696,05	1699,59
15	1735,20	1715,11	1705,07	1715,53	1715,35
16	1734,00	1719,84	1712,92	1716,16	1717,65
17	1748,70	1718,22	1710,91	1715,62	1716,33
18	1732,10	1733,61	1724,29	1731,69	1731,02
19	1743,90	1714,87	1708,69	1709,97	1712,32
20	1736,00	1731,93	1720,56	1732,75	1730,96
21	1730,70	1720,66	1713,08	1716,41	1718,96
22	1733,50	1719,41	1708,83	1719,83	1720,44
23	1726,00	1722,70	1712,00	1721,92	1723,79
24	1731,60	1715,02	1705,09	1714,07	1717,34
25	1705,80	1722,65	1710,68	1723,34	1725,80
26	1679,70	1695,41	1686,82	1692,88	1699,30

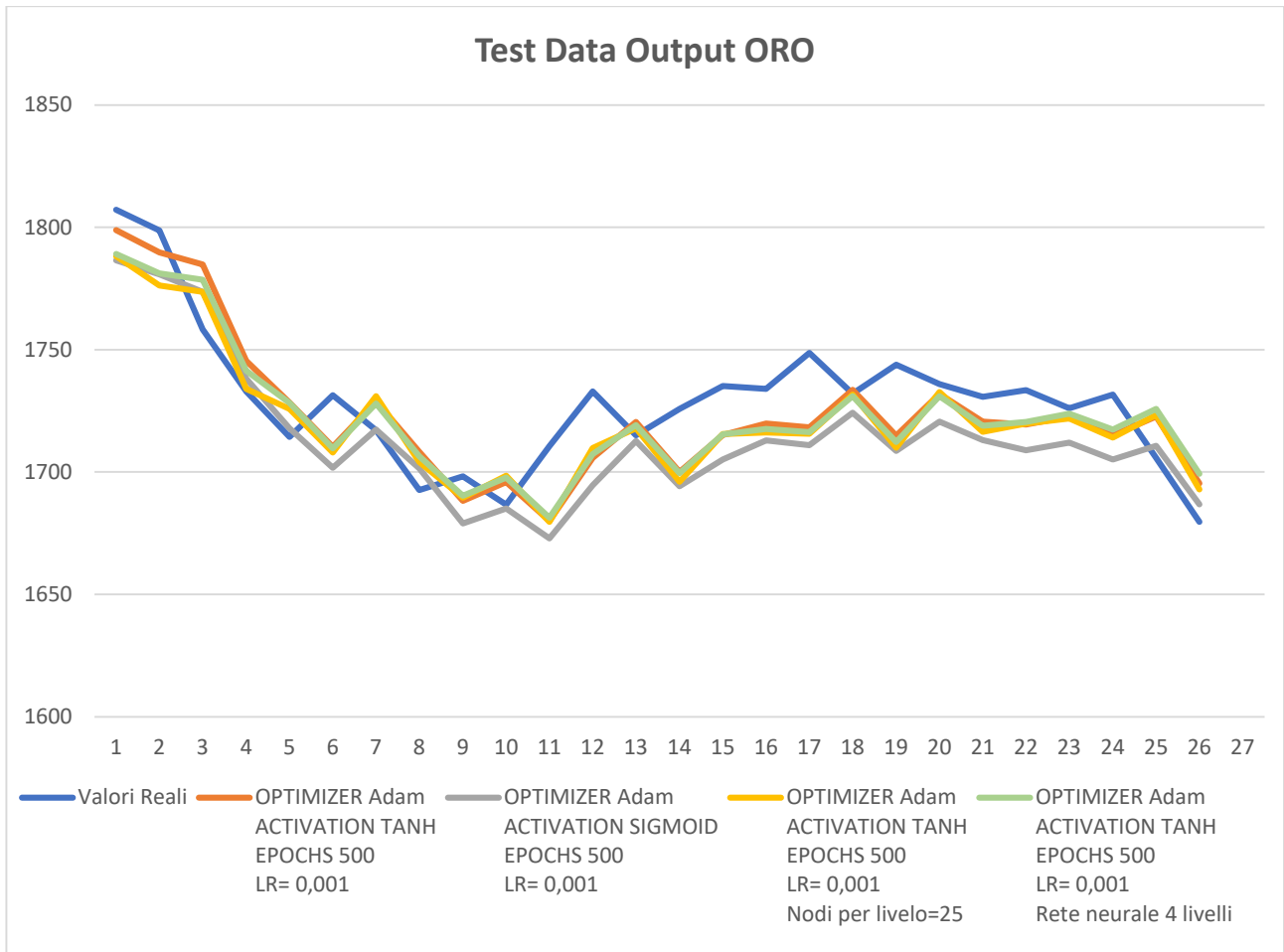


Figura 26. Plot Forecasting commodity Oro

Epoch=500 Optimizer Adam Learning rate=0,01	RMSE	MAE	MAPE
Activation = ReLU	109.7745784367829	105.68045537861494	6.100706883869395
Activation = Sigmoid	102.20650237633635	100.18805461434218	5.777512610638727
Activation= Tanh	38.93160596575173	30.75399057681744	1.7620442391219844

Epoch=500, Optimizer=Adam Learning Rate=0,001	RMSE	MAE	MAPE
Activation = 'ReLU'	86.9616681570858	85.28473586990283	4.921661593439235
Activation ='Sigmoid'	21.894717180688904	18.2475211253533	1.051608094713676
Activation= 'Tanh'	17.50725137426841	14.5146178488548	0.8429774134318896

Epoch=500, Optimizer=Adam Learning Rate=0,0001	RMSE	MAE	MAPE
Activation = ReLU	138.7950713877031	137.1855833289715	7.924438163208716
Activation = Sigmoid	49.91332613449833	40.17279261809129	2.335225059341408
Activation= Tanh	24.897077697488488	21.46047891332553	1.2374305186247945

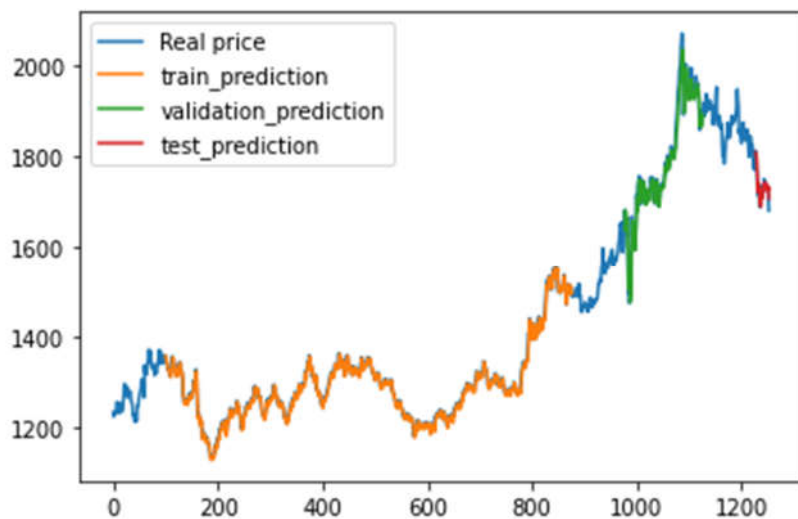
<b>Epoch=500, Optimizer=SGD Learning rate=0,01</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	68.39345423735315	59.42807434613888	3.4560758258873414
<b>Activation = Sigmoid</b>	157.60951922061568	155.09807006670877	8.950562304043592
<b>Activation= Tanh</b>	36.307171154711654	29.05770227083793	1.6912387331674348

<b>Epoch=500, Optimizer=SGD Learning rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	397.62503835267495	396.6726936451515	22.920047116346876
<b>Activation = Sigmoid</b>	430.464759241785	429.5527937801983	24.821462627943987
<b>Activation= Tanh</b>	329.6129138893652	328.5048943127577	18.97801182226369

<b>Epoch=500, Optimizer=SGD Learning rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	451.74960409574544	450.85355399319735	26.05322457043298
<b>Activation = Sigmoid</b>	440.4928752236189	439.59103708484975	25.40197340224415
<b>Activation= Tanh</b>	445.2078920040677	444.30974937001076	25.67484709757887

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	282.1125915651633	256.2120309701332	14.811519436488624
<b>Activation = Sigmoid</b>	21.277350124974394	17.27472593004887	1.00226077231721
<b>Activation= Tanh</b>	18.465396652222644	15.906786973659809	0.9178194306550098

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Architettura 4 livelli</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation= Tanh</b>	18.11255850062155	16.003262049876728	0.9242505302690501



*Figura 27. Plot Funzione di attivazione TANH, optimizer Adam, Learning Rate= 0,01, 500 Epoche - COMMODITY ORO*

## 6.2. COMMODITY RAME

L'analisi delle quotazioni del Rame ha evidenziato che l'*optimizer* Adam è in grado di generare sistematicamente valori di errori di previsione più bassi rispetto al gradiente discendente stocastico (SGD). Nello specifico, applicando l'algoritmo Adam con il suo valore di learning rate di default (Lr=0,001) in corrispondenza delle funzioni di attivazione sigmoide e tangente iperbolica si sono ottenuti rispettivamente i valori di MAPE pari a 1,09% e 1,28%.

Dopo aver effettuato delle prove con learning rate pari a 0,001 e 0,0001, dati gli alti valori di MSE e MAE, si è deciso di eliminare SGD e utilizzare l'algoritmo di ottimizzazione Adam per reti neurali con una diversa struttura.

Per verificare la robustezza del modello, in prima battuta sono stati dimezzati il numero di nodi per ciascun layer, ottenendo errori percentuali medi al di sotto del 2%.

Per la materia prima Rame, le prestazioni della funzione di attivazione sigmoide sono superiori rispetto alla tangente iperbolica. Tuttavia, data la lieve differenza in termini di capacità predittiva e la minore complessità computazionale della tangente iperbolica sono state ripetuti i test alternando entrambe le funzioni di attivazione. Come conferma, eliminando dalla rete neurale il livello nascosto Dense (50), è stato generato un errore percentuale MAPE pari al 2,11% con l'impiego della funzione sigmoide e del 1,59% invece utilizzando la tangente iperbolica.

La tabella n. 18 riporta la previsione dei prezzi della commodity negli ultimi 27 giorni del range temporale analizzato, mentre la figura n. 29 mostra graficamente le oscillazioni rispetto ai valori reali dei valori output generati in output dal modello.

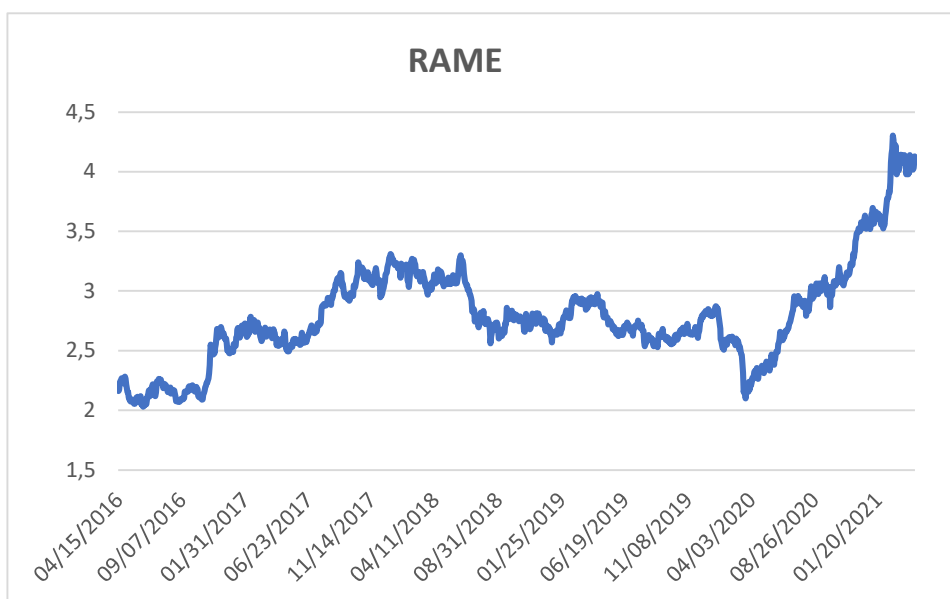


Figura 28. Prezzo Rame in \$/lb 04/2016-04/2021



Tabella 18. Forecasting commodity Rame

<b>Indice</b>	<b>Valori Reali</b>	<b>OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001</b>	<b>OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001</b>	<b>OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Nodi per livello=25</b>	<b>OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001 Nodi per livello=25</b>	<b>OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Rete neurale 4 livelli</b>
1	4,09	4,10	4,04	4,05	4,01	4,03
2	4,01	4,10	4,07	4,01	4,05	4,03
3	4,03	3,99	4,01	3,91	3,97	3,93
4	4,14	4,04	4,02	3,99	3,99	3,98
5	4,14	4,17	4,11	4,11	4,08	4,10
6	4,14	4,14	4,13	4,07	4,10	4,08
7	4,07	4,14	4,13	4,07	4,09	4,08
8	4,12	4,06	4,07	3,98	4,03	3,99
9	4,11	4,14	4,10	4,07	4,06	4,07
10	4,11	4,11	4,10	4,03	4,06	4,04
11	4,14	4,12	4,10	4,05	4,07	4,05
12	4,08	4,15	4,13	4,08	4,09	4,08
13	4,06	4,07	4,08	3,99	4,04	4,00
14	3,98	4,07	4,06	4,00	4,02	4,00
15	4,07	3,97	3,98	3,90	3,95	3,90
16	4,03	4,10	4,05	4,05	4,02	4,03
17	3,98	4,05	4,04	3,97	4,00	3,97
18	4,00	3,98	3,99	3,92	3,95	3,92
19	3,99	4,02	4,00	3,96	3,96	3,95
20	4,14	4,01	4,00	3,95	3,96	3,94
21	4,12	4,19	4,12	4,13	4,09	4,11
22	4,05	4,13	4,12	4,05	4,08	4,06
23	4,09	4,05	4,06	3,99	4,02	3,99
24	4,04	4,12	4,08	4,05	4,05	4,05
25	4,02	4,04	4,04	3,97	4,01	3,97
26	4,03	4,03	4,02	3,96	3,98	3,96
27	4,13	4,05	4,03	3,98	3,99	3,98

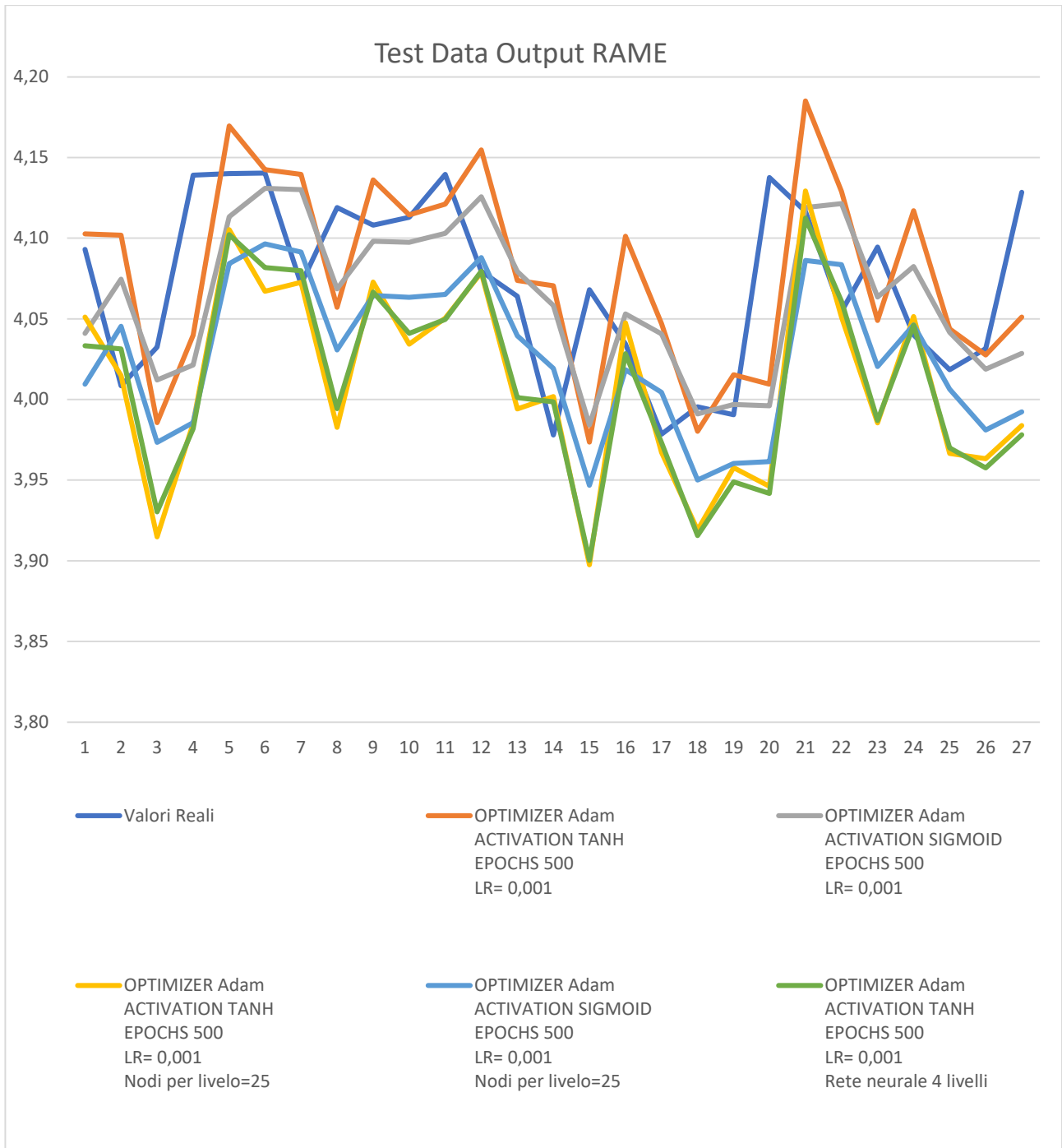


Figura 29. Plot forecasting commodity Rame

<b>Epoch=500 Optimizer Adam Learning rate=0,01</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.46418559467728615	0.4024500466254024	9.85272424225946
<b>Activation = Sigmoid</b>	1.4326261881036144	1.431628597751812	35.15609881710307
<b>Activation= Tanh</b>	0.4918835384318649	0.48535396743483034	11.90889338367082

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.11708076745478549	0.10012984850892313	2.4498287143307844
<b>Activation = Sigmoid</b>	0.057025994772731424	0.04447146523992239	1.090403615292599
<b>Activation= Tanh</b>	0.06259641791094586	0.052071307506826196	1.2796054762150928

<b>Epoch=500, Optimizer=Adam Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.25384692399342357	0.24785754524668052	6.076446017516314
<b>Activation = Sigmoid</b>	0.03434451615481054	0.027347226937611897	3.014065451099108
<b>Activation= Tanh</b>	0.1745083107248558	0.1660121874290484	4.064320346629109

<b>Epoch=500, Optimizer=SGD Learning rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.9083164659366907	0.9065343766190389	22.25439211700028
<b>Activation = Sigmoid</b>	1.216004054011936	1.2148201072436795	29.829318015300753
<b>Activation= Tanh</b>	0.8931320737420445	0.8913391137641891	21.881129533904613

<b>Epoch=500, Optimizer=SGD Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.5689558703049339	0.568473427383988	63.3163715981932
<b>Activation = Sigmoid</b>	1.2927402951427176	1.2916160733158943	31.716026550153924
<b>Activation= Tanh</b>	1.2927402951427176	1.2916160733158943	31.716026550153924

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.1468278434172773	0.13682254154262719	3.345517856042207
<b>Activation = Sigmoid</b>	0.07173635959702168	0.056892056792974544	1.390189218701813
<b>Activation= Tanh</b>	0.08579488223052707	0.06519151306483478	1.5934993584825146

Epoch=500, Optimizer=Adam Learning Rate=0,001 Architettura 4 livelli	RMSE	MAE	MAPE
Activation=Sigmoid	0.10314392635679745	0.08621134168903039	2.1067733725133695
Activation= Tanh	0.08514737865299084	0.0648618646021243	1.5855626324020595

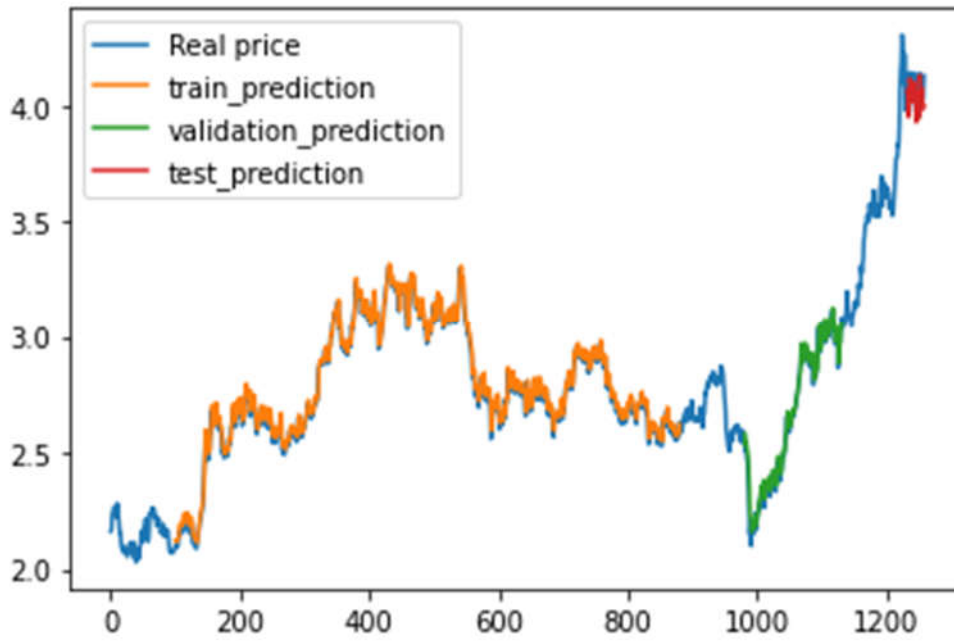


Figura 30. Plot Funzione di attivazione TANH, optimizer Adam, Learning Rate= 0,01, 500 Epoche - COMMODITY RATE

### 6.3. COMMODITY COTONE

L'analisi delle quotazioni del Cotone ha evidenziato che l'*optimizer* Adam è in grado di generare sistematicamente valori di errori di previsione più bassi rispetto al gradiente discendente stocastico (SGD).

A differenza delle altre commodity oggetto dell'analisi, applicando l'algoritmo di ottimizzazione Adam, la miglior performance in termini di errore medio percentuale è stata ottenuta impiegando la funzione di attivazione ReLU, con un valore MAPE pari a 1,58%

Dopo aver effettuato delle prove con learning rate pari a 0,001 e 0,0001, dati gli alti valori di MSE e MAE, si è deciso di eliminare SGD e utilizzare l'algoritmo di ottimizzazione Adam per reti neurali con una diversa struttura.

Per verificare la robustezza del modello, in prima battuta sono stati dimezzati il numero di nodi per ciascun layer. In questo caso tuttavia la funzione di attivazione tangente iperbolica ha generato una previsione delle quotazioni più accurata (MAPE 1,57%).

Infine, il test è stato ripetuto eliminando uno degli *hidden layer*, confermando la robustezza della rete neurale (MAPE 1,68%).

La tabella n. 19 riporta la previsione dei prezzi della commodity negli ultimi 27 giorni del range temporale analizzato, mentre la figura n. 32 mostra graficamente le oscillazioni rispetto ai valori reali dei valori output generati in output dal modello.

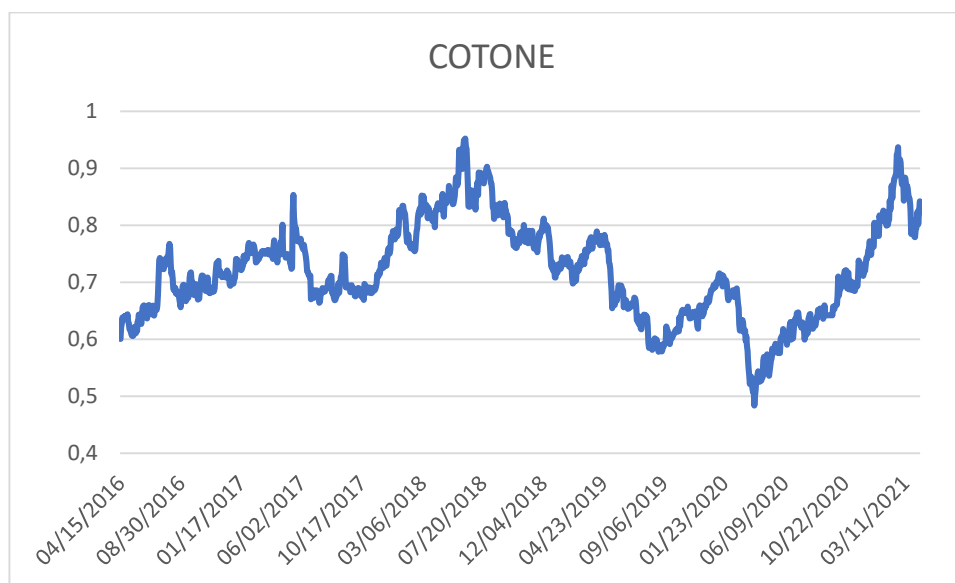


Figura 31. Prezzo Cotone in \$cent/lb - 04/2016-04/2021

Tabella 19. Forecasting commodity cotone

Indice	Valori Reali	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001	OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001	OPTIMIZER Adam ACTIVATION ReLU EPOCHS 500 LR= 0,001	OPTIMIZER Adam ACTIVATION RELU EPOCHS 500 LR= 0,001 Nodi per livello=25	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Nodi per livello=25	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Rete neurale 4 livelli
1	0,88	0,88	0,88	0,88	0,87	0,88	0,88
2	0,84	0,89	0,88	0,84	0,87	0,88	0,88
3	0,85	0,84	0,84	0,85	0,83	0,83	0,83
4	0,88	0,86	0,85	0,88	0,85	0,85	0,85
5	0,88	0,89	0,88	0,88	0,88	0,88	0,89
6	0,87	0,88	0,87	0,87	0,86	0,87	0,87
7	0,87	0,87	0,87	0,87	0,85	0,86	0,86
8	0,87	0,87	0,87	0,87	0,86	0,87	0,87
9	0,85	0,87	0,86	0,85	0,85	0,86	0,86
10	0,85	0,86	0,85	0,85	0,85	0,85	0,85
11	0,85	0,86	0,85	0,85	0,84	0,85	0,85
12	0,84	0,86	0,85	0,84	0,84	0,85	0,85
13	0,82	0,85	0,84	0,82	0,83	0,84	0,84
14	0,78	0,84	0,83	0,78	0,82	0,83	0,83
15	0,80	0,80	0,79	0,80	0,78	0,79	0,79
16	0,80	0,82	0,81	0,80	0,81	0,81	0,81
17	0,81	0,82	0,81	0,81	0,81	0,81	0,81
18	0,81	0,82	0,81	0,81	0,80	0,81	0,81
19	0,78	0,82	0,82	0,78	0,81	0,81	0,81
20	0,78	0,79	0,78	0,78	0,78	0,78	0,79
21	0,79	0,79	0,79	0,79	0,78	0,78	0,78
22	0,80	0,80	0,80	0,80	0,79	0,80	0,79
23	0,81	0,81	0,80	0,81	0,79	0,80	0,80
24	0,82	0,82	0,82	0,82	0,81	0,82	0,81
25	0,80	0,83	0,83	0,80	0,82	0,83	0,82
26	0,82	0,81	0,80	0,82	0,80	0,80	0,80
27	0,84	0,83	0,82	0,84	0,82	0,82	0,82

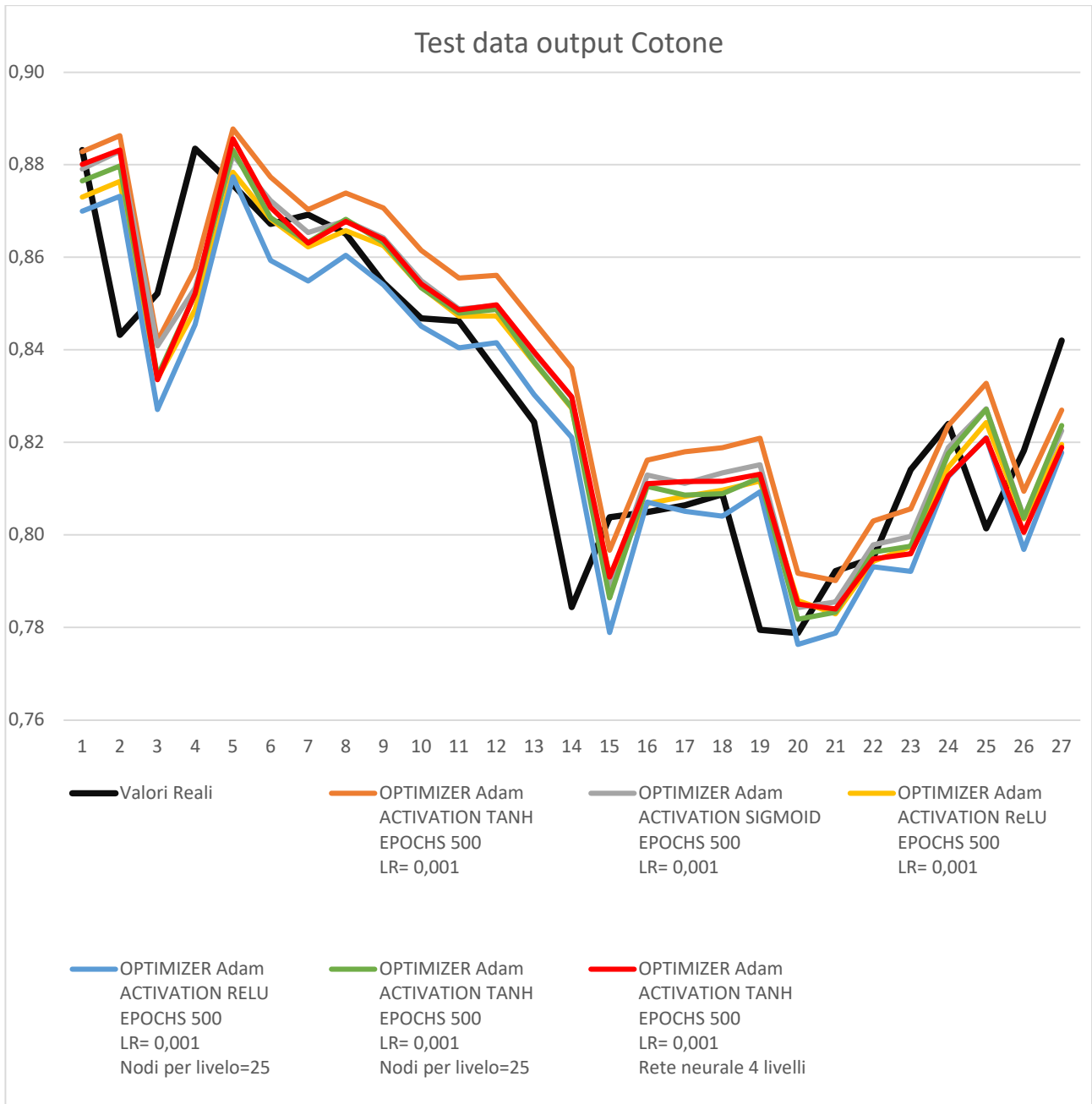


Figura 32. Plot Forecasting commodity cotone

Epoch=500 Optimizer Adam Learning rate=0,01	RMSE	MAE	MAPE
Activation = ReLU	0.021490248963706247	0.01663030870755514	2.0089833612214596
Activation = Sigmoid	0.01938472370380853	0.015509110530217484	1.8719815889907534
Activation= Tanh	0.04556959101783846	0.03607421671902692	4.947039694777104

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.017241698514754968	0.012743828058242788	1.5467078159015522
<b>Activation = Sigmoid</b>	0.017781517926444387	0.013346993843714402	1.624104710772994
<b>Activation= Tanh</b>	0.020020417152436444	0.015354973236719782	1.8697841118339098

<b>Epoch=500, Optimizer=Adam Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.019066643931175523	0.01467096161842348	1.7817536085685024
<b>Activation = Sigmoid</b>	0.04790053723283543	0.03899189277931496	5.354896410043125
<b>Activation= Tanh</b>	0.019585735964034106	0.01502731267611187	1.8244044360288578

<b>Epoch=500, Optimizer=SGD Learning rate=0,01</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.0374385632439764	0.030291454712549844	3.7347873101736746
<b>Activation = Sigmoid</b>	0.035444536856157616	0.028483264287312828	3.51333332732933
<b>Activation= Tanh</b>	0.035578135099968336	0.028540409723917647	3.520994592083086

<b>Epoch=500, Optimizer=SGD Learning rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.03947600564366197	0.03254780165354411	3.8374516725439847
<b>Activation = Sigmoid</b>	0.09183911996052097	0.0862542095184326	10.266218634558227
<b>Activation= Tanh</b>	0.02930568565932354	0.02414359156290691	2.9431489909937545

<b>Epoch=500, Optimizer=SGD Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.08692774832514669	0.08105460397402443	9.639161441583061
<b>Activation = Sigmoid</b>	0.09179834460824221	0.08616405916213987	10.25474513985946
<b>Activation= Tanh</b>	0.08038235555465614	0.07425847816467285	8.821818633631747

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	0.017919424209455928	0.013745185295740762	1.662814617397386
<b>Activation = Sigmoid</b>	0.09045821910212068	0.08476446708043417	10.086162744473532
<b>Activation= Tanh</b>	0.017383981994183995	0.012964671532313045	1.574306373964431



Epoch=500, Optimizer=Adam Learning Rate=0,001 Architettura 4 livelli	RMSE	MAE	MAPE
Activation=Tanh	0.01815805496677665	0.01387304449081421	1.684007979815358

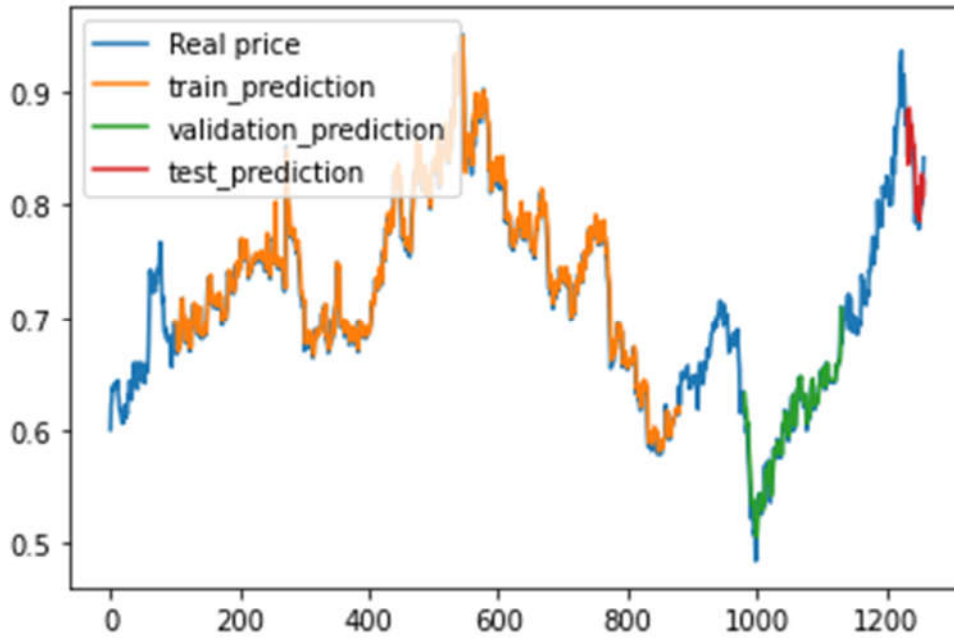


Figura 33. Plot Funzione di attivazione TANH, optimizer Adam, Learning Rate= 0,001, 500 Epoche - COMMODITY COTONE

## 6.4. COMMODITY SOIA

L'analisi effettuata sul prezzo del future della Soia evidenzia che la funzione di attivazione tangente iperbolica è in grado di fornire la miglior previsione sui dati per il test della rete (MAPE 1.22 %).

L'algoritmo di ottimizzazione Adam ha generato sistematicamente dei risultati più accurati. Infatti, per valori di learning rate compresi tra  $10^{-3}$  e  $10^{-4}$ , si registra un errore percentuale inferiore al 10% per tutte le tipologie di funzione di attivazione.

L'algoritmo gradiente discendente stocastico ha rivelato previsioni lontane dai valori reali; per questo motivo è stato deciso di proseguire i test modificando la struttura della rete solamente con l'impiego dell'ottimizzatore Adam. Il test è stato ripetuto dimezzando il numero di nodi per ciascun livello con valori di learning rate  $10^{-3}$  e  $10^{-4}$  per verificare la robustezza del modello. I risultati hanno confermato l'efficienza della funzione di attivazione sigmoide (MAPE 2,07%).

Infine, prendendo come riferimento la rete neurale che ha fornito il risultato più preciso, è stato eliminato un hidden layer di tipo Dense dalla struttura, con bias percentuale rispetto ai dati reali approssimativamente del 3,64%.

La tabella n. 20 riporta la previsione dei prezzi della commodity negli ultimi 27 giorni del range temporale analizzato, mentre la figura n. 35 mostra graficamente le oscillazioni rispetto ai valori reali dei valori output generati in output dal modello.

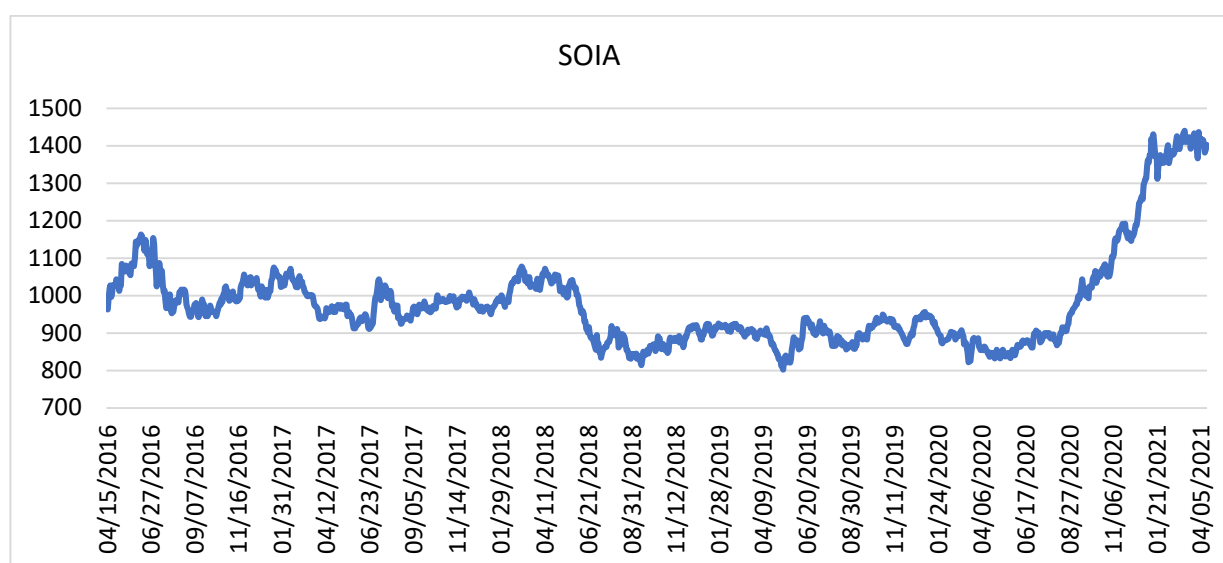


Figura 34. Prezzo soia in cent\$/Bushel - 04/2016-04/2021

Tabella 20. Forecasting commodity Soia

Indice	Valori Reali	OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,0001	OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,0001 Nodi per livello=25	OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,0001 Nodi per livello=25	OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001 Rete neurale 4 livelli
1	1433,75	1392,28	1437,88	1362,37	1402,62	1356,99
2	1440,00	1397,27	1433,39	1365,60	1401,51	1360,57
3	1409,75	1402,19	1439,27	1369,49	1408,39	1365,45
4	1413,50	1384,35	1402,10	1371,31	1371,21	1351,91
5	1413,25	1382,86	1422,27	1372,05	1388,06	1352,75
6	1419,50	1382,28	1422,29	1371,91	1381,74	1351,61
7	1423,25	1386,28	1429,18	1371,60	1393,11	1354,19
8	1417,75	1389,56	1430,47	1371,42	1393,18	1356,23
9	1392,25	1386,84	1421,72	1370,92	1387,26	1353,93
10	1416,25	1369,83	1394,21	1368,44	1359,83	1340,91
11	1417,50	1381,67	1434,82	1366,68	1396,07	1351,12
12	1423,25	1384,88	1425,26	1365,46	1385,58	1351,66
13	1432,75	1389,32	1429,61	1365,12	1397,21	1355,08
14	1414,25	1396,05	1438,45	1366,07	1402,25	1360,65
15	1400,50	1385,93	1412,12	1366,47	1379,95	1352,59
16	1393,00	1374,91	1405,21	1365,58	1371,46	1345,24
17	1366,75	1367,64	1404,09	1363,43	1365,47	1339,74
18	1436,75	1348,68	1376,52	1358,69	1338,75	1324,13
19	1402,00	1390,23	1470,59	1357,76	1427,27	1356,92
20	1412,75	1377,18	1395,46	1356,53	1359,90	1341,79
21	1418,75	1381,49	1420,88	1356,25	1395,93	1347,82
22	1408,75	1385,69	1425,59	1357,02	1385,85	1351,32
23	1415,25	1380,18	1410,21	1357,70	1380,41	1346,95
24	1403,00	1383,10	1423,17	1358,82	1388,24	1350,23
25	1382,00	1375,83	1406,59	1359,22	1372,09	1344,18
26	1389,50	1360,63	1387,48	1357,66	1354,16	1332,70
27	1402,25	1362,58	1406,46	1355,53	1368,46	1334,46

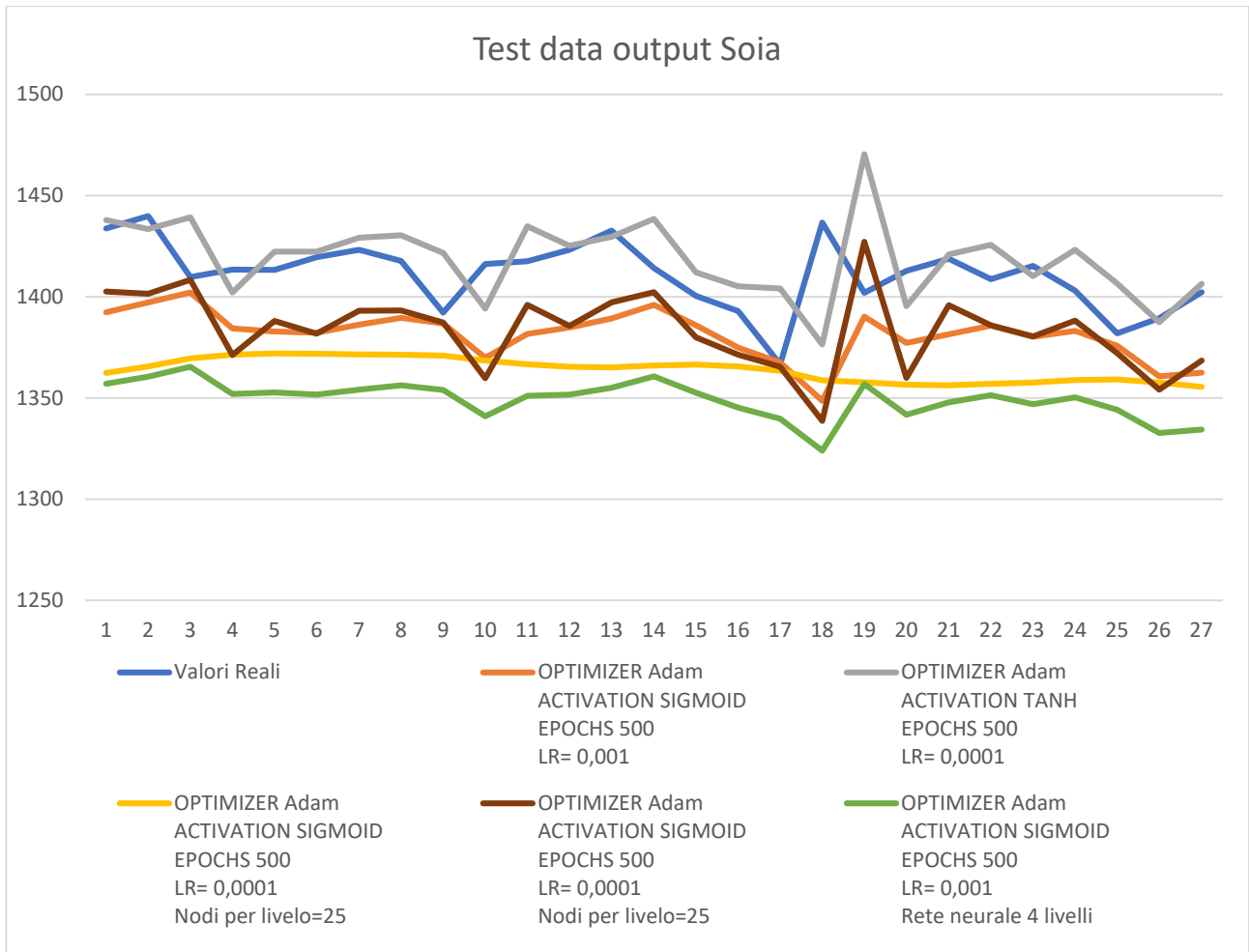


Figura 35. Plot forecasting commodity Soia

Epoch=500 Optimizer Adam Learning rate=0,01	RMSE	MAE	MAPE
Activation = ReLU	161.09826234124023	149.3871500922574	10.602764670278148
Activation = Sigmoid	221.11904344716996	220.548086365064	15.619278273144296
Activation= Tanh	208.49901749917032	207.87312901682324	14.720530717152302

Epoch=500, Optimizer=Adam Learning Rate=0,001	RMSE	MAE	MAPE
Activation = ReLU	99.09950534650613	97.53527036971516	6.90140705896899
Activation = Sigmoid	34.303219617291525	29.640012565586304	2.089897678505208
Activation= Tanh	122.6376042263163	121.33971088462405	8.590101274130538

<b>Epoch=500, Optimizer=Adam Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	87.14264492365693	84.92385827832751	6.010418243265398
<b>Activation = Sigmoid</b>	36.15289862356591	32.897508806652496	2.3204416964595187
<b>Activation= Tanh</b>	23.771957688694663	17.133432626724243	1.2175878157178759

<b>Epoch=500, Optimizer=SGD Learning rate=0,01</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	73.03197840079706	70.50283973415692	4.981701202073419
<b>Activation = Sigmoid</b>	430.956248288566	430.6358470271031	30.50923772218394
<b>Activation= Tanh</b>	90.88150602358624	88.85896934403314	6.282726544388074

<b>Epoch=500, Optimizer=SGD Learning rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	106.82089962115299	105.14503055148654	7.437210758131446
<b>Activation = Sigmoid</b>	407.8938704787726	407.548754538099	28.872715729526906
<b>Activation= Tanh</b>	348.09961188214817	347.659854301148	24.627310048928454

<b>Epoch=500, Optimizer=SGD Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	453.96732970876275	453.6654661926958	32.141608447518486
<b>Activation = Sigmoid</b>	463.0633336136418	462.76856946448487	32.78685579614527
<b>Activation= Tanh</b>	422.2831943697751	421.94886927803356	29.893364248425886

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	245.46774999830245	244.567758589983	17.328850685525378
<b>Activation = Sigmoid</b>	35.064063402325864	29.36459771460957	2.0720760400099802
<b>Activation= Tanh</b>	106.89804075932688	105.50789278414514	7.467167323399194

<b>Epoch=500, Optimizer=Adam Learning Rate=0,0001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	224.38335934173207	222.0055234928926	15.714852215980029
<b>Activation = Sigmoid</b>	49.83563094246343	47.00440242886543	3.3186081778300487
<b>Activation= Tanh</b>	58.24236838825758	55.7961975534757	3.9425565000816745

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Architettura 4 livelli</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = Sigmoid</b>	63.9588985759789	61.74567764004072	4.364646112821458
<b>Activation= Tanh</b>	99.75441700870141	98.04672574003537	6.938065576645328

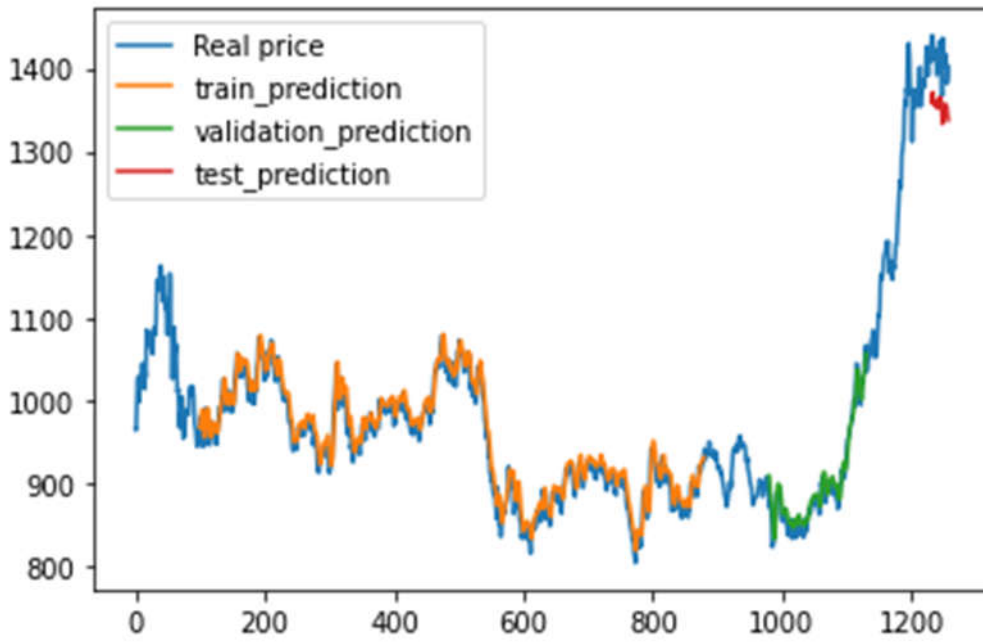


Figura 36. Plot Funzione di attivazione TANH, optimizer Adam, Learning Rate= 0,0001, 500 Epoche - COMMODITY SOLA

## 6.5. COMMODITY PETROLIO

L'analisi del prezzo del barile di petrolio greggio lungo un arco temporale di cinque anni ha confermato che la rete neurale LSTM è in grado di fornire generalmente una buona previsione dei prezzi futuri della commodity.

I risultati in termine di errore assoluto e percentuale sono dello stesso ordine di grandezza al variare della tipologia di funzione di attivazione (ReLU, Sigmoid, e Tanh), impiegando l'algoritmo di ottimizzazione Adam.

Il test è stato ripetuto modificando il valore del learning rate sulla stessa struttura neurale. Le migliori previsioni sono state generate utilizzando l'algoritmo Adam, un learning rate pari a  $10^{-3}$  e funzioni di attivazione sigmoide (MAPE 2,07%) e tangente iperbolica (MAPE 2,09%).

Dopo aver trovato questi risultati, sono stati ripetuti i test su 500 epoche di apprendimento dimezzando il numero di neuroni per ciascun livello, confermando la robustezza del modello con un MAPE pari al 2,10% in corrispondenza della funzione di attivazione tangente iperbolica.

Infine, è stato eliminato un hidden layer di tipo Dense dalla rete; l'errore percentuale si è mantenuto al di sotto del 3% (MAPE SIGMOID 2,90%- MAPE TANH 2,44%)

Successivamente, è stato eseguito il forecasting delle quotazioni con l'algoritmo di ottimizzazione SGD, ma i risultati emersi sono sistematicamente più distanti rispetto ai valori reali, seppur con un errore percentuale costantemente inferiore al 10%.

La tabella n. 21 riporta la previsione dei prezzi della commodity negli ultimi 27 giorni del range temporale analizzato, mentre la figura n. 38 mostra graficamente le oscillazioni rispetto ai valori reali dei valori output generati in output dal modello.

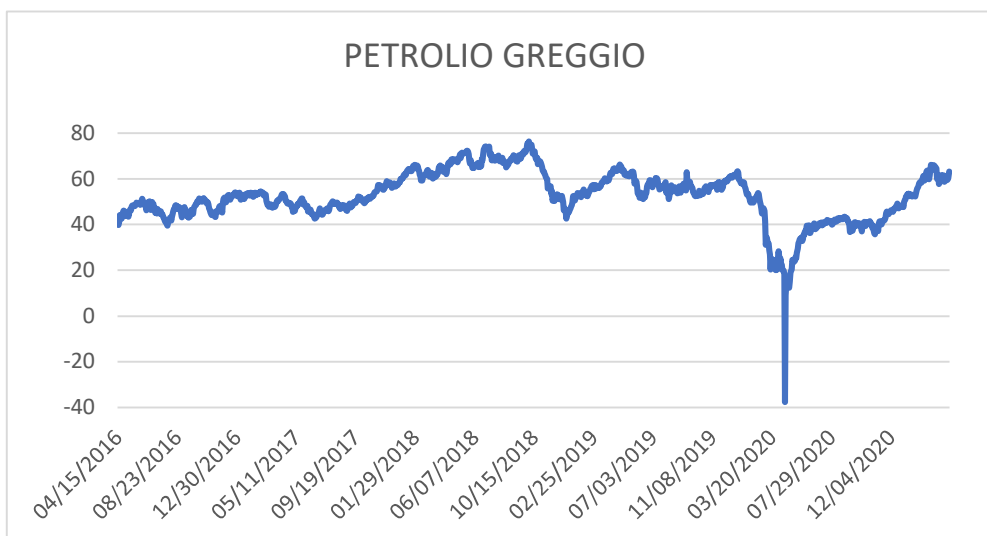


Figura 37. Prezzo Petrolio Greggio al Barile in \$ 04/2016-04/2021

Tabella 21. Forecasting commodity Petrolio Greggio

Indice	Valori Reali	OPTIMIZER Adam ACTIVATION SIGMOID EPOCHS 500 LR= 0,001	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,0001 Nodi per livello=25	OPTIMIZER Adam ACTIVATION TANH EPOCHS 500 LR= 0,001 Rete neurale 4 livelli
1	65,05	66,08	66,26	65,83	66,50
2	64,01	67,18	65,84	65,60	65,83
3	64,44	66,62	64,82	64,55	64,89
4	66,02	66,00	64,93	64,61	65,35
5	65,61	66,54	66,27	65,92	66,82
6	65,39	67,09	66,18	65,94	66,46
7	64,80	67,14	65,98	65,69	66,12
8	64,60	66,69	65,42	65,15	65,65
9	60,00	66,18	65,14	64,85	65,42
10	61,42	63,74	61,15	61,08	61,10
11	61,55	62,32	61,66	61,17	61,93
12	57,76	62,20	61,89	61,54	62,35
13	61,18	60,78	58,83	58,60	58,67
14	58,56	60,97	61,08	60,44	61,33
15	60,97	60,66	59,42	59,18	59,67
16	61,56	61,28	61,08	60,54	61,37
17	60,55	62,35	61,95	61,58	62,35
18	59,16	62,57	61,30	61,00	61,54
19	61,45	61,67	60,00	59,68	60,21
20	58,65	61,89	61,59	61,13	62,24
21	59,33	61,27	59,66	59,46	60,05
22	59,77	60,78	59,81	59,37	60,15
23	59,60	60,85	60,23	59,84	60,76
24	59,32	61,03	60,21	59,84	60,64
25	59,70	60,97	59,98	59,60	60,28
26	60,18	61,03	60,24	59,84	60,63
27	63,15	61,35	60,69	60,30	61,14



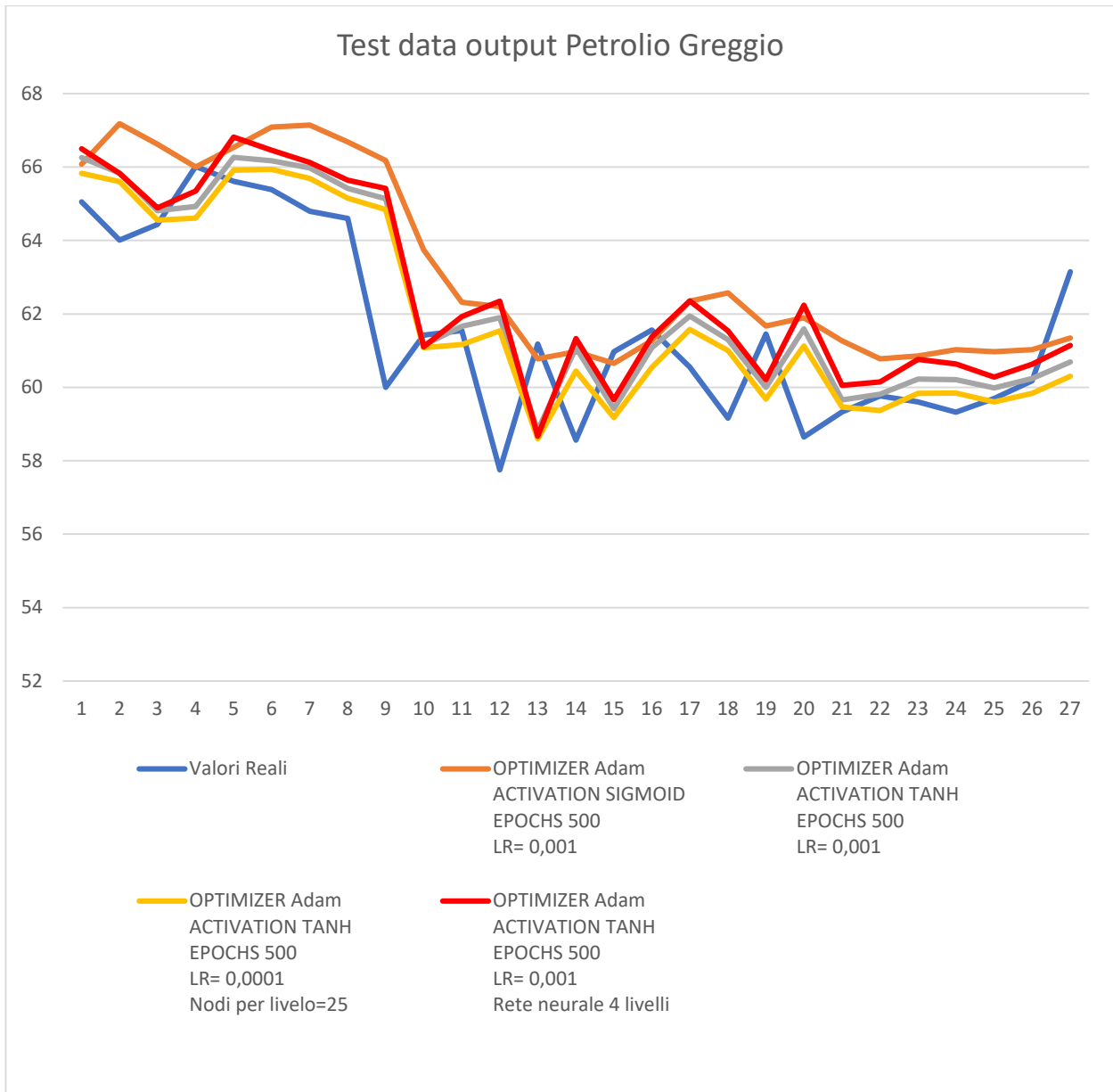


Figura 38. Plot Forecasting commodity Petrolio Greggio

Epoch=500 Optimizer Adam Learning rate=0,01	RMSE	MAE	MAPE
Activation = ReLU	1.8446211128551255	1.3760700858963857	2.260534397805152
Activation = Sigmoid	2.1384757844802134	1.7046056554935585	2.80338531972245
Activation= Tanh	1.9187083008455172	1.4793912680060768	2.425437902212887

Epoch=500, Optimizer=Adam Learning Rate=0,001	RMSE	MAE	MAPE
Activation = ReLU	1.9676729203400192	1.4838694389661151	2.4394314457147437
Activation = Sigmoid	1.6485773434293167	1.2729704458625224	2.0745943397535127
Activation= Tanh	1.7259738858338303	1.2737596895076615	2.093660290945621

<b>Epoch=500, Optimizer=Adam Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	2.54191329381083	2.148168640136719	3.4699155774492887
<b>Activation = Sigmoid</b>	2.5986128387301517	1.9854897903513031	3.2550601051859642
<b>Activation= Tanh</b>	2.316772637075536	1.7477305828200447	2.8320718670975764

<b>Epoch=500, Optimizer=SGD Learning rate=0,01</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	3.357931480441179	2.6154585293487265	4.163484861090547
<b>Activation = Sigmoid</b>	5.643671016401165	5.08447170451836	8.107387462395348
<b>Activation= Tanh</b>	3.1863147448116167	2.8833409136312973	4.659511516786664

<b>Epoch=500, Optimizer=SGD Learning rate=0,001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	4.411130777656748	3.6150677043420294	5.715286573206293
<b>Activation = Sigmoid</b>	5.584227948857805	5.011952423519563	7.988666313547766
<b>Activation= Tanh</b>	4.548371618399903	3.7725174058808224	5.969896186163233

<b>Epoch=500, Optimizer=SGD Learning Rate=0,0001</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	5.011158859651422	4.326242439481948	6.870017275715794
<b>Activation = Sigmoid</b>	5.49218463684958	4.909354777689339	7.8219280090114305
<b>Activation= Tanh</b>	4.408505378104469	3.6315829108379507	5.7438716289084955

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	2.1622459034345454	1.8225956922990303	2.953574118990705
<b>Activation = Sigmoid</b>	2.248078243882002	1.820687661170959	2.989049328720225
<b>Activation= Tanh</b>	1.738117667301435	1.2786243003386037	2.1000102363630035

<b>Epoch=500, Optimizer=Adam Learning Rate=0,0001 Nodi per livello=25</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation = ReLU</b>	2.6086527367797094	2.165768395794762	3.4774841803491676
<b>Activation = Sigmoid</b>	3.099802183453397	2.7992284365053526	4.5917225875711996
<b>Activation= Tanh</b>	2.5359369276411887	1.9408543488714431	3.1662307812542596

<b>Epoch=500, Optimizer=Adam Learning Rate=0,001 Architettura 4 livelli</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE</b>
<b>Activation= Sigmoid</b>	2.124850412531256	1.794947492988022	2.906091048672714
<b>Activation= Tanh</b>	1.946084382008672	1.485376176834105	2.440134625201315

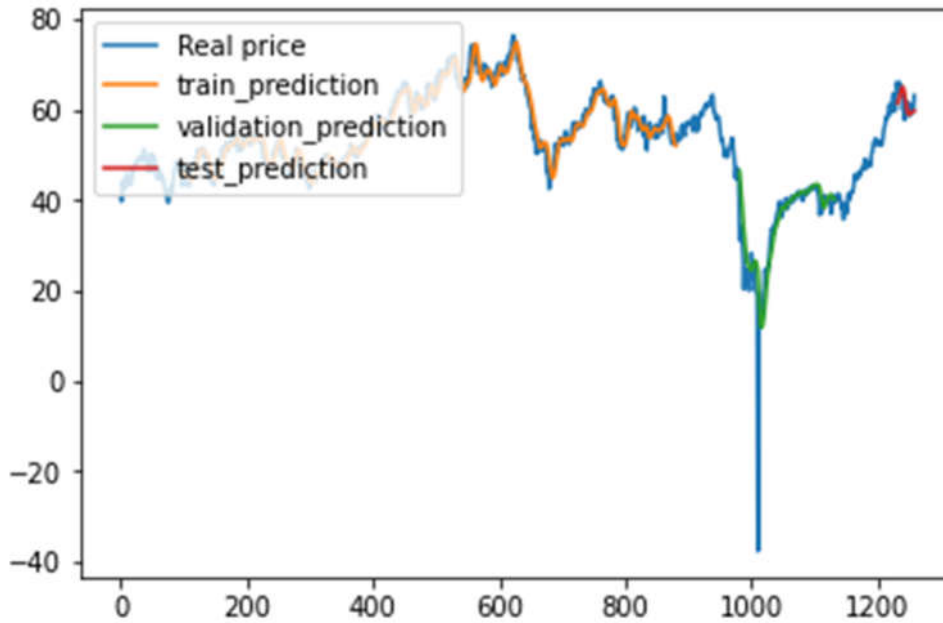


Figura 39. Plot Funzione di attivazione TANH, optimizer Adam, Learning Rate= 0,0001, 500 Epoche - COMMODITY PETROLIO GREGGIO

# CONCLUSIONE

Il presente elaborato fornisce una panoramica completa sul mercato delle materie prime. L'analisi delle quotazioni lungo il quinquennio 2016-2021, ha permesso di testare il comportamento efficiente del mercato. Il test effettuato per verificare la randomicità dei log-rendimenti ha rivelato che mediamente i movimenti dei prezzi si distribuiscono casualmente. Questo risultato conferma la solidità delle fluttuazioni di mercato, nonostante la crisi economica derivata dal Covid-19 abbia avuto un forte impatto sui prezzi del 2020 (ad esempio il crollo del prezzo del petrolio ad Aprile 2020 analizzato nel capitolo 2).

Nella seconda parte dell'elaborato, dopo aver studiato il comportamento matematico delle reti neurali e il comportamento delle reti LSTM, si è creato un modello scritto nel linguaggio di programmazione Python per prevedere i movimenti dei prezzi su un orizzonte temporale oscillante tra 25 e 30 giorni. Per raggiungere questo obiettivo è stato deciso di allenare la rete neurale per 500 epoche, partendo da una struttura neurale di base e modificandola iterativamente per verificare la robustezza degli output. Al fine di ottenere dei risultati più accurati, è stato fondamentale testare la rete applicando diversi algoritmi di ottimizzazione e molteplici funzioni di attivazione dei nodi dell'architettura.

I risultati ottenuti sono stati analizzati principalmente sulla distanza in termini percentuali rispetto ai valori di mercato reali. L'indice MAPE è stato il driver principale nella valutazione delle previsioni.

Generalmente, è stato dimostrato che per l'analisi delle materie prime oggetto della tesi l'algoritmo di ottimizzazione Adam risulta sistematicamente più efficiente del gradiente stocastico discendente, studiato nel capitolo 4.

Un'importante peculiarità emersa dai test riguarda il comportamento delle funzioni di attivazione sigmoide e tangente iperbolica ha fornito nella maggior parte dei casi i risultati più accurati rispetto alla Rectified Linear Unit (ReLU). Tuttavia, nessuna delle due funzioni è sistematicamente più efficiente dell'altra per tutte le serie temporali analizzate, nonostante i tempi di elaborazione del modello siano stati costantemente inferiori con l'utilizzo della tangente iperbolica.

Numericamente, le prestazioni predittive del modello hanno rivelato una capacità di commettere degli errori percentuali medi inferiori al 2,5% nel caso peggiore (MAPE Petrolio Greggio=2,07%) e inferiori all'1% nel best case (MAPE ORO=0,84%).

Lo studio delle reti neurali e del Deep learning per il forecasting delle quotazioni delle *commodities* trova un enorme campo di applicazione e molteplici gradi di libertà nella costruzione dell'architettura

e nella scelta dei parametri per l'allenamento della rete neurale.

Per questa ragione potrebbe essere interessante selezionare architetture neurali diverse, modificando i parametri di apprendimento come ad esempio il Learning Rate dell'algoritmo di ottimizzazione oppure le funzioni di attivazione dei nodi.

# **BIBLIOGRAFIA**

- 1. H. Kent Baker, Greg Filbeck and Jeffrey H. Harris, Commodities: Markets, Performance and Strategies, 2018**
- 2. John Baffes, Global Agricultural Trade and Developing Countries, World Bank Policy Research Working Paper 3218, February 2004.**
- 3. WCB-Commodities Markets Outlook- April 2021**
- 4. Vivek Voora, Cristina Larrea and Steffany Bermudez, Global Market Report, International Institute for Sustainable Development (IISD) (2020)**
- 5. John Y. Campbell, Andrew W. Lo, A. Craig MacKinlay, The Econometrics of Financial Markets, Princeton University, 1996**
- 6. Michael Nielsen, Neural Networks and Deep Learning,2013**

# SITOGRAFIA

1. [www.borsaitaliana.it](http://www.borsaitaliana.it)
2. [www.Justeff.com](http://www.Justeff.com)
3. [www.cotton.org](http://www.cotton.org)
4. <https://www.bancaditalia.it/compiti/riserve-portafoglio-rischi/caratteristiche-oro/index.html>
5. <https://www.treccani.it/enciclopedia/oro/>
6. <https://it.peopleperproject.com/posts/27782-a-look-at-the-largest-copper-smelters>
7. [www.Investopedia.com](http://www.Investopedia.com)
8. <https://www.istor.org/stable/resrep26554> Accessed: 02-06-2021 14:18 UTC
9. <https://www.treccani.it/enciclopedia>
10. <https://www.startingfinance.com/approfondimenti/petrolio-prezzo/>
11. <https://www.strategia-borsa.it>
12. <https://www.opec.org>
13. <https://www.eia.gov/the-world-factbook/field/crude-oil-proved-reserves/country-comparison>
14. <https://www.eia.gov>
15. [www.patreon.com/NEDLEducation](http://www.patreon.com/NEDLEducation)
16. <https://www.nasdaq.com>
17. <https://pandas.pydata.org/>
18. <https://numpy.org/>
19. <https://keras.io/>
20. <https://www.tensorflow.org/>
21. <https://qstack.it/>
22. <https://machinelearningmastery.com/>
23. <https://ichi.pro/it/reti-lstm-una-spiegazione-dettagliata-157979281180665>
24. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>