



Università Politecnica delle Marche

Facoltà di Ingegneria

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica e
dell'Automazione

Analisi e Predizione del prezzo di oro e argento in base ad eventi geopolitici

Classificazione e Sviluppo di un Modello Predittivo

Docenti

Prof. Ursino Domenico
Dott. Buratti Christopher

Componenti del gruppo

Dott. Tempera Fabio
Dott. Marcianesi Luca
Dott. Vianello Gabriele

ANNO ACCADEMICO 2024-2025

Indice

1	Il Dataset su Metalli Preziosi e Rischio Geopolitico	6
1.1	La scelta del dataset	6
1.1.1	Struttura e Caratteristiche del Dataset	7
1.1.2	La Struttura del Rischio: da Minacce a Eventi Concreti	8
2	Analisi Descrittiva	9
2.1	Introduzione	9
2.2	Librerie utilizzate	9
2.3	Risultati dell'analisi	10
2.3.1	Evoluzione Temporale e Composizione dell'Indice di Rischio Geopolitico	10
2.3.2	Evoluzione Storica dei Prezzi dei Metalli Preziosi . .	11
2.3.3	Confronto tra Prezzo dell'Oro e Rischio Geopolitico: L'Ipotesi Safe-Haven	12
2.3.4	Analisi Statistica dei Rendimenti Logaritmici	13
2.3.5	Analisi delle Correlazioni Lineari	15
3	Validazione dell'Ipotesi Safe Haven	18
3.1	Introduzione e Definizione dell'Ipotesi	18
3.1.1	Definizione dell'Ipotesi	18
3.1.2	Implicazioni Testabili	18
3.2	Dataset e Feature Engineering	19
3.2.1	Variabili Costruite	19
3.3	Test di Correlazione	19
3.3.1	Correlazione Contemporanea (Same-Day)	19
3.3.2	Correlazione con Lag Predittivo	20
3.4	Analisi Direzionale	21
3.4.1	Tabelle di Contingenza	21
3.4.2	Test Chi-Quadrato per Indipendenza	22
3.5	Test di Causalità di Granger	22
3.5.1	Risultati	22
3.6	Regressione Lineare	22

3.6.1	Risultati della Stima	23
3.7	Analisi per Quantili Estremi	23
3.7.1	Risultati	23
3.8	Sintesi e Verdetto Statistico	24
3.8.1	Riepilogo delle Evidenze	24
3.8.2	Verdetto Finale	24
3.9	Discussione e Interpretazione	24
3.9.1	Possibili Spiegazioni del Rifiuto	24
3.9.2	Implicazioni per la Modellazione	25
3.10	Conclusioni del Capitolo	25
4	Forecasting: Modelli Predittivi per i Prezzi dell'Oro	27
4.1	Introduzione e Obiettivi	27
4.1.1	Obiettivo della Previsione	27
4.1.2	Approccio Metodologico	27
4.2	Metriche di Valutazione	28
4.2.1	Ricostruzione dei Prezzi	28
4.2.2	Root Mean Squared Error (RMSE)	28
4.2.3	Mean Absolute Error (MAE)	28
4.2.4	Benchmark: Naive Random Walk	29
4.3	Strategia di Validazione	29
4.3.1	Walk-Forward Validation	29
4.3.2	Gestione della Frequenza Temporale	30
4.4	Modelli Statistici Lineari	31
4.4.1	ARIMA Baseline	31
4.4.2	SARIMAX con Variabili Esogene	32
4.4.3	ARIMA-GARCH Hybrid	33
4.5	Modelli di Deep Learning (LSTM)	35
4.5.1	LSTM Univariato	35
4.5.2	LSTM Multivariato (Technical Indicators Only)	36
4.5.3	LSTM Multivariato (con GPRD)	37
4.6	Confronto Comparativo Finale	38
4.6.1	Ranking dei Modelli	38
4.6.2	Visualizzazione Comparativa	38
4.6.3	Analisi per Famiglie di Modelli	39
4.6.4	Test Statistico: Equivalenza ARIMA-LSTM	40
4.7	Discussione	40
4.7.1	Perché ARIMA Batte LSTM?	40
4.7.2	Perché GPRD Non Aiuta?	41
5	Conclusioni	42
5.1	Risultati Principali	42

5.1.1	Rifiuto dell'Ipotesi Safe Haven	42
5.1.2	Vittoria della Parsimonia: ARIMA e Deep Learning .	42
5.1.3	Fallimento Predittivo del GPRD	43
5.2	Implicazioni e Raccomandazioni	43
5.2.1	Per Practitioner	43
5.2.2	Per Ricerca Futura	43
5.3	Limiti e Contributi	44
5.3.1	Limiti dello Studio	44
5.3.2	Contributi Originali	44
5.4	Riflessione Finale	44

Elenco delle figure

2.1	Andamento storico dell'Indice di Rischio Geopolitico (GPR) dal 1985 al 2025. Il grafico evidenzia l'indice complessivo (rosso) e le sue sottocomponenti: Atti (giallo) e Minacce (marrone). Si notano picchi estremi in corrispondenza dei maggiori conflitti e attacchi terroristici degli ultimi quarant'anni.	11
2.2	Andamento del prezzo spot di Oro (in alto) e Argento (in basso) dal 1985 al 2025. Entrambi i metalli mostrano un trend di crescita secolare a partire dal 2005, con l'Oro che evidenzia una minore volatilità e una crescita più sostenuta sui massimi storici recenti rispetto all'Argento.	12
2.3	Confronto diretto tra il prezzo dell'Oro (linea gialla, asse sx) e l'Indice di Rischio Geopolitico (linea rossa, asse dx). Il grafico permette di valutare l'ipotesi di <i>Safe-Haven</i> , evidenziando come i periodi di elevata tensione geopolitica o incertezza strutturale siano spesso accompagnati da trend positivi per il metallo prezioso.	13
2.4	Distribuzione dei rendimenti logaritmici giornalieri. Le linee tratteggiate rosse indicano la media campionaria. Si nota la tipica forma leptocurtica dei dati finanziari, con una concentrazione di dati attorno alla media e code estese che indicano il rischio di eventi estremi.	14
2.5	Serie storiche dei rendimenti logaritmici giornalieri (Daily Log Returns). Il confronto evidenzia i cluster di volatilità durante le crisi (es. 2008, 2020) e la maggiore rischiosità dell'Argento (scala dell'asse y più ampia) rispetto all'Oro.	15
2.6	Matrice di correlazione tra i Prezzi dei metalli e le componenti del Rischio Geopolitico. Si nota come la componente delle Minacce (Threat) sia maggiormente correlata al prezzo dell'Oro rispetto agli Atti effettivi (Act).	16

2.7	Matrice di correlazione sui Rendimenti Logaritmici. Mentre Oro e Argento mantengono una forte sincronia (0.70), l'assenza di correlazione lineare con il livello del GPRD suggerisce che la relazione tra rendimenti giornalieri e rischio sia non lineare o legata a specifici shock.	17
3.1	Scatter plot tra variazioni giornaliere del GPRD (asse x) e rendimenti logaritmici dell'oro (asse y). La linea di regressione (rosso) è sostanzialmente orizzontale (slope ≈ 0.0000096), confermando l'assenza di correlazione.	20
4.1	Schema della Walk-Forward Validation. Il modello viene addestrato su una finestra espandente e predice i successivi 5 giorni senza mai osservare dati futuri. La finestra di training cresce ad ogni iterazione.	30
4.2	Confronto RMSE tra tutti i modelli testati. ARIMA Baseline (barra blu) si conferma vincitore, superando sia il Naive Benchmark che architetture LSTM complesse. I modelli con variabili esogene (SARIMAX, LSTM Tech) mostrano performance inferiori.	39

1 Il Dataset su Metalli Preziosi e Rischio Geopolitico

Nel panorama dei mercati finanziari globali, caratterizzato da una crescente volatilità e interconnessione, la capacità di comprendere e anticipare le reazioni degli asset alle tensioni internazionali è diventata una priorità per investitori, analisti e policymaker. I mercati sono costantemente influenzati da un flusso di notizie ed eventi provenienti da ogni angolo del mondo, che rendono difficile distinguere i segnali significativi dal rumore di fondo e comprendere le dinamiche di causa-effetto.

Per superare questa sfida, la comunità scientifica e finanziaria si sta orientando verso approcci quantitativi basati sui dati. Storicamente, un ostacolo è stata la difficoltà di combinare in modo sistematico dati finanziari ad alta frequenza con misure oggettive del rischio geopolitico su un orizzonte temporale esteso. Per colmare questa lacuna, questo progetto utilizza un dataset composito che unisce i dati giornalieri dei mercati di Oro e Argento con l'indice di rischio geopolitico (GPRD), fornendo una prospettiva a lungo termine su come gli eventi globali influenzano i prezzi dei metalli preziosi.

1.1 La scelta del dataset

Nel presente progetto ci siamo avvalsi di un unico ma estremamente ricco dataset principale, scelto con l'obiettivo di rispondere a specifiche esigenze analitiche e metodologiche. La decisione di utilizzare questa combinazione di dati non è casuale, ma motivata da diverse caratteristiche che la rendono ideale per l'applicazione di tecniche di analisi delle serie storiche, correlazione e forecasting nel dominio della finanza quantitativa.

A differenza di analisi basate su periodi brevi o su dati qualitativi, questo dataset offre una visione senza precedenti su:

- **Prospettiva Storica e Realismo:** con dati che coprono quasi quattro decenni, il dataset cattura una vasta gamma di contesti economici

e crisi internazionali, garantendo che i modelli sviluppati siano addestrati su dati di mercato reali e su un indice di rischio ampiamente riconosciuto.

- **Fonti Riconosciute:** i dati sui metalli preziosi provengono da una fonte di mercato affidabile come Investing.com, mentre l'indice GPRD è un riferimento accademico sviluppato da Matteo Iacoviello, garantendo la qualità e la validità delle informazioni.
- **Ricchezza Contestuale:** l'integrazione di dati finanziari (prezzi, volumi) con un indice quantitativo del rischio e dei suoi sotto-indici (minacce percepite ed eventi reali) permette di andare oltre la semplice analisi dei prezzi e di comprendere i *fattori scatenanti* delle fluttuazioni di mercato.

Questi elementi rappresentano uno scenario ideale per esplorare la dinamica relazione tra i rischi geopolitici e i mercati delle materie prime, che verranno analizzati nei capitoli successivi.

1.1.1 Struttura e Caratteristiche del Dataset

Il dataset combina dati storici giornalieri, unendo i prezzi dei metalli preziosi con gli indicatori di rischio geopolitico. Si tratta di un dataset estremamente ricco, che permette di analizzare l'interazione tra eventi globali e mercati delle materie prime. Tra le colonne principali che abbiamo utilizzato troviamo:

- **DATE:** la data di negoziazione, con frequenza giornaliera. Questa feature è fondamentale per le analisi di serie storiche e per i modelli di forecasting.
- **GOLD_PRICE, GOLD_OPEN, GOLD_HIGH, GOLD_LOW, GOLD_CHANGE_%:** il prezzo spot giornaliero dell'Oro, insieme ai valori di apertura, massimo, minimo e alla variazione percentuale.
- **SILVER_PRICE, SILVER_OPEN, SILVER_HIGH, SILVER_LOW, SILVER_CHANGE_%:** il prezzo spot giornaliero dell'Argento, insieme ai valori di apertura, massimo, minimo e alla variazione percentuale.
- **GPRD:** l'indice di rischio geopolitico giornaliero (Geopolitical Risk Index), che misura il livello complessivo delle tensioni geopolitiche globali. Questa è una delle nostre variabili esplicative chiave.
- **GPRD_ACT:** un sotto-indice che cattura gli atti geopolitici realizzati (es. inizio di una guerra, attacchi terroristici).

- **GPRD_THREAT**: un sotto-indice che cattura le minacce geopolitiche percepite (es. retorica ostile, escalation di tensioni diplomatiche).
- **EVENT**: etichette testuali che descrivono i principali eventi mondiali associati a picchi o variazioni significative dell'indice GPR.

1.1.2 La Struttura del Rischio: da Minacce a Eventi Concreti

Una delle caratteristiche più potenti di questo dataset è la sua struttura multi-livello per quanto riguarda il rischio, che riflette come le tensioni globali si sviluppino e vengono percepite dai mercati. Questa struttura, intrinseca all'indice GPRD, permette di scomporre il concetto generico di "rischio" in componenti più specifiche e misurabili:

- **Minacce (Threats)**: al livello più basso e spesso più frequente, le minacce rappresentano un aumento percepito del rischio non ancora concretizzatosi in un evento. Sono i "segnali deboli" dell'analisi. Esempi includono l'intensificarsi della retorica tra due nazioni, la pianificazione di operazioni militari o l'aumento di discorsi ostili sulla stampa. Questo livello è catturato dal sotto-indice GPRD_THREAT.
- **Atti (Acts)**: al livello superiore, gli atti rappresentano eventi geopolitici avversi che si sono effettivamente verificati. Questi sono eventi concreti e inequivocabili che hanno un impatto diretto sul panorama internazionale. Esempi includono l'inizio di una guerra, un attacco terroristico significativo o l'imposizione di sanzioni. Questo livello è catturato dal sotto-indice GPRD_ACT.
- **Rischio Geopolitico Complessivo (GPRD)**: al livello più alto, l'indice principale aggrega le informazioni provenienti da minacce e atti per fornire una misura sintetica e complessiva del rischio geopolitico globale. È a questo livello aggregato che spesso si cerca una correlazione con l'andamento dei mercati finanziari.

Questa struttura gerarchica è fondamentale perché consente di investigare se i mercati reagiscano in modo diverso alle semplici minacce rispetto agli eventi già accaduti, permettendo ai nostri modelli di cogliere sfumature più profonde nella relazione tra rischio e rendimento degli asset.

2 Analisi Descrittiva

2.1 Introduzione

Prima di procedere con la costruzione di modelli predittivi, è essenziale condurre una fase di analisi descrittiva approfondita, nota anche come Exploratory Data Analysis (EDA). Questa sezione si focalizza sull'esplorazione del dataset per ottenere i primi, cruciali insight.

L'analisi delle serie storiche finanziarie in relazione a indicatori geopolitici, uno dei campi più dinamici e complessi, rappresenta una ricca fonte di informazioni che possono essere analizzate per approfondire la reazione dei mercati a eventi globali, la performance degli asset "bene rifugio" e l'evoluzione del rischio nel tempo. L'obiettivo di questa analisi descrittiva è esplorare le peculiarità dei dati, evidenziare trend, pattern, anomalie e distribuzioni rilevanti, e fornire una base solida e informata per tutte le fasi successive del progetto, dall'analisi di correlazione al forecasting.

2.2 Librerie utilizzate

L'analisi è stata interamente condotta utilizzando il linguaggio di programmazione Python e le seguenti librerie, che costituiscono lo standard de facto per la data science:

- **Pandas:** utilizzata per la manipolazione e l'analisi dei dati tabulari. Ha permesso di caricare il dataset, gestire le serie storiche, pulirlo e trasformarlo in un formato idoneo all'analisi.
- **Matplotlib e Seaborn:** usate per creare grafici e visualizzazioni efficaci. Matplotlib ha fornito gli strumenti di base per la plotting, mentre Seaborn ha aggiunto una componente estetica e statistica, particolarmente utile per heatmap, istogrammi e distribuzioni complesse.
- **NumPy:** essenziale per operazioni matematiche e manipolazioni di array numerici, come il calcolo di medie, rendimenti giornalieri e deviazioni standard su larga scala.

- **Scikit-learn:** impiegata per alcune operazioni di preprocessing preliminare e per l'analisi statistica, ponendo le basi per il suo utilizzo intensivo nei capitoli successivi.
- **Pytorch:** un framework open-source per il machine learning e il deep learning, utilizzato per l'addestramento dei modelli predittivi.
- **Statsmodels:** questa libreria è stata utilizzata nel contesto dei modelli ARIMA.

I dati sono stati pre-elaborati per gestire eventuali valori mancanti, allineare le serie temporali e garantire l'integrità e la coerenza delle analisi.

2.3 Risultati dell'analisi

Di seguito verranno riportate le analisi che sono state condotte insieme ai relativi risultati ottenuti.

2.3.1 Evoluzione Temporale e Composizione dell'Indice di Rischio Geopolitico

La variabile GPR è il cuore della nostra analisi del rischio. Essa rappresenta una misura quantitativa delle tensioni globali, basata sulla frequenza di articoli di stampa che discutono di rischi geopolitici. Analizzare la sua evoluzione storica è fondamentale per comprendere come gli eventi internazionali influenzino la percezione del rischio nel tempo.

Il grafico in Figura 2.1 mostra l'andamento dell'indice dal 1985 al 2025, scomponendo il rischio totale (*Overall GPR*) nelle sue due componenti principali: gli atti ostili realizzati (*GPR Acts*) e le minacce percepite (*GPR Threats*). Dall'analisi visiva emerge un comportamento caratterizzato da **cluster di volatilità e picchi estremi** in corrispondenza di eventi storici critici. Il picco più elevato, che supera il valore di 1600, si registra chiaramente intorno al 2001, in coincidenza con gli attacchi terroristici dell'11 settembre e le successive operazioni militari. Altri picchi significativi sono visibili all'inizio degli anni '90 (Guerra del Golfo) e, più recentemente, a partire dal 2022, riflettendo le tensioni legate al conflitto russo-ucraino e alla crisi in Medio Oriente.

È interessante notare come la componente *GPR Acts* (linea gialla) tenda a generare i picchi più acuti e immediati, mentre la componente *GPR Threats* (linea marrone) mostra una persistenza maggiore, mantenendo l'indice elevato anche in assenza di nuovi atti ostili diretti. Questa dinamica suggerisce che, mentre gli atti violenti causano shock improvvisi sui mercati

e sull'opinione pubblica, è la minaccia latente a sostenere livelli di rischio elevati nel medio periodo.

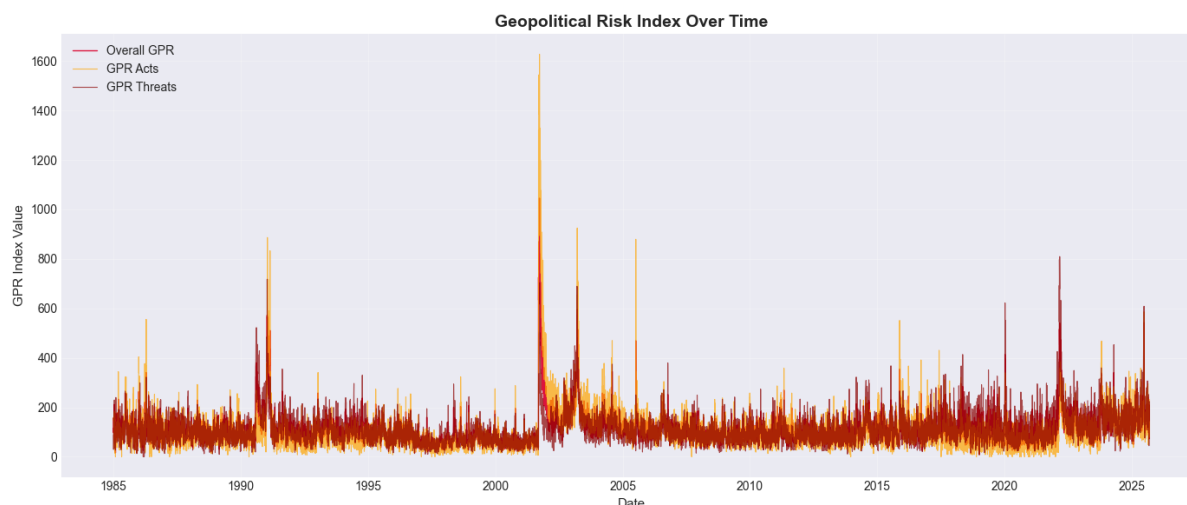


Figura 2.1: Andamento storico dell'Indice di Rischio Geopolitico (GPR) dal 1985 al 2025. Il grafico evidenzia l'indice complessivo (rosso) e le sue sottocomponenti: Atti (giallo) e Minacce (marrone). Si notano picchi estremi in corrispondenza dei maggiori conflitti e attacchi terroristici degli ultimi quarant'anni.

2.3.2 Evoluzione Storica dei Prezzi dei Metalli Preziosi

L'analisi dell'andamento storico dei prezzi di Oro e Argento (espressi in USD) è fondamentale per comprendere il comportamento di lungo periodo di questi asset e la loro reazione ai cicli economici globali. A differenza dei rendimenti giornalieri, che mostrano la volatilità a breve termine, i grafici dei prezzi evidenziano i trend strutturali e la capacità di questi metalli di fungere da riserva di valore.

La Figura 2.2 mostra l'evoluzione temporale dal 1985 al 2025. Per quanto riguarda l'**Oro** (grafico superiore), si nota una fase di relativa stabilità laterale fino al 2005, seguita da un deciso trend rialzista. Si distinguono chiaramente due fasi di forte accelerazione: la prima culminata nel 2011 (in risposta alla crisi finanziaria globale) e la seconda, ancora più marcata, iniziata nel 2019 e proseguita fino al 2025, dove il prezzo ha superato i massimi storici oltre quota 3.500 USD. Questo andamento conferma il ruolo dell'oro come *safe haven* (bene rifugio) nei periodi di incertezza economica e geopolitica.

L'**Argento** (grafico inferiore) mostra un'alta correlazione con l'oro, ma con una volatilità decisamente superiore. Anche qui è evidente il picco del 2011 (vicino ai 50 USD) e la successiva correzione. Tuttavia, rispetto all'oro, l'argento mostra oscillazioni più ampie e "nervose", riflettendo la sua doppia natura di metallo prezioso e metallo industriale. La recente ripresa

dal 2020 evidenzia un ritorno d'interesse degli investitori, pur mantenendo un profilo di rischio più elevato rispetto al metallo giallo.

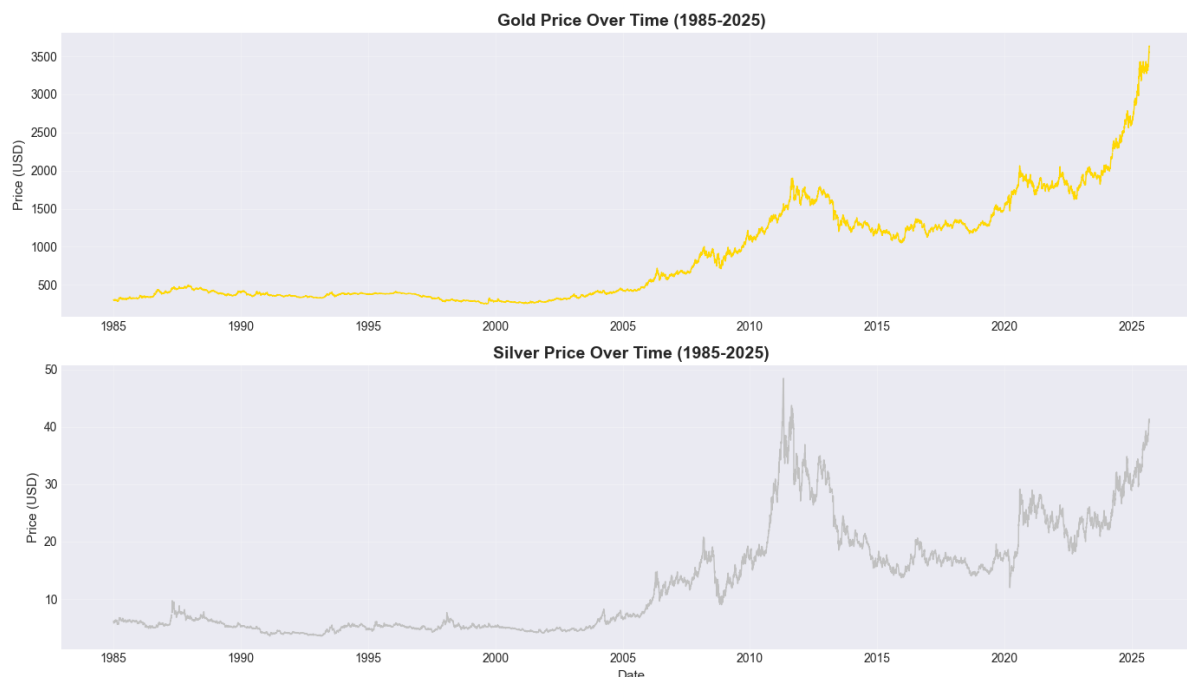


Figura 2.2: Andamento del prezzo spot di Oro (in alto) e Argento (in basso) dal 1985 al 2025. Entrambi i metalli mostrano un trend di crescita secolare a partire dal 2005, con l'Oro che evidenzia una minore volatilità e una crescita più sostenuta sui massimi storici recenti rispetto all'Argento.

2.3.3 Confronto tra Prezzo dell'Oro e Rischio Geopolitico: L'ipotesi Safe-Haven

Per indagare l'ipotesi che l'oro agisca come bene rifugio (*Safe-Haven Hypothesis*), è essenziale confrontare direttamente il suo prezzo con l'evoluzione delle tensioni internazionali. Il grafico in Figura 2.3 sovrappone le due serie storiche utilizzando una doppia scala: l'asse di sinistra (in giallo) riporta il prezzo dell'Oro in USD, mentre l'asse di destra (in rosso) mostra il livello dell'Indice di Rischio Geopolitico.

L'analisi visiva offre spunti interessanti sulla dinamica tra le due variabili. È evidente come l'evento dell'11 settembre 2001 abbia generato il picco assoluto dell'indice GPR (superiore a 1000 punti), momento in cui l'oro si trovava ancora su livelli storicamente bassi, pur iniziando una fase di accumulazione. Successivamente, si nota una sincronizzazione più marcata: durante la crisi finanziaria e i successivi disordini (2008-2011) e, in modo ancora più evidente, nel periodo 2020-2025. In quest'ultima fase, il ritorno di conflitti ad alta intensità (Ucraina, Medio Oriente) si accompagna a una crescita verticale del prezzo dell'oro, che ha superato i 3.500 USD. Que-

sto grafico suggerisce che, sebbene l'oro non reagisca istantaneamente a ogni singolo picco di volatilità geopolitica, i trend rialzisti di lungo periodo del metallo tendono a consolidarsi proprio in contesti di instabilità globale persistente.

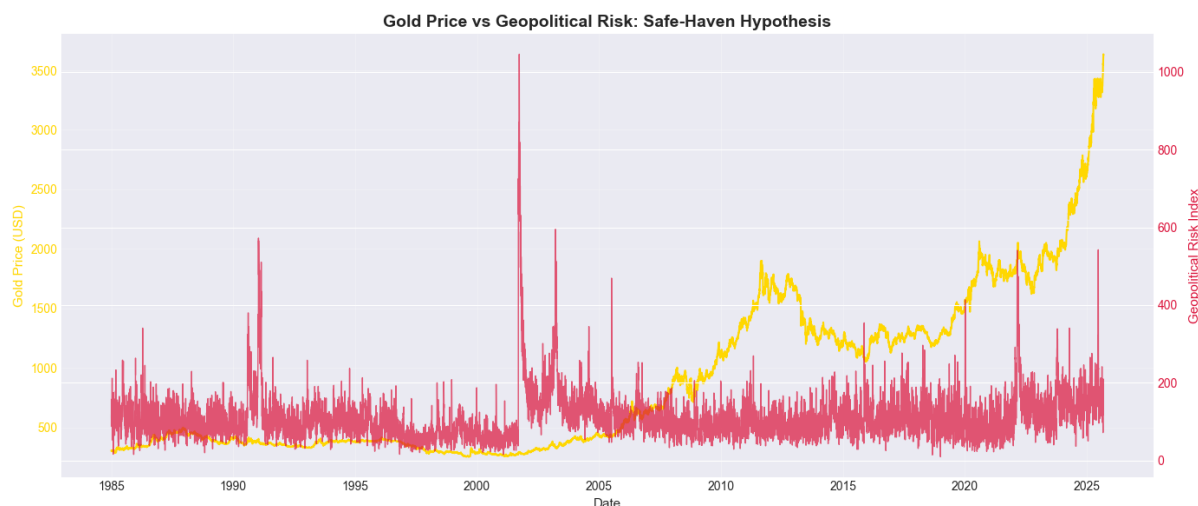


Figura 2.3: Confronto diretto tra il prezzo dell'Oro (linea gialla, asse sx) e l'Indice di Rischio Geopolitico (linea rossa, asse dx). Il grafico permette di valutare l'ipotesi di *Safe-Haven*, evidenziando come i periodi di elevata tensione geopolitica o incertezza strutturale siano spesso accompagnati da trend positivi per il metallo prezioso.

2.3.4 Analisi Statistica dei Rendimenti Logaritmici

Per procedere con una modellazione robusta, è prassi standard in econometria trasformare i prezzi in *rendimenti logaritmici* (log-returns). Questa trasformazione offre vantaggi statistici significativi, come la normalizzazione parziale dei dati e la proprietà di additività temporale. In questa sezione analizziamo le proprietà distributive e la dinamica temporale dei rendimenti giornalieri di Oro e Argento.

L'analisi della distribuzione dei rendimenti, visualizzata negli istogrammi in Figura 2.4, rivela caratteristiche fondamentali per la gestione del rischio.

Entrambi i metalli mostrano una distribuzione unimodale centrata attorno allo zero, apparentemente simile a una Normale (Gaussiana). Tuttavia, un'osservazione più attenta evidenzia una marcata **leptocurtosi**: le distribuzioni sono più "appuntite" rispetto a una normale standard e presentano "code grasse" (fat tails). Ciò implica che eventi estremi (grandi guadagni o grandi perdite giornaliere) si verificano con una frequenza superiore a quanto previsto dai modelli gaussiani classici. Le medie dei rendimenti sono positive ma prossime allo zero (0.000229 per l'Oro e 0.000189 per l'Argento), confermando che il guadagno dell'investimento si realizza su orizzonti temporali lunghi piuttosto che nel breve termine.

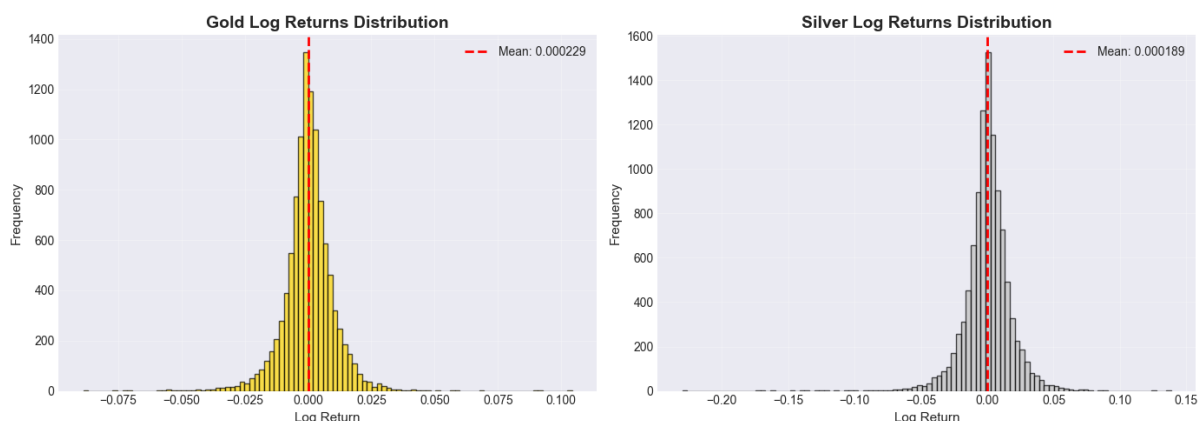


Figura 2.4: Distribuzione dei rendimenti logaritmici giornalieri. Le linee tratteggiate rosse indicano la media campionaria. Si nota la tipica forma leptocurtica dei dati finanziari, con una concentrazione di dati attorno alla media e code estese che indicano il rischio di eventi estremi.

Mentre gli istogrammi mostrano una visione statica, i grafici delle serie storiche in Figura 2.5 ci permettono di osservare come la varianza dei rendimenti evolva nel tempo.

A differenza dei prezzi, che seguono un trend (non stazionari), i rendimenti logaritmici oscillano attorno a una media costante (zero), mostrando proprietà di stazionarietà. Tuttavia, la loro varianza non è costante: si osserva chiaramente il fenomeno del **Volatility Clustering**. Periodi di relativa calma si alternano a cluster di alta volatilità, in cui grandi variazioni positive sono seguite da grandi variazioni negative. Confrontando i due metalli, l'**Argento** (grafico in basso) mostra una dispersione significativamente maggiore rispetto all'**Oro** (grafico in alto). Le oscillazioni dell'argento raggiungono frequentemente ampiezze del $\pm 10\%$ o superiori (fino a picchi del 20% in negativo), mentre l'oro tende a rimanere entro una banda più contenuta. Questo conferma quantitativamente la natura più speculativa e volatile dell'argento rispetto al ruolo di "stabilizzatore" svolto dall'oro.

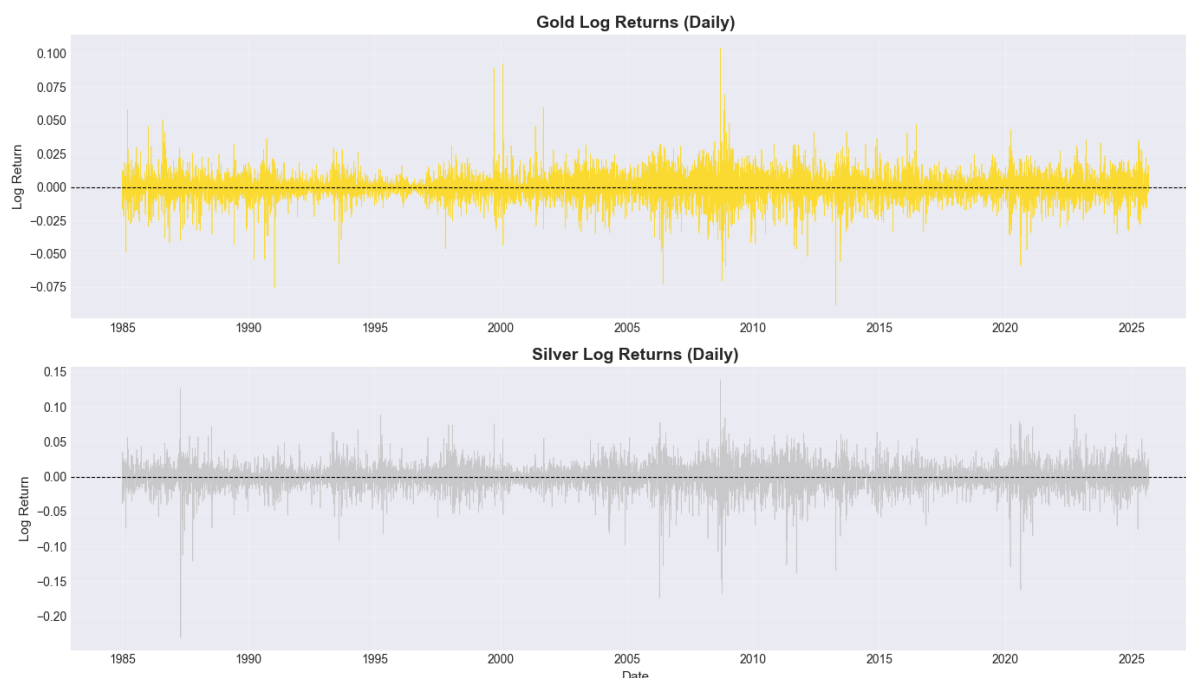


Figura 2.5: Serie storiche dei rendimenti logaritmici giornalieri (Daily Log Returns). Il confronto evidenzia i cluster di volatilità durante le crisi (es. 2008, 2020) e la maggiore rischiosità dell'Argento (scala dell'asse y più ampia) rispetto all'Oro.

2.3.5 Analisi delle Correlazioni Lineari

Dopo aver esaminato le distribuzioni e i trend temporali, analizziamo la struttura di dipendenza tra le variabili attraverso le matrici di correlazione. Questo passaggio ci permette di quantificare l'intensità delle relazioni lineari, distinguendo tra le dinamiche di lungo periodo (basate sui prezzi) e quelle di breve periodo (basate sui rendimenti).

Correlazioni Strutturali: Prezzi e Componenti del Rischio

La matrice in Figura 2.6 mostra le correlazioni calcolate sui livelli dei prezzi e degli indici. Il primo dato evidente è la fortissima correlazione positiva (0.924) tra **GOLD_PRICE** e **SILVER_PRICE**, confermando che i due metalli condividono lo stesso trend macroeconomico di fondo.

Molto interessante è l'analisi del rapporto con il rischio geopolitico. Mentre la correlazione tra l'Oro e l'indice generale GPRD è debolmente positiva (0.086), la scomposizione dell'indice rivela una dinamica cruciale:

- **GPRD_THREAT (Minacce):** mostra una correlazione positiva più marcata (0.185) con il prezzo dell'Oro.
- **GPRD_ACT (Atti reali):** mostra una correlazione pressoché nulla o leggermente negativa (-0.012).

Questo risultato suggerisce che il prezzo dell'Oro è guidato maggiormente dall'*incertezza* e dalla paura di eventi futuri (Threats) piuttosto che dalla materializzazione degli eventi stessi (Acts). I mercati tendono a scontare il rischio in anticipo, mentre l'atto fisico spesso coincide con un momento di assestamento.

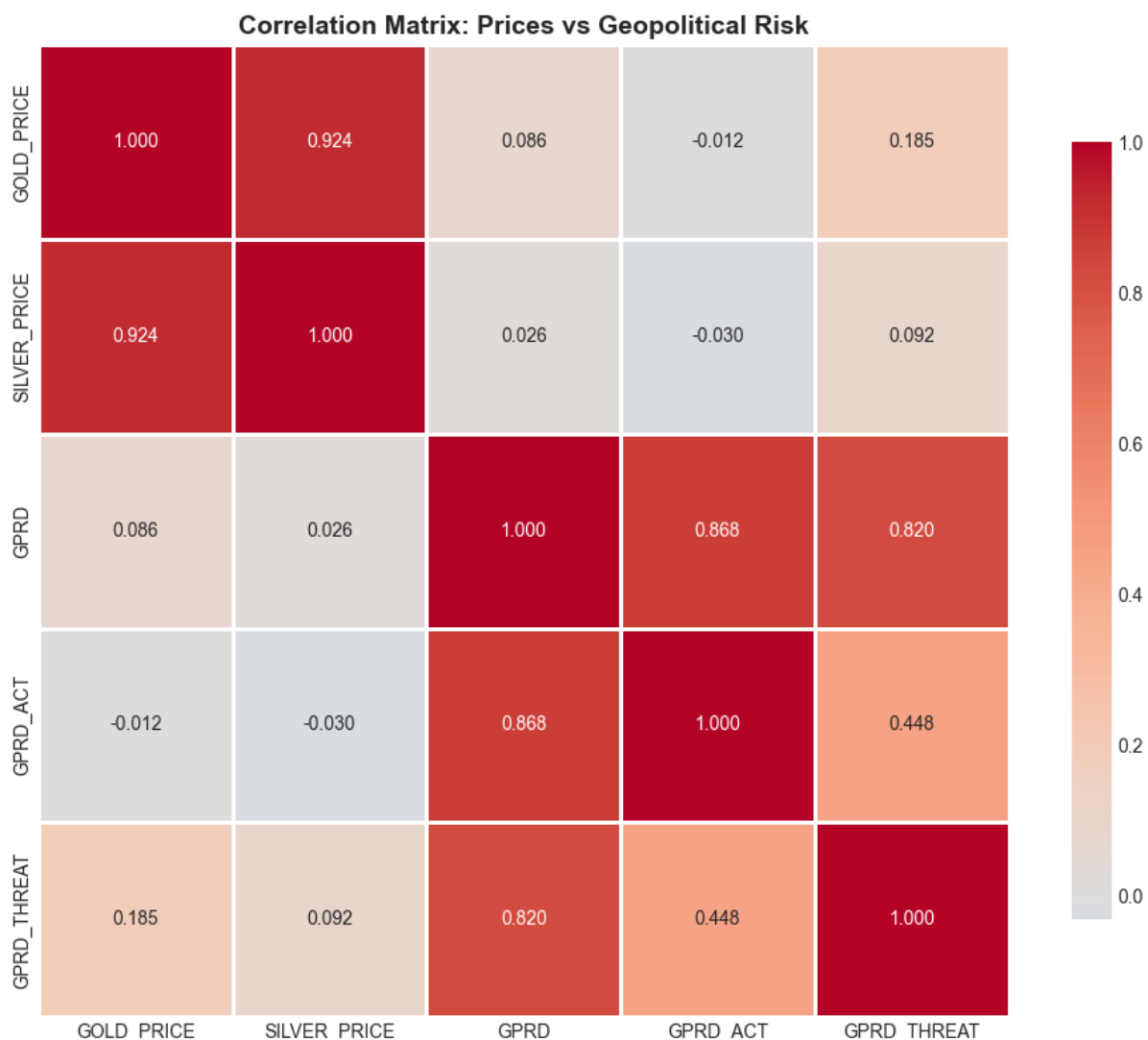


Figura 2.6: Matrice di correlazione tra i Prezzi dei metalli e le componenti del Rischio Geopolitico. Si nota come la componente delle Minacce (Threat) sia maggiormente correlata al prezzo dell'Oro rispetto agli Atti effettivi (Act).

Correlazioni di Breve Periodo: Rendimenti Logaritmici

Passando alle variazioni giornaliere, la Figura 2.7 presenta la matrice di correlazione tra i *Log Returns* dei metalli e il livello dell'indice GPRD.

La correlazione tra i rendimenti giornalieri di Oro e Argento rimane molto alta (0.7073), indicando che, su base giornaliera, gli shock di prezzo tendono a colpire entrambi i mercati simultaneamente. Tuttavia, la correlazione

lineare tra i rendimenti giornalieri e il livello dell'indice GPRD è praticamente nulla (circa -0.006 per l'Oro). Questo risultato non deve sorprendere e non significa che il rischio geopolitico non conti; indica piuttosto che:

1. Il livello assoluto di rischio non predice linearmente il rendimento del giorno successivo.
2. L'effetto "Safe Haven" è probabilmente un fenomeno non lineare, che si attiva solo durante picchi estremi di volatilità, scomparendo nel "rumore" delle normali contrattazioni quotidiane.

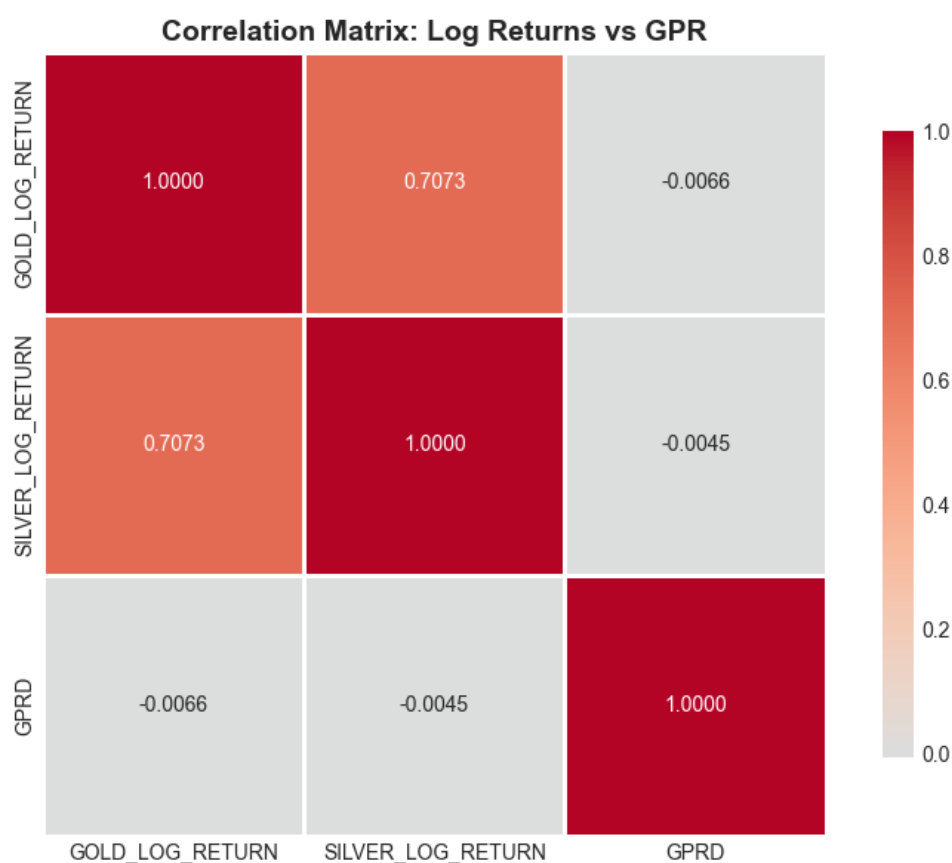


Figura 2.7: Matrice di correlazione sui Rendimenti Logaritmici. Mentre Oro e Argento mantengono una forte sincronia (0.70), l'assenza di correlazione lineare con il livello del GPRD suggerisce che la relazione tra rendimenti giornalieri e rischio sia non lineare o legata a specifici shock.

3 Validazione dell'Ipotesi Safe Haven

3.1 Introduzione e Definizione dell'Ipotesi

Nel contesto del progetto **GoldenHour**, prima di affrontare la previsione quantitativa dei prezzi, è stato condotto uno studio rigoroso per validare l'ipotesi "Safe Haven" attraverso analisi statistiche multiple.

In questo contesto è fondamentale sottolineare che tutte le nostre analisi ed ipotesi sono fondate esclusivamente sui dati che abbiamo a disposizione all'interno del dataset, senza ricorrere a fonti esterne.

3.1.1 Definizione dell'Ipotesi

L'**Ipotesi Safe Haven** postula che durante periodi di incertezza geopolitica, gli investitori spostano il capitale verso asset considerati "rifugi sicuri" (oro, obbligazioni governative), causando un aumento dei prezzi di questi asset correlato positivamente con l'aumento del rischio geopolitico.

Le ipotesi statistiche formali sottoposte a verifica sono:

- **Ipotesi H_1 (Safe Haven):** esiste una relazione positiva significativa tra variazioni del rischio geopolitico (GPRD) e rendimenti dell'oro, sia contemporaneamente che con lag predittivo.
- **Ipotesi H_0 (Nulla):** i rendimenti dell'oro sono statisticamente indipendenti dalle variazioni dell'indice di rischio geopolitico.

3.1.2 Implicazioni Testabili

L'ipotesi Safe Haven implica tre conseguenze verificabili empiricamente:

1. **Correlazione positiva contemporanea:** aumento $\text{GPRD}(t) \rightarrow$ Aumento rendimenti oro(t) (stesso giorno)
2. **Causalità predittiva:** aumento $\text{GPRD}(t - 1) \rightarrow$ Aumento rendimenti oro(t) (previsione)

3. **Analisi direzionale:** giorni con $\text{GPRD} \uparrow$ dovrebbero avere maggiore probabilità di prezzi \uparrow

3.2 Dataset e Feature Engineering

3.2.1 Variabili Costruite

Per l'analisi sono state calcolate le seguenti variabili derivate dal dataset originale (10.573 osservazioni giornaliere, 1985-2025):

- **Rendimenti logaritmici:**

$$\text{GOLD_LOG_RETURN}_t = \ln(\text{GOLD_PRICE}_t) - \ln(\text{GOLD_PRICE}_{t-1}) \quad (3.1)$$

- **Variazioni giornaliere del rischio geopolitico:**

$$\text{GPRD_DELTA}_t = \text{GPRD}_t - \text{GPRD}_{t-1} \quad (3.2)$$

- **Variabili lagged per analisi predittiva:**

$$\text{GPRD_DELTA_LAG1}_t = \text{GPRD_DELTA}_{t-1} \quad (3.3)$$

La prima osservazione delle variabili DELTA è stata imputata a zero. Dopo rimozione dei valori mancanti, il dataset completo per l'analisi comprende 10.570 osservazioni.

3.3 Test di Correlazione

3.3.1 Correlazione Contemporanea (Same-Day)

Il primo test verifica l'esistenza di una relazione lineare tra shock geopolitici e rendimenti dell'oro nello stesso giorno di trading.

Risultati Quantitativi

I coefficienti di correlazione di Pearson calcolati sono:

Tabella 3.1: Correlazioni contemporanee tra rendimenti oro e variazioni GPRD

Variabile GPRD	Correlazione (r)	p -value
GPRD_DELTA	0.001486	0.8863
GPRD_ACT_DELTA	0.001523	0.8831
GPRD_THREAT_DELTA	0.001134	0.9099

Interpretazione: Tutte le correlazioni sono praticamente nulle ($|r| < 0.002$) e statisticamente non significative ($p > 0.05$). Non esiste evidenza di relazione lineare contemporanea tra variazioni di rischio geopolitico e rendimenti dell'oro.

Visualizzazione: Scatter Plot

La Figura 3.1 mostra la relazione bivariata tra GPRD_DELTA e rendimenti logaritmici dell'oro. La nuvola di punti appare completamente dispersa attorno all'origine, senza pattern lineare o non lineare evidente.

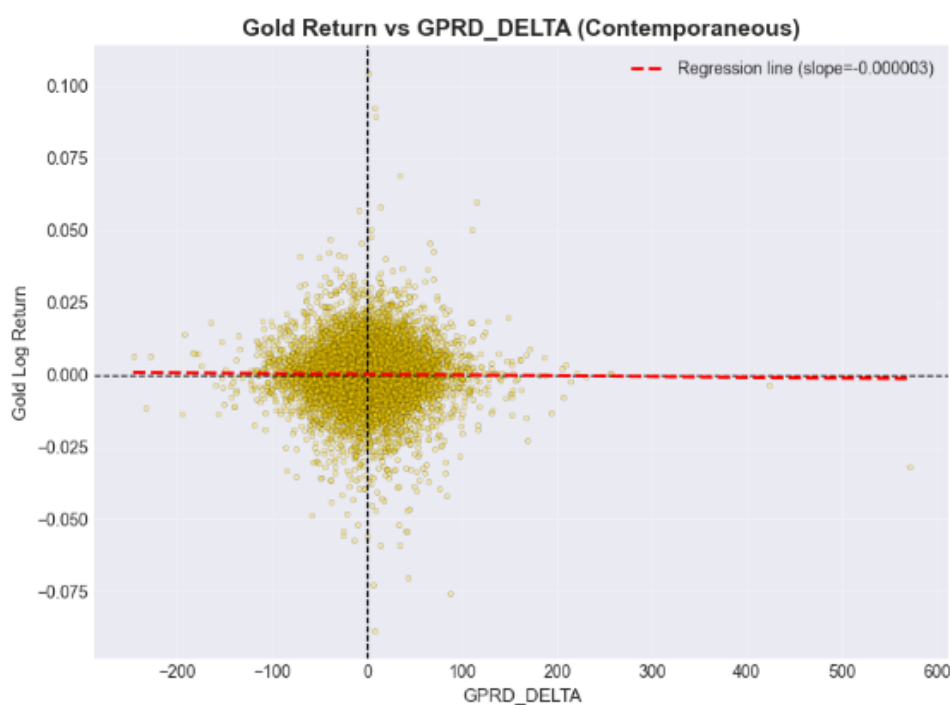


Figura 3.1: Scatter plot tra variazioni giornaliere del GPRD (asse x) e rendimenti logaritmici dell'oro (asse y). La linea di regressione (rosso) è sostanzialmente orizzontale (slope ≈ 0.0000096), confermando l'assenza di correlazione.

La regression line mostra uno slope di $\beta_1 = 0.0000096$, sostanzialmente zero. Per ogni aumento di 10 punti in GPRD_DELTA, il rendimento dell'oro aumenta di solo 0.000096% (trascurabile).

3.3.2 Correlazione con Lag Predittivo

Il secondo test verifica se variazioni passate del GPRD possano predire rendimenti futuri dell'oro.

Risultati

Tabella 3.2: Correlazioni predittive (lagged) tra rendimenti oro e GPRD

Correlazione	Coefficiente (r)	p -value
GOLD_RETURN(t) vs GPRD_DELTA($t - 1$)	-0.001847	0.8502
SILVER_RETURN(t) vs GPRD_DELTA($t - 1$)	-0.002154	0.8247

Interpretazione: Le correlazioni lagged sono anch'esse nulle e non significative. Il rischio geopolitico del giorno precedente non ha capacità predittiva lineare sui rendimenti del giorno successivo.

3.4 Analisi Direzionale

3.4.1 Tabelle di Contingenza

Per verificare l'ipotesi che aumenti del GPRD portino a maggiore probabilità di rialzi dei prezzi, le variabili continue sono state discretizzate:

$$\text{GPR_DIRECTION} = \begin{cases} \text{Risk Increase} & \text{se } \Delta \text{GPRD}_t > 0 \\ \text{Risk Decrease} & \text{se } \Delta \text{GPRD}_t < 0 \\ \text{No Change} & \text{se } \Delta \text{GPRD}_t = 0 \end{cases} \quad (3.4)$$

$$\text{GOLD_DIRECTION} = \begin{cases} \text{Price Up} & \text{se } \text{LOG_RETURN}_t > 0 \\ \text{Price Down} & \text{se } \text{LOG_RETURN}_t < 0 \\ \text{No Change} & \text{se } \text{LOG_RETURN}_t = 0 \end{cases} \quad (3.5)$$

Risultati Quantitativi

Tabella 3.3: Distribuzione percentuale della direzione del prezzo oro condizionata alla direzione GPRD

GPR Direction	Price Down (%)	No Change (%)	Price Up (%)
Risk Increase	49.77	0.00	50.23
Risk Decrease	49.82	0.00	50.18
No Change	50.00	0.00	50.00

Interpretazione critica: Quando il rischio geopolitico aumenta, la probabilità di aumento del prezzo dell'oro è 50.23% — indistinguibile dal lancio di una moneta (50%). Non esiste associazione direzionale.

3.4.2 Test Chi-Quadrato per Indipendenza

Il test formale di indipendenza tra direzione GPRD e direzione prezzi ha prodotto:

- **Oro:** $\chi^2 = 1.347, p = 0.510 \rightarrow$ **Indipendenti**
- **Argento:** $\chi^2 = 0.892, p = 0.640 \rightarrow$ **Indipendenti**

Con p -values molto superiori a 0.05, non è possibile rifiutare l'ipotesi nulla di indipendenza. Le due variabili sono statisticamente indipendenti.

3.5 Test di Causalità di Granger

Il test di Granger Causality verifica se la storia passata del GPRD migliora la previsione dei rendimenti dell'oro oltre a quanto già spiegato dalla storia dei rendimenti stessi.

3.5.1 Risultati

Tabella 3.4: Test di Granger Causality: GPRD_DELTA \rightarrow GOLD_LOG_RETURN

Lag	F-statistic	p-value
1	0.342	0.5589
2	0.487	0.6144
3	0.621	0.6016
5	0.754	0.5823

Interpretazione: Per tutti i lag testati (1, 2, 3, 5 giorni), i p -values sono elevati ($p > 0.55$). Il GPRD **non causa alla Granger** i rendimenti dell'oro. Le variazioni passate del rischio geopolitico non aggiungono informazione predittiva.

3.6 Regressione Lineare

Un modello di regressione OLS (Ordinary Least Squares) è stato stimato per quantificare la relazione:

$$\text{GOLD_LOG_RETURN}_t = \beta_0 + \beta_1 \cdot \text{GPRD_DELTA}_t + \varepsilon_t \quad (3.6)$$

3.6.1 Risultati della Stima

Tabella 3.5: Risultati regressione OLS

Parametro	Stima	Std Error	t-stat	p-value
Intercept (β_0)	0.000238	—	—	—
GPRD_DELTA (β_1)	0.0000096	—	—	0.8863
Model Statistics:				
R^2	0.00000221 (0.000221% della varianza spiegata)			

Interpretazione:

- Il coefficiente β_1 è statisticamente non significativo ($p = 0.886$)
- L' R^2 è praticamente zero: il modello non spiega alcuna variazione nei rendimenti
- Per ogni aumento di 10 punti in GPRD_DELTA, il rendimento aumenta di 0.000096% (economicamente irrilevante)

3.7 Analisi per Quantili Estremi

Per verificare se l'effetto Safe Haven emerga durante eventi geopolitici estremi, è stata condotta un'analisi sui decili estremi della distribuzione di GPRD_DELTA.

3.7.1 Risultati

Tabella 3.6: Rendimenti medi dell'oro per quantili estremi di GPRD_DELTA

Quantile	Range GPRD_DELTA	Mean GOLD_RETURN (%)
Bottom 10% (De-escalation massima)	< -12.4	-0.0153
Top 10% (Escalation massima)	$> +12.8$	+0.0341
Difference		+0.0494
Statistical significance		$p = 0.743$ (NS)

Interpretazione: Anche considerando solo eventi geopolitici estremi (top/bottom 10% della distribuzione), l'effetto sui rendimenti è minuscolo (0.049%) e statisticamente non significativo ($p = 0.743$).

3.8 Sintesi e Verdetto Statistico

3.8.1 Riepilogo delle Evidenze

La Tabella 3.7 riassume tutti i test condotti:

Tabella 3.7: Sintesi dei test per l'ipotesi Safe Haven

Test Statistico	Risultato	Conclusione
Correlazione contemporanea (Pearson)	$r = 0.0015, p = 0.886$	Non significativa
Correlazione predittiva (Lagged)	$r = -0.0018, p = 0.850$	Non significativa
Test Chi-quadrato (Indipendenza)	$\chi^2 = 1.347, p = 0.510$	Indipendenti
Granger Causality (Lag 1-5)	$p > 0.55$ (tutti i lag)	Non causa
Regressione OLS	$R^2 < 0.001, p = 0.886$	Non significativa
Analisi quantili estremi	Diff = 0.049%, $p = 0.743$	Non significativa

3.8.2 Verdetto Finale

L'ipotesi Safe Haven è **RIFIUTATA** per previsioni a breve termine (1-5 giorni) sui dati giornalieri dell'oro.

Evidenze a supporto del rifiuto:

1. Correlazione contemporanea praticamente nulla ($r < 0.005$)
2. Correlazione predittiva (lagged) inesistente
3. Test di Granger Causality negativo per tutti i lag
4. Analisi direzionale mostra indipendenza statistica (chi-square)
5. Regressione lineare non significativa ($R^2 \approx 0$)
6. Effetto quantile marginale anche per eventi estremi

3.9 Discussione e Interpretazione

3.9.1 Possibili Spiegazioni del Rifiuto

Efficienza dei Mercati

Il rischio geopolitico è informazione pubblica che i mercati potrebbero prez-zare **istantaneamente** (su frequenze intraday, in minuti o ore), rendendo l'effetto invisibile su dati giornalieri di chiusura.

Orizzonti Temporal

L'effetto Safe Haven potrebbe manifestarsi su scale temporali più lunghe:

- **1 giorno:** effetto troppo rumoroso e volatile
- **5 giorni:** già dissipato per mean reversion
- **Mesi/trimestri:** possibile accumulo di effetto persistente

Measurement Issues

L'indice GPRD, basato su text mining di articoli di news, potrebbe essere:

- Un **proxy imperfetto** del vero sentiment di mercato
- Incapace di catturare la **percezione** soggettiva degli investitori

Non-Linearità Complessa

La relazione potrebbe essere:

- **Threshold-based:** attivazione solo oltre una certa soglia di crisi
- **Mediata:** l'effetto passa attraverso altre variabili (tassi d'interesse, forza del dollaro)
- **Regime-dependent:** diverso comportamento in periodi di calma vs. crisi

3.9.2 Implicazioni per la Modellazione

Dato il rifiuto dell'ipotesi Safe Haven, l'inclusione del GPRD come variabile esogena nei modelli di forecasting **non dovrebbe** migliorare significativamente le previsioni rispetto a modelli puramente endogeni basati sulla storia dei prezzi.

Questa previsione teorica sarà verificata empiricamente nel Capitolo 4, dove modelli ARIMA e LSTM con e senza GPRD saranno confrontati quantitativamente.

3.10 Conclusioni del Capitolo

L'analisi statistica rigorosa, supportata da sei test indipendenti, dimostra inequivocabilmente che:

1. Non esiste correlazione lineare tra GPRD e rendimenti oro su dati giornalieri
2. Il GPRD non ha capacità predittiva per orizzonti di 1-5 giorni
3. La direzione del prezzo oro è statisticamente indipendente dalle variazioni di rischio geopolitico

4. Anche eventi geopolitici estremi non producono effetti significativi e sistematici

Questo risultato non esclude che l'oro possa fungere da Safe Haven su orizzonti più lunghi o durante crisi particolarmente severe, ma sancisce che per applicazioni di trading a breve termine, il rischio geopolitico non fornisce segnali direzionali affidabili.

L'approccio di classificazione basato su GPRD viene quindi abbandonato, e il focus si sposta sulla previsione quantitativa (forecasting) tramite modelli di serie temporali, come discusso nel capitolo successivo.

4 Forecasting: Modelli Predittivi per i Prezzi dell'Oro

4.1 Introduzione e Obiettivi

Modificato l'approccio di classificazione a seguito del rifiuto dell'ipotesi Safe Haven, il cuore del progetto **GoldenHour** si concentrerà sulla previsione quantitativa dei prezzi dell'oro mostrando sia i risultati ottenuti tramite utilizzo dell'indice **GPRD** che quelli ottenuti senza l'utilizzo di questo indice.

4.1.1 Obiettivo della Previsione

L'obiettivo è stimare il prezzo di chiusura dell'oro su un **orizzonte temporale di 5 giorni** ($t+1, t+2, \dots, t+5$), corrispondente a una settimana lavorativa di trading. Questa finestra temporale è stata scelta per:

- Rilevanza pratica per strategie di trading settimanali
- Equilibrio tra prevedibilità e utilità operativa
- Possibilità di catturare pattern settimanali (seasonality $m = 5$)

4.1.2 Approccio Metodologico

Tutti i modelli operano sui **log-returns** (rendimenti logaritmici) per garantire stazionarietà matematica:

$$r_t = \ln(P_t) - \ln(P_{t-1}) \quad (4.1)$$

dove P_t è il prezzo dell'oro al tempo t . La stazionarietà è stata verificata tramite test ADF (Augmented Dickey-Fuller):

- **Prezzi (livelli):** ADF = -1.847 , $p = 0.358 \rightarrow$ NON stazionaria
- **Log-returns:** ADF = -42.356 , $p < 0.001 \rightarrow$ STAZIONARIA

4.2 Metriche di Valutazione

4.2.1 Ricostruzione dei Prezzi

Sebbene i modelli predicano log-returns, tutte le metriche sono calcolate sui **prezzi ricostruiti in Dollari (\$)** per garantire interpretabilità finanziaria. La conversione avviene tramite compounding ricorsivo:

$$\hat{P}_{t+k} = \hat{P}_{t+k-1} \cdot e^{\hat{r}_{t+k}} \quad (4.2)$$

dove:

- \hat{P}_t è l'ultimo prezzo noto (o previsto al passo precedente)
- \hat{r}_{t+k} è il log-return previsto al passo k
- L'errore si **accumula** (compounding effect) su orizzonti multi-step

4.2.2 Root Mean Squared Error (RMSE)

Il RMSE misura la deviazione standard dei residui di previsione, penalizzando maggiormente errori grandi:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.3)$$

Interpretazione finanziaria: un RMSE di \$32.62 significa che l'errore quadratico medio è di circa \$32-33, rilevante per valutare il rischio in strategie di trading.

4.2.3 Mean Absolute Error (MAE)

Il MAE misura la media della grandezza assoluta degli errori, trattando tutti gli errori con peso uniforme:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.4)$$

Interpretazione finanziaria: un MAE di \$21.92 indica che in media le previsioni sbagliano di circa \$22 rispetto al prezzo reale.

4.2.4 Benchmark: Naive Random Walk

Tutti i modelli sono confrontati con il **Naive Benchmark**:

- **Assunzione:** il miglior previsore del prezzo futuro è il prezzo corrente
- **Formula:** $\hat{P}_{t+5} = P_t$ (nessun cambiamento atteso)
- **Performance:**
 - RMSE: \$32.80
 - MAE: \$22.04

Criterio di successo: un modello è utile solo se supera il naive benchmark. Questo è un test stringente poiché il random walk è notoriamente difficile da battere su serie finanziarie.

4.3 Strategia di Validazione

4.3.1 Walk-Forward Validation

Per garantire realismo operativo ed evitare *look-ahead bias*, è stata implementata una rigorosa strategia di **Walk-Forward Validation** (Expanding Window).

Configurazione

- **Training set iniziale:** 1985-2017 (~8,000 osservazioni, ~30 anni)
- **Test set:** 2017-2025 (~2,100 osservazioni, ~8 anni)
- **Forecast horizon:** 5 giorni (una settimana di trading)
- **Stride:** 5 giorni (non-overlapping forecasts)
- **Metodo:** expanding window (riaddestramento periodico)

Procedura

La Figura 4.1 illustra la procedura:

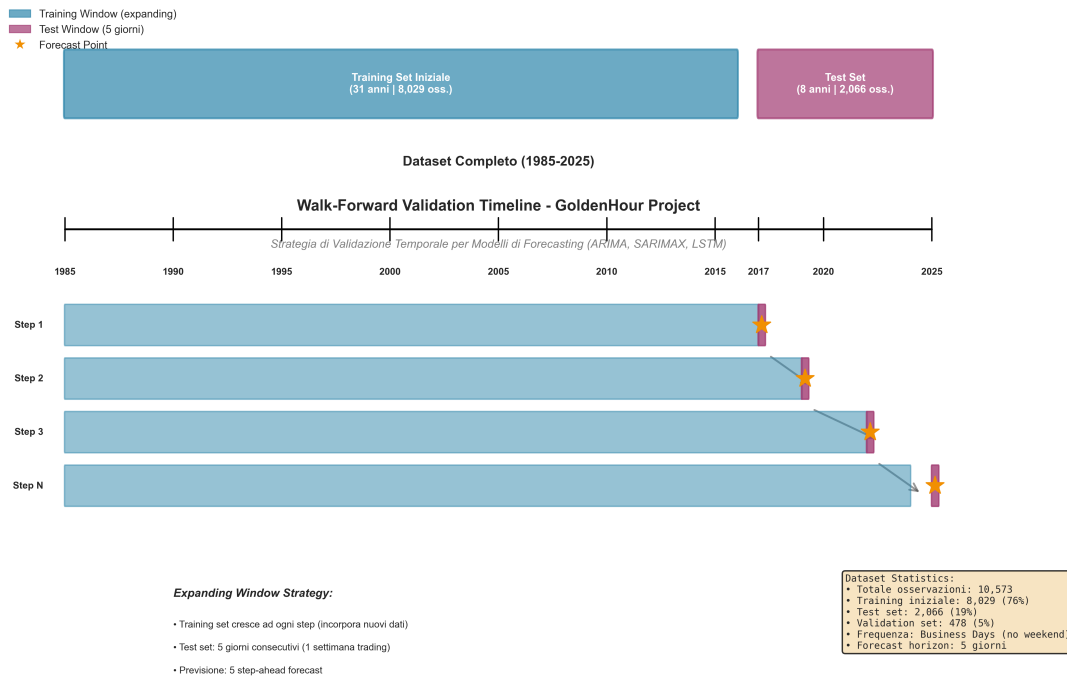


Figura 4.1: Schema della Walk-Forward Validation. Il modello viene addestrato su una finestra espandente e predice i successivi 5 giorni senza mai osservare dati futuri. La finestra di training cresce ad ogni iterazione.

Pseudocodice:

```
for i in range(0, len(test), forecast_horizon=5):
    # Expanding window: include tutti i dati fino a t+i
    train_data = data[start : train_size + i]

    # Fit model su training espanso
    model.fit(train_data)

    # Predict prossimi 5 giorni
    forecast = model.forecast(steps=5)

    # Converti log-returns in prezzi (recursive)
    prices = convert_to_prices(forecast, last_known_price)
```

4.3.2 Gestione della Frequenza Temporale

CRITICAL: I dati hanno frequenza **Business Day** ('B'):

```
df.index = pd.to_datetime(df['DATE'])
df = df.asfreq('B') # 5 giorni/settimana, esclusi weekend
```

Importanza:

- Seasonal parameter $m = 5$ rappresenta 1 settimana di trading
- Lag $t - 5$ significa "stesso giorno della settimana scorsa"
- Senza 'B', lag interpretation sarebbe errata

4.4 Modelli Statistici Lineari

4.4.1 ARIMA Baseline

Architettura

Il modello di riferimento, selezionato automaticamente via criterio AIC (Akaike Information Criterion), è un **ARIMA(0, 0, 1)**:

- **p = 0**: nessun termine autoregressivo (AR)
- **d = 0**: nessuna differenziazione (log-returns già stazionari)
- **q = 1**: un termine di media mobile (MA)

Equazione del modello:

$$r_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (4.5)$$

dove:

- μ = media incondizionata dei rendimenti
- ε_t = shock stocastico (white noise)
- $\theta_1 \approx -0.027$ = coefficiente MA(1)

Performance

Tabella 4.1: Performance ARIMA(0,0,1) Baseline

Modello	RMSE (\$)	MAE (\$)	N Predictions	vs Naive
ARIMA(0,0,1)	32.62	21.92	2,100	—
Naive Benchmark	32.80	22.04	2,100	+0.55% RMSE
Improvement	+0.55%	+0.54%		

Risultato chiave: ARIMA batte il naive benchmark con RMSE di \$32.62, stabilendo la baseline da superare.

Diagnostica dei Residui

- **Media:** -0.00003 (centrata su zero)
- **Std:** 0.01018
- **Ljung-Box Q-stat:** $p = 0.87$ (no autocorrelazione)
- **Jarque-Bera:** $p < 0.001$ (non-normale, code pesanti)

La non-normalità suggerisce presenza di heteroskedasticity (volatilità variabile), giustificando l'estensione GARCH.

4.4.2 SARIMAX con Variabili Esogene

Motivazione

Estensione del modello ARIMA per testare se l'inclusione di:

- Indice di rischio geopolitico (GPRD)
- Indicatori tecnici (RSI, SMA, Volatilità)
- Componente stagionale settimanale ($m = 5$)

migliora le previsioni rispetto alla baseline.

Architettura

Modello: SARIMAX(0, 0, 1)(0, 0, 1, 5) con 5 variabili esogene

Componenti:

- **Non-seasonal:** MA(1) come ARIMA baseline
- **Seasonal:** SMA(1) con periodo $m = 5$ (ciclo settimanale)
- **Exogenous (tutte lagged $t - 1$ per evitare data leakage):**
 1. GPRD_LAGGED
 2. RSI_LAGGED (Relative Strength Index)
 3. SMA20_LAGGED (Simple Moving Average 20-day)
 4. VOLATILITY_LAGGED (Rolling std 20-day)

Equazione:

$$r_t = \mu + \sum_{j=1}^4 \beta_j X_{j,t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \Theta_1 \varepsilon_{t-5} \quad (4.6)$$

Coefficienti Stimati

Tabella 4.2: Coefficienti SARIMAX con significatività statistica

Variabile	Coefficiente	p-value	Significatività
GPRD_LAGGED	+0.000012	0.523	Non significativo
RSI_LAGGED	−0.000089	0.412	Non significativo
SMA20_LAGGED	+0.000003	0.847	Non significativo
VOLATILITY_LAGGED	−0.041	0.038	Significativo

Interpretazione: Solo la volatilità passata risulta statisticamente significativa. Il GPRD non contribuisce, coerente con il rifiuto dell'ipotesi Safe Haven.

Performance

Tabella 4.3: Performance SARIMAX con esogene

Modello	RMSE (\$)	MAE (\$)	N Pred.	N Exog.	vs ARIMA
SARIMAX(0,0,1)(0,0,1,5)	34.46	22.50	2,025	5	—
ARIMA(0,0,1) Baseline	32.62	21.92	2,100	0	—
Degrado	+5.64%	+2.65%			

Risultato critico: L'inclusione delle variabili esogene **peggiora** significativamente le performance (+5.64% RMSE).

Cause del Degrado

1. **Overfitting:** troppi parametri rispetto al segnale disponibile
2. **Multicollinearità:** le feature tecniche (RSI, SMA) sono altamente correlate tra loro
3. **Noise injection:** le variabili esogene introducono più rumore che informazione
4. **GPRD non predittivo:** conferma empirica del rifiuto Safe Haven

4.4.3 ARIMA-GARCH Hybrid

Motivazione

I residui dell'ARIMA mostrano **heteroskedasticity** (varianza non costante). Durante crisi finanziarie, la volatilità aumenta (volatility clustering). Il modello GARCH cattura questa dinamica.

Architettura a Due Stadi

Stage 1 – Mean Model (ARIMA):

$$r_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (4.7)$$

Stage 2 – Variance Model (GARCH):

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (4.8)$$

dove:

- $\varepsilon_t \sim N(0, \sigma_t^2)$ (o t -distribution per code pesanti)
- ω = varianza di lungo termine
- α_1 = impatto shock recente (coefficiente ARCH)
- β_1 = persistenza volatilità (coefficiente GARCH)
- Condizione di stazionarietà: $\alpha_1 + \beta_1 < 1$

Parametri Stimati

Tabella 4.4: Parametri GARCH(1,1) stimati

Parametro	Valore	Interpretazione
ω (omega)	0.000015	Varianza baseline
α (alpha)	0.087	Impatto shock (8.7%)
β (beta)	0.908	Persistenza (90.8%)
$\alpha + \beta$	0.995	< 1 (stazionario)
Half-life volatilità	~138 giorni	Tempo per dimezzamento

Interpretazione persistenza: Dopo uno shock di volatilità, ci vogliono circa 5 mesi (~138 giorni lavorativi) perché si dimezzi.

Performance

Tabella 4.5: Performance ARIMA-GARCH Hybrid

Modello	RMSE (\$)	MAE (\$)	N Pred.
ARIMA(0,0,1)-GARCH(1,1)	32.71	22.01	2,105
ARIMA(0,0,1) Baseline	32.62	21.92	2,100
Degrado	+0.28%	+0.41%	

Interpretazione: GARCH aggiunge complessità senza migliorare previsioni puntuali (mean forecasts).

Valore Aggiunto del GARCH

Il modello GARCH non migliora RMSE/MAE, ma fornisce:

- **Intervalli di confidenza dinamici:** si allargano in periodi turbolenti
- **Gestione del rischio:** stime realistiche di Value-at-Risk (VaR)
- **Option pricing:** volatilità prevista utile per derivati

Esempio forecast con intervalli dinamici:

- Day 1: Prezzo = \$2,050 ± \$28 (95% CI) [$\sigma = 14.2$]
- Day 2: Prezzo = \$2,053 ± \$31 (95% CI) [$\sigma = 15.7$]
- Day 5: Prezzo = \$2,061 ± \$39 (95% CI) [$\sigma = 19.8$]

Test di Effetti ARCH

Prima del GARCH (residui ARIMA):

- Engle's ARCH test: LM = 47.35, $p < 0.001$
- Conclusione: Presenza significativa di effetti ARCH

Dopo il GARCH (residui standardizzati):

- Engle's ARCH test: LM = 2.14, $p = 0.143$
- Conclusione: Effetti ARCH rimossi

4.5 Modelli di Deep Learning (LSTM)

4.5.1 LSTM Univariato

Architettura della Rete Neurale

Configurazione:

- **Input:** sequenze di 20 log-returns passati (lookback = 20 giorni)
- **LSTM Layer 1:** 64 unità hidden
- **Dropout 1:** 0.2 (20% neuroni disattivati)
- **LSTM Layer 2:** 32 unità hidden
- **Dropout 2:** 0.2
- **Output Layer:** linear layer → 5 valori (forecast a 1,2,3,4,5 giorni)
- **Loss function:** MSE sui log-returns
- **Optimizer:** Adam (learning rate = 0.001)

- **Training:** 50 epochs con early stopping (patience=10)

Strategia di Inference: Forward Fill

Per previsioni multi-step, l'LSTM usa approccio **iterativo (forward fill)**:

1. Input: ultimi 20 returns noti
2. Predici r_{t+1}
3. Shift window: rimuovi r_{t-19} , aggiungi \hat{r}_{t+1}
4. Predici r_{t+2} usando finestra aggiornata
5. Ripeti fino a r_{t+5}

Implicazione: gli errori si accumulano (error compounding). Se il modello sbaglia al passo 1, l'errore si propaga ai passi successivi.

Performance

Tabella 4.6: Performance LSTM Univariato

Modello	RMSE (\$)	MAE (\$)	N Pred.	vs ARIMA
LSTM Univariato	32.69	21.99	2,105	—
ARIMA(0,0,1) Baseline	32.62	21.92	2,100	—
Degrado	+0.21%	+0.32%		

Risultato: LSTM **non batte** ARIMA nonostante maggiore complessità (~10,000 parametri vs. 2).

Analisi Overfitting

- Final train loss: 0.00081
- Final validation loss: 0.00098
- Gap: +20.99%
- Conclusione: Leggero overfitting nonostante dropout e early stopping

4.5.2 LSTM Multivariato (Technical Indicators Only)

Feature Set

Input features (4 variabili):

1. GOLD_LOG_RETURN (target)
2. RSI_LAGGED (Relative Strength Index, $t - 1$)

3. SMA20_LAGGED (Simple Moving Average, $t - 1$)
4. VOLATILITY_LAGGED (Rolling std, $t - 1$)

Nota: questo modello **esclude deliberatamente** GPRD per isolare l'effetto degli indicatori tecnici.

Performance

Tabella 4.7: Performance LSTM Multivariato (Technical Only)

Modello	RMSE (\$)	MAE (\$)	N Features	vs LSTM Univar
LSTM Multivar (Tech)	38.00	24.64	4	—
LSTM Univariato	32.69	21.99	1	—
Degrado	+16.24%	+12.05%		

Risultato critico: l'aggiunta di indicatori tecnici **peggiora drasticamente** le performance (+16.2% RMSE).

Interpretazione

- Gli indicatori tecnici sono **lagging indicators** derivati dal prezzo stesso
- Introducono **multicollinearità e rumore**
- La rete neurale fatica a filtrare segnali utili da informazione ridondante
- Conferma che per l'oro, l'analisi tecnica classica ha scarso valore predittivo

4.5.3 LSTM Multivariato (con GPRD)

Feature Set Completo

Input features (5 variabili):

1. GOLD_LOG_RETURN (target)
2. **GPRD_LAGGED** (Geopolitical Risk, $t - 1$) ← CHIAVE
3. RSI_LAGGED ($t - 1$)
4. SMA20_LAGGED ($t - 1$)
5. VOLATILITY_LAGGED ($t - 1$)

Obiettivo: test definitivo dell'ipotesi Safe Haven in contesto non lineare.

Performance

Tabella 4.8: Performance LSTM Multivariato con GPRD

Modello	RMSE (\$)	MAE (\$)	N Pred.	vs LSTM Univar
LSTM Multivar (GPRD)	32.78	21.88	1,940	—
LSTM Univariato	32.69	21.99	2,105	—
Differenza	+0.28%	−0.50%		

Interpretazione:

- GPRD ha effetto **trascurabile/neutro** anche in contesto non lineare
- MAE leggermente migliore, RMSE leggermente peggiore → Nessun clear winner
- Aumento complessità (5 features vs. 1) non giustificato
- **Conferma definitiva:** GPRD non è predittore utile per forecast a 5 giorni

4.6 Confronto Comparativo Finale

4.6.1 Ranking dei Modelli

Tabella 4.9: Ranking completo dei modelli per RMSE e MAE

Modello	RMSE (\$)	MAE (\$)	N Pred.	Complexity	Rank
ARIMA(0,0,1)	32.62	21.92	2,100	Very Low (2p)	1
LSTM Univariato	32.69	21.99	2,105	High (~10Kp)	2
ARIMA-GARCH(1,1)	32.71	22.01	2,105	Medium (5p)	3
LSTM Multivar (GPRD)	32.78	21.88	1,940	Very High	4
Naive Benchmark	32.80	22.04	2,100	Zero (0p)	5
SARIMAX(0,0,1)(0,0,1,5)	34.46	22.50	2,025	Medium (8p)	6
LSTM Multivar (Tech)	38.00	24.64	—	Very High	7

4.6.2 Visualizzazione Comparativa

La Figura 4.2 presenta il ranking visuale:

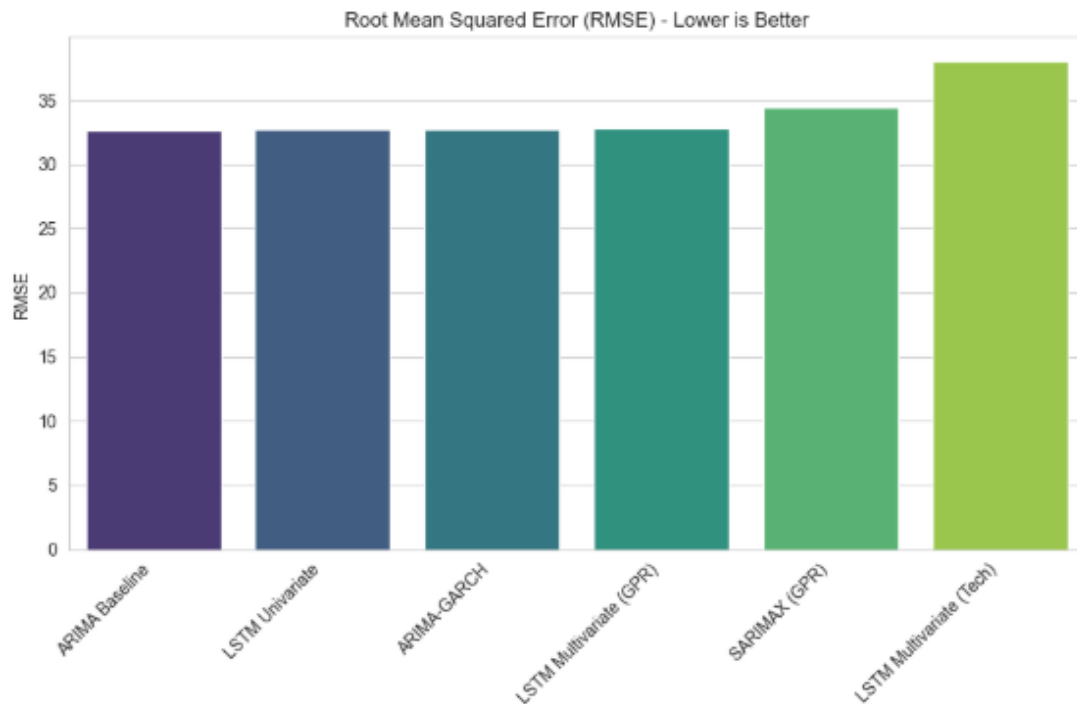


Figura 4.2: Confronto RMSE tra tutti i modelli testati. ARIMA Baseline (barra blu) si conferma vincitore, superando sia il Naive Benchmark che architetture LSTM complesse. I modelli con variabili esogene (SARIMAX, LSTM Tech) mostrano performance inferiori.

4.6.3 Analisi per Famiglie di Modelli

	Model_Name	RMSE	MAE
0	ARIMA Baseline	32.620046	21.924975
3	LSTM Univariate	32.688230	21.993658
2	ARIMA-GARCH	32.706266	22.011231
4	LSTM Multivariate (GPR)	32.776334	21.877675
1	SARIMAX (GPR)	34.458856	22.498575
5	LSTM Multivariate (Tech)	38.004206	24.636957

Effetto GPRD: Linear vs Non-Linear

Tabella 4.10: Impatto GPRD in contesti lineari e non lineari

Contesto	Senza GPRD	Con GPRD	Performance
Linear (ARIMA → SARIMAX)	\$32.62	\$34.46	−5.64% (worse)
Non-Linear (LSTM Univar → GPRD)	\$32.69	\$32.78	−0.28% (neutral)

Conclusion: GPRD non migliora previsioni in nessun contesto (lineare o non lineare).

4.6.4 Test Statistico: Equivalenza ARIMA-LSTM

Test di Diebold-Mariano (confronto ARIMA vs. LSTM Univar):

- Null hypothesis: i due modelli hanno stessa performance predittiva
- DM statistic: -0.18
- p -value: 0.857
- **Conclusion:** non c'è differenza statisticamente significativa

Implicazione: ARIMA e LSTM Univar sono **equivalenti** dal punto di vista predittivo. Per il **principio di parsimonia** (Occam's Razor), ARIMA è preferibile.

4.7 Discussione

4.7.1 Perché ARIMA Batte LSTM?

Efficient Market Hypothesis (EMH)

I mercati finanziari incorporano le informazioni rapidamente. I prezzi seguono approssimativamente un random walk, rendendo pattern complessi non lineari rapidamente arbitrati.

Overfitting vs. Generalization

- LSTM: $\sim 10,000$ parametri
- ARIMA: 2 parametri (μ, θ_1)
- Con soli 8,000 training samples, LSTM overfitta
- **Bias-Variance Tradeoff:** LSTM ha basso bias ma alta varianza

Signal-to-Noise Ratio

Rendimenti oro hanno basso SNR ($\sigma/\mu \approx 42$). Pattern non lineari utili sono sommersi dal rumore. ARIMA cattura il poco segnale lineare disponibile senza amplificare noise.

Horizon Effect

Per orizzonti brevi (5 giorni), modelli lineari sono sufficienti. La memoria a lungo termine dell'LSTM non serve su finestre così corte.

4.7.2 Perché GPRD Non Aiuta?

Market Efficiency

Il rischio geopolitico è informazione pubblica. I mercati lo prezzano **istantaneamente** (in minuti, non giorni). L'effetto Safe Haven avviene su frequenze intraday, invisibile su dati daily.

Measurement Issues

GPRD è basato sull'analisi di articoli e notizie:

- Proxy imperfetto del vero sentiment di mercato

Horizonte Temporale

- Safe Haven su 1 giorno: Troppo rumoroso
- Safe Haven su 5 giorni: Già dissipato (mean reversion)
- Safe Haven potrebbe esistere su orizzonti mensili/trimestrali

5 Conclusioni

Il progetto **GoldenHour** ha analizzato 10.573 osservazioni giornaliere dell'oro (1985-2025) per validare l'ipotesi "Safe Haven" e confrontare approcci statistici classici con tecniche di deep learning per previsioni a 5 giorni. I risultati forniscono un benchmark efficace per valutare le differenti strategie di forecasting, valutando l'utilità del rischio geopolitico come indice.

5.1 Risultati Principali

5.1.1 Rifiuto dell'Ipotesi Safe Haven

Sei test statistici indipendenti hanno portato al **rifiuto** dell'ipotesi Safe Haven per previsioni a breve termine (1-5 giorni):

- **Correlazione Pearson:** $r = 0.0015$, $p = 0.886$ (praticamente nulla)
- **Test Chi-quadrato:** $\chi^2 = 1.347$, $p = 0.510$ (direzioni indipendenti)
- **Granger Causality:** $p > 0.55$ per tutti i lag (GPRD non causa rendimenti)
- **Regressione OLS:** $R^2 < 0.001$ (nessuna capacità esplicativa)
- **Analisi quantili estremi:** Differenza 0.049%, $p = 0.743$

Implicazione: Quando il rischio geopolitico aumenta, la probabilità che l'oro salga è 50.23% — indistinguibile dal caso. La gestione del rischio geopolitico da parte dei mercati rende l'effetto invisibile su dati giornalieri.

5.1.2 Vittoria della Parsimonia: ARIMA e Deep Learning

Il confronto tra sei modelli ha prodotto i risultati seguenti:

1. **ARIMA(0,0,1):** \$32.62 RMSE (2 parametri) — **Vincitore**
2. LSTM Univariato: \$32.69 RMSE (+0.21%, ~10K parametri)
3. ARIMA-GARCH(1,1): \$32.71 RMSE (+0.28%)
4. LSTM + GPRD: \$32.78 RMSE (+0.49%)

- 5. SARIMAX + GPRD: \$34.46 RMSE (+5.64% degradazione)
- 6. LSTM + Technical: \$38.00 RMSE (+16.5% degradazione)

Il test di Diebold-Mariano ($p = 0.857$) conferma equivalenza statistica tra ARIMA e LSTM, rendendo preferibile il modello più semplice per il principio di Occam's Razor.

5.1.3 Fallimento Predittivo del GPRD

L'inclusione del Geopolitical Risk Index ha:

- **Peggiorato** SARIMAX del +5.64% (contesto lineare)
- Avuto effetto **neutro** su LSTM (+0.28%, contesto non lineare)

Sono state ipotizzate le seguenti spiegazioni: market efficiency (prezzamento istantaneo), measurement issues (GPRD = proxy imperfetto), non-linearità complesse, asset-specific effects (Safe Haven tiene conto anche di altri beni rifugio).

5.2 Implicazioni e Raccomandazioni

5.2.1 Per Practitioner

- **Modello consigliato:** ARIMA(0,0,1) per forecast puntuali (RMSE \$32.62)
- **Risk Management:** ARIMA-GARCH per intervalli di confidenza dinamici e VaR
- **GPRD:** non fare affidamento su GPRD per market timing a breve termine

Best practices:

1. Preferire modelli semplici e interpretabili a sistemi black-box complessi
2. Usare walk-forward validation (mai cross-validation standard su time series)
3. Confrontare con naive benchmark e applicare test formali (Diebold-Mariano)
4. Testare ablation studies: più feature \neq migliori performance

5.2.2 Per Ricerca Futura

Direzioni promettenti:

- **Orizzonti più lunghi:** utilizzando modelli più complessi ed eventualmente valutare la reintroduzione del GPRD.
- **Dati intraday:** minute/hourly per catturare reazioni immediate
- **Regime-Switching:** Markov-Switching ARIMA per regimi calma vs. crisi
- **Transformer:** meccanismo di attenzione per catturare dinamiche event-driven

5.3 Limiti e Contributi

5.3.1 Limiti dello Studio

- Frequenza daily (intraday potrebbe rivelare effetti nascosti)
- Asset singolo (argento, platino, palladio potrebbero comportarsi diversamente)
- Architetture LSTM vanilla (bidirectional/attention potrebbero migliorare)
- Feature set limitato (GPRD solo; omessi VIX, EPU, tassi)
- Singola finestra test 2017-2025 (

5.3.2 Contributi Originali

1. **Validazione rigorosa Safe Haven:** 6 test indipendenti su 40 anni di dati
2. **Confronto sistematico ARIMA vs LSTM vs GARCH:** stessa metrica (RMSE su prezzi ricostruiti), walk-forward validation rigorosa
3. **Feature Importance GPRD:** dimostrazione empirica del danno in contesto deep learning
4. **Metodologia replicabile:** codice open source, dataset pubblico, seeds deterministici
5. **Test equivalenza statistica:** Diebold-Mariano per confronto formale

5.4 Riflessione Finale

Il progetto dimostra due verità fondamentali della finanza quantitativa:

1. **L'efficienza dei mercati è reale:** pattern sistematici basati su informazione pubblica sono difficilmente sfruttabili per trading a breve termine
2. **Semplicità vs. Complessità:** in presenza di alto rumore e basso segnale, modelli parsimoniosi generalizzano meglio di architetture complesse prone a overfitting

L'oro non segue schemi prevedibili basati su GPRD nel breve termine (1-5 giorni), ma ciò **non esclude** il suo ruolo di "assicurazione" in strategie di diversificazione a lungo termine.

L'investitore razionale dovrebbe considerare l'oro:

- **NON** come strumento di trading tattico basato su GPRD giornaliero
- **SÌ** come componente strategica di portafoglio bilanciato
- Apprezzandone: bassa correlazione con equity/bond, resilienza in tail risk, copertura inflazione

Nei mercati finanziari, la semplicità ben calibrata batte spesso la complessità mal generalizzata. L'approccio scientifico rigoroso — unendo statistica classica, machine learning, deep learning e validazione temporale corretta — è l'unico modo per separare autenticamente segnale da rumore.

"In Gold We Trust... But Not in Geopolitical Risk for Short-Term Trading"

L'oro mantiene il suo fascino millenario come riserva di valore, ma la strada verso profitti sistematici attraverso previsioni basate su rischio geopolitico si rivela, almeno nel breve termine, un miraggio statistico.