



UNIVERSITÀ DEGLI STUDI DI PALERMO

Corso di Laurea Magistrale – Statistica e Data Science (Classe LM-82).
Dipartimento di Scienze Economiche, Aziendali e Statistiche

“Variazioni della Volatilità Storica del
Sottostante e il loro Impatto sui Modelli
Parametrici e Non-Parametrici per il Pricing
delle Opzioni: Un’Analisi Comparativa”

TESI DI LAUREA DI
GIOVANNI MANLIO GAROFALO

RELATORE
GIADA ADELFI

ANNO ACCADEMICO 2023 - 2024

Sommario

In questa tesi viene effettuato un confronto tra diverse metodologie per la determinazione dei prezzi delle opzioni finanziarie, focalizzandosi su tre approcci principali: il modello di Black-Scholes, il metodo delle simulazioni Monte Carlo e l'algoritmo di Machine Learning XGBoost. L'analisi empirica si basa su dati reali di mercato, con l'obiettivo di valutare la performance predittiva di ciascun modello rispetto ai prezzi di scambio e teorici delle opzioni. In particolare, viene esplorato l'impatto della scelta della finestra temporale per il calcolo della volatilità storica sui risultati dei modelli. I risultati dell'analisi conducono all'identificazione delle combinazioni ottimali tra modello di pricing e finestra temporale per il calcolo della volatilità, in relazione alle due diverse tipologie di opzioni e di prezzi. In questo modo, si forniscono indicazioni pratiche su come selezionare il metodo più adatto per la valutazione delle opzioni, contribuendo a migliorare l'affidabilità delle previsioni e l'efficienza della gestione del rischio finanziario.

Indice

1	Introduzione	4
2	Le opzioni	6
2.1	Posizioni su opzioni	7
2.2	Tipologie di Opzioni	8
2.2.1	Classificazione delle Opzioni	8
2.3	Valore Intrinseco e Valore Temporale	8
2.4	Payoff delle Opzioni	9
2.5	Fattori che Influenzano i Prezzi delle Opzioni	12
2.6	Misure di Sensibilità del Contratto di Opzione	14
2.6.1	Delta (Δ)	14
2.6.2	Gamma (Γ)	14
2.6.3	Vega (ν)	15
2.6.4	Theta (Θ)	15
2.6.5	Rho (ρ)	15
3	Modello di Black-Scholes e la teoria della valutazione delle opzioni	16
3.1	Moto Browniano	16
3.1.1	Moto Browniano come Processo Markoviano	17
3.1.2	Moto Browniano come Martingala	17
3.1.3	Incrementi Stazionari e Indipendenti	18
3.2	Moto Browniano Geometrico ed Equazioni Differenziali Sto- castiche	19
3.3	Calcolo di Itô e Lemma di Itô	19
3.4	Derivazione dell'equazione di Black-Scholes	20
3.4.1	Soluzione dell'equazione di Black-Scholes	21
3.4.2	Ulteriori chiarimenti	21
3.4.3	Ipotesi del modello	22
3.5	Valutazione neutrale del rischio	23

4	Metodo Monte Carlo per la valutazione delle opzioni	25
4.1	Introduzione	25
4.2	I passi della procedura	26
4.2.1	Simulazione dell'evoluzione del prezzo del sottostante	26
4.3	Dipendenza Markoviana e Suddivisione degli Intervalli	27
4.3.1	Costruzione dei Sentieri Temporali	27
4.3.2	Calcolo del Payoff	28
4.3.3	Misura Neutrale al Rischio	28
4.3.4	Attualizzazione del Payoff	29
4.4	Ottimizzazione del Numero di Simulazioni	29
4.5	Generazione di Numeri Casuali e Quasi-Casuali	30
4.6	Tecniche di Riduzione della Varianza	30
5	Extra Gradient Boost (XGBoost)	32
5.1	L'Ensemble Learning	32
5.1.1	Bagging	32
5.1.2	Boosting	33
5.2	Gradient Boosting	33
5.3	Inizializzazione del Modello con la Funzione di Perdita	33
5.3.1	Discesa del Gradiente in XGBoost	34
5.4	Iperparametri di XGBoost	35
5.5	Cross-Validation: Il Metodo dei 5 Folds	35
6	Dati e Analisi Esplorativa	37
6.1	Dati Originali	37
6.2	Calcolo della Volatilità Storica degli Asset Sottostanti	39
6.2.1	Definizione delle finestre Temporali	40
6.2.2	Selezione dei Dati	40
6.2.3	Calcolo della Volatilità Annuale	41
6.2.4	Applicazione della Funzione	41
6.3	Analisi Outliers con Local Outlier Factor	41
6.4	Analisi Esplorativa	42
6.4.1	Sintesi Analisi Esplorativa	54
7	Analisi Comparativa Modelli di Pricing	56
7.1	Modello Black-Scholes	56
7.2	Metodo Monte Carlo	60
7.3	XGBoost	65

8 Conclusioni	73
8.1 Introduzione ai risultati	73
8.2 Modello di Black-Scholes	74
8.3 Metodo Monte Carlo	75
8.4 Modello di Machine Learning XGBoost	75
8.5 Sintesi	76
8.6 Estensione dell'Analisi a un Contesto Out of Sample	77
.1 Listings Codice R: Importazione e processamento dati	80
.2 Listings Codice R: Analisi Outliers - LOF	81
.3 Listings Codice R: Modello BLACK SCHOLES	83
.4 Listings Codice R: Metodo Montecarlo	84
.5 Listings Codice R: XGBoost	86

Capitolo 1

Introduzione

La finanza rappresenta un settore in costante evoluzione, caratterizzato da una crescita significativa e da una complessità crescente. I mercati finanziari, in particolare, hanno visto un’espansione dei prodotti derivati, strumenti finanziari che offrono agli investitori nuove opportunità, ma anche nuove sfide in termini di gestione del rischio. Tra questi, le opzioni sono tra i derivati più utilizzati, apprezzate per la loro flessibilità. Tuttavia, la valutazione precisa delle opzioni è cruciale, poiché errori in questa fase possono portare a decisioni di investimento non ottimali e a rischi mal gestiti. La determinazione del prezzo delle opzioni è oggetto di ampio studio in letteratura, con due principali approcci: i modelli parametrici e i modelli non parametrici. Tra i modelli parametrici, il modello di Black-Scholes, introdotto nel 1973, è stato rivoluzionario per la sua capacità di fornire una formula di valutazione delle opzioni in forma chiusa, permettendo agli investitori di determinare rapidamente il prezzo teorico delle opzioni. Questo modello, tuttavia, si basa su assunzioni semplificative, come quella della volatilità costante, che non sempre riflettono accuratamente le dinamiche di mercato osservate. Ad esempio, per le opzioni “at the money”, si osserva spesso un fenomeno noto come “smile volatility”, dove la volatilità implicita mostra un andamento parabolico, contrastando l’assunzione di volatilità costante del modello di Black-Scholes.

Per affrontare queste limitazioni, sono stati sviluppati modelli evolutivi, tra cui il modello di Heston[5], che introduce una volatilità stocastica, fornendo una rappresentazione più realistica delle dinamiche di mercato.

Accanto ai modelli parametrici, la crescente complessità dei mercati e la disponibilità di grandi quantità di dati hanno favorito l’emergere di approcci numerici e tecniche di machine learning. Questi approcci, grazie alla loro capacità di adattarsi a pattern complessi senza richiedere ipotesi rigide sui dati, stanno guadagnando terreno come strumenti efficaci per la valutazione delle opzioni. In particolare, il modello XGBoost, una potente tecnica

di gradient boosting, ha dimostrato eccellenti capacità previsive in numerosi contesti finanziari[7] [11] [12] [14]. Nell’ambito di questo studio, verrà impiegato il modello XGBoost, combinato con il metodo *5-folds cross validation*, al fine di valutare la performance predittiva nella determinazione dei prezzi delle opzioni.

Nel corso di questo elaborato, il Capitolo 2 offrirà una panoramica sulle opzioni finanziarie, fornendo le basi teoriche necessarie. Nel Capitolo 3, verrà presentata una derivazione formale del modello di Black-Scholes, con un focus sulle assunzioni di base e sul concetto di neutralità del rischio. Il Capitolo 4 esplorerà il metodo delle simulazioni Monte Carlo come approccio numerico computazionale per la determinazione dei prezzi delle opzioni. Il Capitolo 5 introdurrà il modello di machine learning XGBoost e la 5-fold cross validation quale metodo per la selezione della combinazione degli iper-parametri del modello utilizzato nell’analisi.

Analisi che verrà condotta e discussa nei capitoli 6 e 7. Infine, il Capitolo 8 presenterà i risultati ottenuti e le conclusioni tratte da questo studio.

Capitolo 2

Le opzioni

In finanza, un'opzione è un contratto che attribuisce all'acquirente il diritto, ma non l'obbligo, di comprare o vendere un titolo noto come strumento sottostante entro una data di scadenza. L'acquirente acquisisce tale diritto corrispondendo il pagamento di un premio. Poiché si tratta di uno strumento costruito su un asset preesistente, le opzioni rientrano tra gli strumenti “derivati”[10]. Ci sono due tipi principali di opzioni:

- **Opzioni call (call options):** danno al titolare il diritto di acquistare un'attività entro una data ed a un prezzo stabilito a priori.
- **Opzioni put (put options):** danno al titolare il diritto di vendere un'attività entro una certa data per un certo prezzo.

Il prezzo specificato nel contratto è noto come **prezzo d'esercizio (exercise price)** o **prezzo base (strike price)**; la data stabilita a priori nel contratto è nota come **data di estinzione (expiration date)** o **scadenza (maturity)**.

In genere, ogni contratto d'opzione riguarda il diritto di comprare o vendere cento azioni. Tuttavia, il titolo sottostante può consistere in altri asset: obbligazioni, titoli su materie prime, valute, indici, ecc.

Le opzioni conferiscono al detentore il diritto ad eseguire un'azione[9]. Il detentore non è vincolato a esercitarlo in ogni caso. Questo è ciò che distingue le opzioni dai contratti a termine e dai futures, in cui ci si impegna a comprare o vendere l'attività sottostante. Inoltre, diversamente dall'acquisto di un contratto a termine o di un contratto future, per acquistare un'opzione si sostiene il pagamento di un premio.

2.1 Posizioni su opzioni

In ogni contratto di opzione, ci sono due parti coinvolte. Da un lato, c'è l'investitore che ha assunto la posizione lunga (cioè, chi ha acquistato l'opzione). Dall'altro lato, c'è l'investitore che ha assunto la posizione corta (cioè, chi ha venduto l'opzione). Chi vende l'opzione riceve un introito iniziale ma è esposto a una potenziale perdita. Il suo profitto (o la sua perdita) è pari alla perdita (o al profitto) di chi ha acquistato l'opzione.

Esistono, dunque, quattro tipi di posizioni su opzioni:

1. una posizione lunga su una call;
2. una posizione lunga su una put;
3. una posizione corta su una call;
4. una posizione corta su una put.

Le opzioni possono essere utilizzate per scopi di copertura. Questi contratti, infatti, rappresentano una sorta di assicurazione. Consentono agli investitori di proteggersi dai movimenti sfavorevoli dei prezzi senza privarli della possibilità di beneficiare dei movimenti favorevoli.

Le opzioni sono utilizzate anche per scopi speculativi. Esse offrono una sorta di leva finanziaria. In tal senso, l'utilizzo delle opzioni amplifica le possibili conseguenze finanziarie di un certo investimento. I risultati buoni diventano ottimi, mentre i risultati cattivi diventano pessimi.

In ogni momento, per ogni titolo, si possono negoziare diverse opzioni. Tutte le opzioni dello stesso genere (call o put) costituiscono una “classe di opzioni” (option class).

Tutte le opzioni di una determinata classe, con la stessa scadenza e lo stesso prezzo d'esercizio concorrono a definire una “serie di opzioni” (option series). Ad esempio, una serie di opzioni potrebbe includere tutte le opzioni call su Apple con una data di scadenza fissata al 20 gennaio 2025 e un prezzo di esercizio di \$150.

2.2 Tipologie di Opzioni

Le opzioni possono essere di tipo europeo o americano:

- **Opzioni europee (European options):** possono essere esercitate solo alla scadenza.
- **Opzioni americane (American options):** possono essere esercitate in qualsiasi momento durante la loro durata fino alla scadenza.

Chi sottoscrive una call prevede che il valore dell'azione sottostante salga. Se l'opzione è di tipo europeo, l'investitore ha la possibilità di esercitarla solo alla scadenza. Se a quella data il prezzo dell'azione è inferiore al prezzo d'esercizio, l'opzione non sarà esercitata. In questo caso l'investitore perde l'intero investimento iniziale corrisposto sotto forma di premio.

Chi sottoscrive una put prevede che il prezzo dell'azione sottostante diminuisca. Se in tale data il prezzo dell'azione sottostante è minore dello strike price, l'opzione verrà esercitata realizzando così un profitto (al netto dei costi di transazione).

2.2.1 Classificazione delle Opzioni

Sulla base della profittabilità, le opzioni possono essere classificate come segue:

- **In the money:** l'opzione che genererebbe un flusso di cassa positivo se fosse esercitata;
- **At the money:** l'opzione che genererebbe un flusso di cassa nullo;
- **Out of the money:** l'opzione che genererebbe un flusso di cassa negativo.

2.3 Valore Intrinseco e Valore Temporale

Il “valore intrinseco” (intrinsic value) di un'opzione è definito come il massimo tra 0 e il valore che l'opzione avrebbe se fosse esercitata immediatamente. Pertanto, il valore intrinseco di una call è dato dalla seguente formula:

$$\text{Valore intrinseco di una call} = \max(S - K, 0) \quad (2.1)$$

Il valore intrinseco di una put, invece, è dato da:

$$\text{Valore intrinseco di una put} = \max(K - S, 0) \quad (2.2)$$

2.4 Payoff delle Opzioni

Il payoff per un'opzione call è definito come la differenza positiva tra il prezzo dell'azione alla scadenza (S_T) e il prezzo di esercizio (K). Allo stesso modo, per un'opzione put è definito come la differenza tra il prezzo di esercizio (K) e il prezzo dell'azione alla scadenza (S_T), se questa differenza è positiva; altrimenti, è zero.

- Per una posizione lunga su una call europea:

$$\text{Payoff} = \max(S_T - K, 0) \quad (2.3)$$

- Per una posizione lunga su una put europea:

$$\text{Payoff} = \max(K - S_T, 0) \quad (2.4)$$

- Il payoff di una posizione corta su una call europea è pari a:

$$\text{Payoff} = -\max(S_T - K, 0) = \min(K - S_T, 0) \quad (2.5)$$

- Il payoff di una posizione corta su una put europea è pari a:

$$\text{Payoff} = -\max(K - S_T, 0) = \min(S_T - K, 0) \quad (2.6)$$

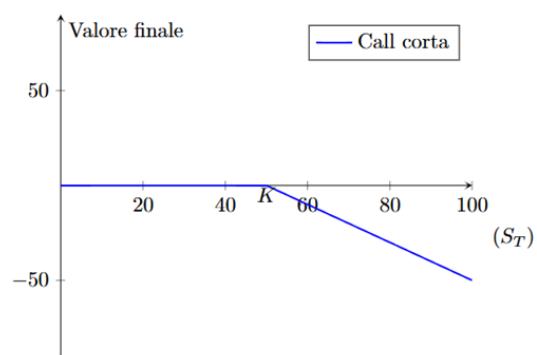
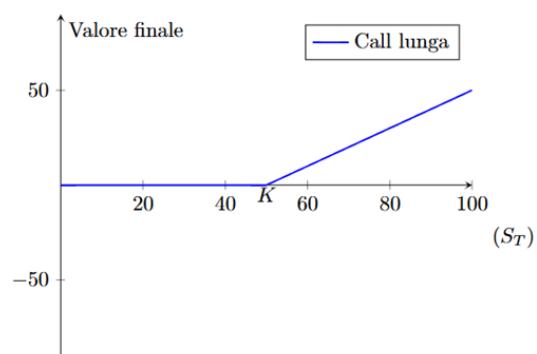


Figura 2.1: Payoff su posizioni call

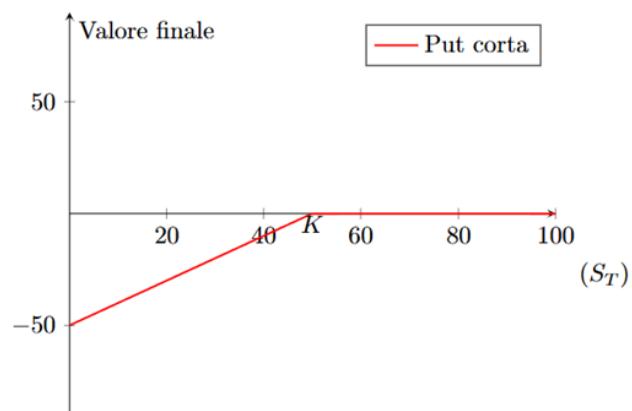
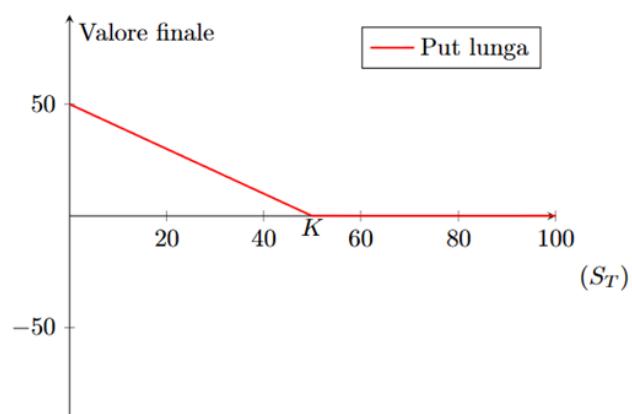


Figura 2.2: Payoff su posizioni put

2.5 Fattori che Influenzano i Prezzi delle Opzioni

Come già discusso, un'opzione è un prodotto derivato il cui valore dipende dalla performance attuale ed attesa del suo sottostante: sia esso un'azione, un indice o altro tipo di asset finanziario. I fattori più influenti sul prezzo di un'opzione scritta su un'azione sono:

- Il prezzo corrente dell'azione (S_t);
- Il prezzo d'esercizio (K);
- La vita residua (T);
- La volatilità storica del prezzo dell'azione (σ);
- Il tasso d'interesse privo di rischio (r);
- I dividendi attesi durante la vita dell'opzione il cui valore attuale è (D_0).

In tabella 1 si riporta una sintesi di come questi agiscono sul prezzo finale di un'opzione.

Tabella 2.1: Fattori che influenzano i prezzi delle opzioni.

Variabile	Call europea	Put europea	Call americana	Put americana
Prezzo dell'azione	+	-	+	-
Prezzo d'esercizio	-	+	-	+
Vita residua	?	?	+	+
Volatilità	+	+	+	+
Tasso d'interesse	+	-	+	-
Dividendi	-	+	-	+

Nota: I simboli "+", "-", "?" indicano che l'effetto sul prezzo dell'opzione di un aumento della variabile è, rispettivamente, positivo, negativo o incerto. E = europea; A = americana

Il valore finale di una call esercitata corrisponde alla differenza tra il prezzo dell'azione e il prezzo d'esercizio. Di conseguenza, le call aumentano di valore con l'aumento del **prezzo dell'azione** e diminuiscono di valore con l'aumento del **prezzo d'esercizio**. Al contrario, il valore finale di una put è dato dalla differenza tra il prezzo d'esercizio e il prezzo dell'azione. Pertanto,

le put si comportano in modo opposto alle call: il loro valore diminuisce con l'aumento del prezzo dell'azione e aumenta con l'aumento del prezzo d'esercizio.

Sia le call , sia le put americane aumentano di valore con l'aumento della **vita residua**. Per comprendere il motivo, consideriamo due opzioni che differiscono solo per la data di scadenza. Il possessore dell'opzione con la vita residua maggiore ha tutte le opportunità di esercizio del possessore dell'opzione con la vita residua minore, e altre ancora. Pertanto, l'opzione a lungo termine deve valere almeno quanto l'opzione a breve termine. Generalmente, anche le call e le put europee aumentano di valore con l'aumento della vita residua, ma non è una relazione certa.

La volatilità rappresenta un indicatore cruciale dell'incertezza associata al futuro andamento del prezzo di un titolo. Un incremento della volatilità segnala una maggiore probabilità che il titolo possa subire variazioni di prezzo significative, sia in positivo che in negativo. Questo comportamento riflette la natura del mercato finanziario, dove eventi imprevisti o cambiamenti nelle condizioni economiche possono amplificare le fluttuazioni dei prezzi. Per coloro che detengono opzioni call, l'aumento della volatilità è particolarmente vantaggioso poiché incrementa le possibilità di un significativo rialzo del prezzo del titolo. In questo scenario, il potenziale di guadagno è teoricamente illimitato, mentre il rischio di ribasso è confinato al solo premio pagato per l'opzione. Analogamente, i possessori di opzioni put beneficiano di un aumento della volatilità perché cresce la probabilità di un calo pronunciato del prezzo del titolo. Anche in questo caso, il rischio è limitato al premio pagato, poiché l'opzione put permette di vendere il titolo a un prezzo prefissato, proteggendo così l'investimento. Di conseguenza, l'aumento della volatilità incrementa il valore sia delle call che delle put, poiché entrambe le tipologie di opzioni diventano più preziose con l'aumentare dell'incertezza e delle potenziali oscillazioni dei prezzi del titolo sottostante. Emergono, dunque, sia la funzione di copertura che quella speculativa.

Il tasso d'interesse privo di rischio ha un impatto complesso sul prezzo delle opzioni. Quando i tassi d'interesse aumentano, il tasso di crescita atteso del prezzo delle azioni tende a crescere. Ma, per i detentori di opzioni, il valore attuale dei futuri flussi di cassa diminuisce. Questi effetti combinati fanno sì che il valore di una put diminuisca, per cui il prezzo di una put scende con l'aumento del tasso d'interesse privo di rischio. Nel caso delle call, l'aumento del tasso d'interesse tende a incrementare il prezzo dell'opzione, mentre la riduzione del valore attuale dei futuri flussi di cassa tende a deprimere. È dimostrabile che il primo effetto prevale sul secondo, quindi il prezzo di una call aumenta con l'aumento del tasso d'interesse privo di rischio.

Queste conclusioni si basano sull'ipotesi che tutte le altre variabili rimangano

costanti.

Quando una società distribuisce **dividendi**, il prezzo delle sue azioni tende a diminuire nel giorno di stacco del dividendo. Questo fenomeno si verifica perché il valore del dividendo viene sottratto dal prezzo dell'azione. Per i detentori di opzioni call, ciò rappresenta una notizia sfavorevole. La diminuzione del prezzo dell'azione riduce il valore della call, poiché si riduce la differenza tra il prezzo di mercato dell'azione e il prezzo di esercizio della call. Al contrario, per i detentori di opzioni put, questa è una notizia favorevole. La riduzione del prezzo dell'azione aumenta il valore della put, poiché si amplia la differenza tra il prezzo di esercizio della put e il prezzo di mercato dell'azione. Di conseguenza, esiste una relazione negativa tra il valore di una call e l'importo dei dividendi attesi: all'aumentare dei dividendi attesi, il valore della call diminuisce. D'altro canto, esiste una relazione positiva tra il valore di una put e l'importo dei dividendi attesi: all'aumentare dei dividendi attesi, il valore della put aumenta.

2.6 Misure di Sensibilità del Contratto di Opzione

Le misure di sensibilità del contratto di opzione, comunemente note come Greche, sono parametri utilizzati per valutare come il prezzo di un'opzione cambia in risposta alle fluttuazioni dei prezzi, alle fluttuazioni della volatilità e al trascorrere del tempo.

2.6.1 Delta (Δ)

Delta è la prima delle lettere greche utilizzate nel mercato delle opzioni e rappresenta il tasso di variazione tra il prezzo dell'opzione e una variazione di \$1 nel prezzo dell'attività sottostante. Questo descrive la sensibilità al prezzo dell'opzione rispetto al suo asset sottostante. Ad esempio, un Delta di 0,5 significa che se il prezzo dell'azione sottostante aumenta di \$1, il prezzo dell'opzione aumenterà di \$0,50.

2.6.2 Gamma (Γ)

Gamma misura la variazione attesa del delta per ogni variazione di \$1 nel prezzo dell'attività sottostante. Questo fattore misura la sensibilità del Delta rispetto ai cambiamenti nel prezzo dell'attività sottostante. Un Gamma elevato significa che il Delta cambia rapidamente quando il prezzo del sotto-

stante varia, il che può comportare una maggiore variabilità nei prezzi delle opzioni.

2.6.3 Vega (ν)

Il Vega misura la sensibilità del prezzo dell'opzione rispetto alle variazioni della volatilità implicita dell'attività sottostante.

2.6.4 Theta (Θ)

Il Theta misura la sensibilità del prezzo dell'opzione rispetto al passare del tempo, noto anche come "decadimento temporale". Un Theta negativo significa che il prezzo dell'opzione tende a diminuire con il passare del tempo, riflettendo la perdita di valore temporale delle opzioni. Ad esempio, un Theta di -0,05 significa che il prezzo dell'opzione scenderà di 0,05 unità per ogni giorno che passa.

2.6.5 Rho (ρ)

Il Rho misura la sensibilità del prezzo dell'opzione rispetto ai cambiamenti nei tassi di interesse. Un Rho positivo significa che un aumento dei tassi di interesse tende ad aumentare il prezzo dell'opzione call e a diminuire il prezzo dell'opzione put. Ad esempio, un Rho di 0,1 per una call significa che se i tassi di interesse aumentano di 1 punto percentuale, il prezzo della call aumenterà di \$0,10.

Capitolo 3

Modello di Black-Scholes e la teoria della valutazione delle opzioni

La teoria della valutazione delle opzioni stima il valore di un contratto di opzione assegnando un prezzo, noto come premio, basato sulla probabilità calcolata che il contratto terminerà *in the money* alla scadenza. Più sono alte le probabilità che l'opzione terminerà ITM, più questa sarà redditizia. Di conseguenza, maggiore sarà il valore dell'opzione, e viceversa.

Inoltre, i modelli di pricing sono funzionali per derivare anche i fattori di rischio precedentemente esposti, che sono noti come “Greche” di un’opzione. Il modello di Black-Scholes è uno dei modelli di pricing più importanti per la valutazione delle opzioni. Esso si basa su assunzioni specifiche e su risultati significativi derivanti dall'applicazione dei processi stocastici nella finanza matematica. Tra questi, i concetti chiave comprendono il moto browniano (o processo di Wiener), le equazioni differenziali stocastiche, il calcolo di Itô e il moto browniano geometrico.

3.1 Moto Browniano

Nato da osservazioni fisiche, Il moto browniano è evoluto in un modello stocastico rigoroso che sottende gran parte della teoria finanziaria moderna. Partendo dalla sua definizione come limite di un Random Walk, discutiamo le sue principali proprietà estendendo il concetto al Moto Browniano Geometrico (GBM) e alle sue applicazioni in finanza.

Il moto browniano, noto anche come processo di Wiener, può essere compreso come il limite di un Random Walk simmetrico. Consideriamo uno scenario

in cui un individuo compie un passo alla volta, a sinistra o a destra con probabilità uguale. Dividiamo l'insieme dei passi $[0, \infty)$ in piccolissimi intervalli di lunghezza δ . Per ciascun intervallo, definiamo una variabile casuale X_i come segue:

$$X_i = \begin{cases} \sqrt{\delta} & \text{con probabilità } \frac{1}{2}, \\ -\sqrt{\delta} & \text{con probabilità } \frac{1}{2}. \end{cases} \quad (3.1)$$

Queste variabili casuali X_i sono indipendenti, e il processo $Z(t)$ è definito come la somma di queste:

$$Z(t) = \sum_{i=1}^n X_i, \quad \text{dove } t = n\delta. \quad (3.2)$$

Man mano che n tende all'infinito e δ tende a zero, $Z(t)$ approssima una distribuzione normale $Z(t) \sim N(0, t)$, definendo il processo stocastico noto come moto Browniano.

La media e la varianza del processo $Z(t)$ sono date da:

$$\mathbb{E}[Z(t)] = 0 \quad (3.3)$$

$$\text{Var}(Z(t)) = t \quad (3.4)$$

Questo implica che con il passare del tempo, la posizione attesa rimane a zero e la varianza aumenta con il tempo, riflettendo l'incertezza crescente man mano che il processo evolve.

3.1.1 Moto Browniano come Processo Markoviano

Il moto browniano è un processo Markoviano, il che significa che la sua evoluzione futura dipende solo dal suo stato presente, non dal percorso che ha seguito per arrivarci. Formalmente, per qualsiasi $s < t$:

$$\mathbb{P}(Z(t) \leq z | \mathcal{F}_s) = \mathbb{P}(Z(t) \leq z | Z(s)) \quad (3.5)$$

dove \mathcal{F}_s è il filtro fino al tempo s . Questa proprietà è cruciale per descrivere scenari in cui il futuro è indipendente dal passato, dato il presente.

3.1.2 Moto Browniano come Martingala

Oltre ad essere un processo Markoviano, il moto browniano è anche una martingala. Questo significa che il suo valore atteso in qualsiasi tempo futuro, dato il presente, è uguale al suo valore attuale. Formalmente:

$$\mathbb{E}[Z(t) | \mathcal{F}_s] = Z(s), \quad \text{per } s \leq t \quad (3.6)$$

Questa proprietà riflette la natura di "gioco equo" del moto browniano, rendendolo particolarmente utile nella definizione di scenari dove non esiste una tendenza prevedibile, come nella valutazione di determinati derivati finanziari.

3.1.3 Incrementi Stazionari e Indipendenti

Il moto browniano possiede due proprietà aggiuntive chiave: incrementi stazionari ed indipendenti.

Incrementi Indipendenti: gli incrementi del processo $Z(t)$ sono indipendenti. Per una qualsiasi sequenza di tempi $0 \leq t_1 < t_2 < \dots < t_n$, gli incrementi

$$Z(t_2) - Z(t_1), \quad Z(t_3) - Z(t_2), \quad \dots, \quad Z(t_n) - Z(t_{n-1})$$

sono variabili casuali indipendenti. Questo deriva dall'indipendenza dei singoli X_i che definiscono ciascun incremento.

Incrementi Stazionari: Un processo $X(t)$ ha incrementi stazionari se la distribuzione dell'incremento

$$X(t_2) - X(t_1)$$

dipende solo dalla differenza $t_2 - t_1$, e non dai tempi specifici t_1 e t_2 . Per $Z(t)$, calcoliamo:

$$Z(t_2) - Z(t_1) = \sum_{i=n_1+1}^{n_2} X_i, \tag{3.7}$$

dove $t_1 = n_1\delta$ e $t_2 = n_2\delta$. La media di questa differenza è:

$$\mathbb{E}[Z(t_2) - Z(t_1)] = \sum_{i=n_1+1}^{n_2} \mathbb{E}[X_i] = 0. \tag{3.8}$$

La varianza della differenza è:

$$\text{Var}(Z(t_2) - Z(t_1)) = \sum_{i=n_1+1}^{n_2} \text{Var}(X_i) = (n_2 - n_1)\delta = t_2 - t_1. \tag{3.9}$$

Pertanto, l'incremento $Z(t_2) - Z(t_1)$ è distribuito normalmente $N(0, t_2 - t_1)$, dimostrando che la distribuzione dipende solo dalla lunghezza dell'intervallo e non dai valori specifici di t_1 e t_2 .

Come conseguenza diretta, la covarianza tra $Z(s)$ e $Z(t)$ è data da:

$$\text{Cov}(Z(s), Z(t)) = \min(s, t) \tag{3.10}$$

3.2 Moto Browniano Geometrico ed Equazioni Differenziali Stocastiche

Il Moto Browniano Geometrico (GBM) rappresenta un'estensione fondamentale del moto browniano per la descrizione dei prezzi degli asset. Mentre il moto browniano, o processo di Wiener, è adatto per descrivere l'evoluzione di un processo sotcastico in un contesto generale, i prezzi degli strumenti finanziari richiedono un modello che garantisca valori non negativi e possa descrivere tendenze esponenziali, tipiche delle dinamiche di capitalizzazione. Questo porta alla formulazione del GBM, in cui il logaritmo del prezzo dell'asset segue un moto browniano.

Il prezzo S_t di un'azione è modellato tramite una equazione differenziale stocastica (SDE) del tipo:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (3.11)$$

dove:

- $\mu S_t dt$ rappresenta la componente deterministica e descrive il rendimento medio atteso.
- σ è la volatilità, che misura la variabilità o l'incertezza associata al prezzo dell'azione.
- $S_t dW_t$ è un moto browniano standard, che descrive l'elemento di incertezza, o componente stocastica, nel prezzo.

Combinando le proprietà del GBM con il calcolo stocastico (specificamente, il Lemma di Ito), si deriva l'equazione di Black-Scholes, fornendo una soluzione in forma chiusa per il prezzo delle opzioni call e put europee. Il prezzo dipende dal prezzo corrente dell'attivo, dal prezzo di esercizio, dal tempo alla scadenza, dal tasso privo di rischio e dalla volatilità dell'attivo.

3.3 Calcolo di Itô e Lemma di Itô

Il calcolo di Itô è una tecnica fondamentale per trattare le SDE. Il lemma di Itô permette di determinare come una funzione del processo stocastico evolve nel tempo. Sia $f(S_t, t)$ una funzione del processo S_t e del tempo t . Il lemma di Itô afferma che:

$$df(S_t, t) = \left(\frac{\partial f}{\partial t} + \mu S_t \frac{\partial f}{\partial S_t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 f}{\partial S_t^2} \right) dt + \sigma S_t \frac{\partial f}{\partial S_t} dW_t \quad (3.12)$$

3.4 Derivazione dell'equazione di Black-Scholes

Per un'opzione europea di tipo call o put, il valore dell'opzione $C(S_t, t)$ dipende dal prezzo dell'azione S_t e dal tempo t . Utilizziamo il lemma di Itô sulla funzione $C(S_t, t)$:

$$dC = \left(\frac{\partial C}{\partial t} + \mu S_t \frac{\partial C}{\partial S_t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} \right) dt + \sigma S_t \frac{\partial C}{\partial S_t} dW_t \quad (3.13)$$

Costruiamo un portafoglio privo di rischio che consiste in una posizione lunga in un'opzione call e una posizione corta in $\frac{\partial C}{\partial S_t}$ azioni. La possibilità di creare un portafoglio privo di rischio deriva dal fatto che il prezzo delle azioni e delle opzioni è influenzato dalla stessa fonte di incertezza: le variazioni del prezzo dell'azione sottostante. In ogni breve intervallo di tempo, il prezzo di una call è perfettamente correlato in modo positivo con il prezzo dell'azione sottostante, mentre il prezzo di una put è perfettamente correlato in modo negativo. Pertanto, quando si crea un portafoglio adeguato di azioni e opzioni, le perdite su una posizione sono compensate dai profitti sull'altra, garantendo che il valore complessivo del portafoglio alla fine di ogni breve intervallo di tempo sia sempre noto con certezza. Il valore di questo portafoglio Π è:

$$\Pi = C - \frac{\partial C}{\partial S_t} S_t \quad (3.14)$$

La variazione del valore del portafoglio è:

$$d\Pi = dC - \frac{\partial C}{\partial S_t} dS_t \quad (3.15)$$

Sostituendo dC e dS_t nella (3.15), otteniamo:

$$d\Pi = \left(\frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} \right) dt \quad (3.16)$$

Poiché il portafoglio è privo di rischio, deve crescere al tasso di interesse privo di rischio r :

$$d\Pi = r\Pi dt \quad (3.17)$$

Equivalentemente:

$$\left(\frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} \right) dt = r \left(C - \frac{\partial C}{\partial S_t} S_t \right) dt \quad (3.18)$$

$$\frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} + r S_t \frac{\partial C}{\partial S_t} - r C = 0 \quad (3.19)$$

Quest'ultima è l'equazione di Black-Scholes [1].

3.4.1 Soluzione dell'equazione di Black-Scholes

Per un'opzione call europea, la soluzione dell'equazione di Black-Scholes è:

$$C(S_t, t) = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \quad (3.20)$$

Per un'opzione put europea, la soluzione è:

$$P(S_t, t) = K e^{-r(T-t)} N(-d_2) - S_t N(-d_1) \quad (3.21)$$

dove:

$$d_1 = \frac{\ln(S_t/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad (3.22)$$

$$d_2 = d_1 - \sigma\sqrt{T-t} \quad (3.23)$$

e $N(\cdot)$ è la funzione cumulativa della distribuzione normale standardizzata.

Spiegazione delle variabili

Le variabili presenti nelle equazioni dei prezzi delle opzioni call e put sono descritte come segue:

- S_t : Il prezzo corrente dell'asset sottostante al tempo t .
- K : Il prezzo di esercizio dell'opzione (strike price), cioè il prezzo al quale l'opzione può essere esercitata.
- r : Il tasso di interesse privo di rischio, che rappresenta il tasso di rendimento di un investimento privo di rischio.
- T : La scadenza dell'opzione, espressa in anni.
- t : Il tempo corrente, espresso in anni.
- σ : La volatilità storica del prezzo del sottostante che misura della variazione del prezzo nel tempo.

3.4.2 Ulteriori chiarimenti

- $N(d_2)$: Rappresenta la probabilità, in un mondo neutrale verso il rischio, che l'opzione call venga esercitata.

- $N(d_1)$: Non è altrettanto semplice da interpretare. Il termine $S_t N(d_1) e^{-rT}$ è il valore atteso, in un mondo neutrale verso il rischio, di una variabile che è pari a S_T se $S_T > K$ e che è pari a zero altrimenti. Lo strike K viene pagato solo se $S_T < K$ e la probabilità che ciò accada è pari a $N(d_2)$.

Pertanto, il valore atteso della call in un mondo neutrale verso il rischio è pari a:

$$S_t N(d_1) e^{-rT} - K N(d_2) e^{-rT}$$

Questo rappresenta la differenza tra il valore atteso del prezzo dell'azione al momento della scadenza (moltiplicato per la probabilità che l'opzione venga esercitata) e il valore atteso dello strike price K (anch'esso moltiplicato per la probabilità che l'opzione venga esercitata). Attualizzando il valore atteso in base al tasso d'interesse privo di rischio, si ottiene la formula di Black-Scholes per il valore corrente della call:

$$C = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

3.4.3 Ipotesi del modello

Le ipotesi alla base del modello di Black-Scholes includono:

1. La dinamica del prezzo dell'azione è descritta da un moto browniano geometrico con drift μ e volatilità σ costanti.
2. Le vendite allo scoperto sono consentite senza restrizioni sull'utilizzo dei relativi proventi.
3. Non esistono costi di transazione o tasse, e i titoli sono perfettamente divisibili.
4. L'azione non paga dividendi durante la vita del derivato.
5. Non esistono opportunità di arbitraggio prive di rischio.
6. I titoli vengono negoziati continuamente.
7. Il tasso d'interesse a breve termine privo di rischio, r , è uguale per tutte le scadenze.

Chiaramente, alcune di queste ipotesi non si verificano sempre o anche la maggior parte del tempo. Ad esempio, il modello assume anche che la volatilità rimanga costante durante la vita dell'opzione. Questo è irrealistico e normalmente inverosimile, perché la volatilità fluttua con il livello di offerta

e domanda.

Queste ipotesi possono essere mitigate in alcuni casi. Ad esempio, la volatilità σ e il tasso d'interesse r possono essere funzioni note del tempo t . Inoltre, si può ipotizzare che i tassi d'interesse siano stocastici, purché la distribuzione del prezzo dell'azione alla scadenza dell'opzione rimanga log-normale. Questi adattamenti consentono una maggiore flessibilità del modello rispetto alle condizioni reali di mercato, pur mantenendo la coerenza della valutazione dei derivati.

3.5 Valutazione neutrale del rischio

La "valutazione neutrale verso il rischio" (risk-neutral valuation) rappresenta un principio fondamentale nella valutazione dei derivati, basato sull'assunto che gli investitori siano neutrali verso il rischio, ovvero che non richiedano un tasso di rendimento atteso più elevato come compenso per il rischio assunto. In un ipotetico "mondo neutrale verso il rischio", i prezzi dei derivati sono uguali a quelli del mondo reale. Gli investitori sono indifferenti al rischio e valutano i titoli in modo tale che il loro valore atteso, scontato al tasso privo di rischio, è il prezzo corrente. Questa condizione teorica elimina la necessità di considerare il grado di avversione al rischio degli investitori, semplificando notevolmente il processo di valutazione dei derivati. Sebbene nel mondo reale gli investitori richiedano rendimenti maggiori per compensare i rischi, nel contesto neutrale verso il rischio questa complessità viene eliminata, mantenendo invariati i prezzi dei derivati. Anche in presenza di una maggiore avversione al rischio da parte degli investitori e di una conseguente diminuzione dei prezzi delle azioni, la formula che lega i prezzi dei derivati ai prezzi delle azioni sottostanti rimane costante. Questo approccio teorico rappresenta un'esemplificazione che facilita la determinazione dei prezzi dei derivati, prescindendo dalle preferenze individuali degli investitori nei confronti del rischio, e fornendo un quadro teorico chiaro e meno complesso per la valutazione di questi strumenti finanziari.

Quando, invece, non si adotta l'ipotesi di valutazione neutrale al rischio, diventa necessario incorporare le preferenze degli investitori nel modello. Questo richiede l'uso di funzioni di utilità o di modelli comportamentali per tener conto di come gli investitori valutano i rischi e i rendimenti. Gli investitori, infatti, non sono sempre disposti ad accettare un rischio elevato senza una compensazione in termini di rendimento atteso maggiore, il che introduce l'avversione al rischio nelle decisioni di investimento. Di conseguenza, senza l'ipotesi di neutralità al rischio, il prezzo di un'opzione non sarà necessariamente uguale al valore atteso del payoff scontato a un tasso privo di rischio.

Invece, sarà necessario utilizzare un tasso di sconto aggiustato per il rischio, che includa un premio per il rischio specifico per quel tipo di strumento e per la percezione del rischio da parte degli investitori.

Nella pratica finanziaria, i modelli che non assumono la neutralità al rischio vengono utilizzati per modellare il comportamento reale degli investitori. Ad esempio, nella *finanza comportamentale*, si cerca di spiegare le anomalie di mercato che non possono essere giustificate dai modelli classici, come l'eccessiva volatilità o i boom e crash di mercato. In questo contesto, si tengono in considerazione fattori psicologici e cognitivi, come la paura o l'euforia [4]. La *teoria dell'utilità attesa* o modelli più avanzati, come la *teoria del prospetto* proposta da Kahneman e Tversky, cercano di rappresentare come gli investitori percepiscono realmente guadagni e perdite, e come le loro scelte di investimento siano influenzate dalle probabilità soggettive che assegnano agli eventi [3].

Capitolo 4

Metodo Monte Carlo per la valutazione delle opzioni

4.1 Introduzione

Nel contesto della finanza computazionale, la valutazione delle opzioni finanziarie rappresenta una sfida cruciale, soprattutto quando si ha a che fare con modelli complessi che non consentono di ottenere soluzioni analitiche chiuse. Tra gli approcci numerici sviluppati per affrontare questo problema, il **Metodo Monte Carlo** occupa una posizione di rilievo, grazie alla sua capacità di fornire stime precise anche in presenza di elevata complessità modellistica. Il metodo Monte Carlo è un approccio numerico basato sul campionamento casuale. Quando applicato alla valutazione delle opzioni, il Metodo Monte Carlo prevede la simulazione di molteplici scenari per l'evoluzione del prezzo dell'attività sottostante, ipotizzando quest'ultimo segua un *moto browniano geometrico*. Il fine è quello di calcolare una stima del prezzo dell'opzione, ottenuta come la media dei payoff attualizzati, assumendo una valutazione neutrale al rischio. Il principale vantaggio di questo metodo è la sua capacità di pervenire ad una valutazione finale dello strumento derivato che considera l'intero percorso temporale seguito dall'asset sottostante, S , e non solo il suo valore finale. Ecco perché il Metodo Monte Carlo è particolarmente utile nella valutazione di opzioni esotiche o di derivati finanziari il cui payoff dipende dall'intero percorso del prezzo, piuttosto che dal suo valore alla scadenza. Tuttavia, un importante svantaggio del metodo Monte Carlo consiste nel suo elevato costo computazionale e la difficoltà nell'affrontare situazioni che prevedono l'esercizio anticipato delle opzioni. Per questo motivo, il metodo è generalmente utilizzato per la valutazione delle sole opzioni europee. Il motivo di questa limitazione risiede nel fatto che, come già detto, le opzioni

americane hanno maggiore flessibilità d'esercizio durante la vita del contratto. Quindi, è fondamentale stabilire il momento ottimale per l'esercizio anticipato, il che richiede il calcolo del prezzo dell'opzione in tutti i possibili valori del prezzo dell'asset e del tempo, fino alla scadenza.

In questo capitolo, esamineremo i passaggi fondamentali delle simulazioni Monte Carlo e successivamente, nel capitolo dedicato all'analisi e al confronto con altri approcci, valuteremo i risultati ottenuti dall'applicazione del metodo per la valutazione di opzioni europee Call e Put.

4.2 I passi della procedura

Con una procedura di campionamento, si calcola il valore atteso dell'opzione in un mondo neutrale verso il rischio e lo si attualizza in base al tasso d'interesse privo di rischio. Consideriamo un derivato, con scadenza in T , che dipende dal solo prezzo del sottostante, S . Se i tassi d'interesse sono costanti, il derivato può essere così valutato:

1. si simula un sentiero temporale per S in un mondo neutrale verso il rischio;
2. si calcola il valore finale del derivato;
3. si ripetono i passi 1 e 2 per ottenere una serie di valori finali campionari;
4. si stima il valore finale atteso del derivato, in un mondo neutrale verso il rischio, come media aritmetica dei valori finali campionari;
5. si determina il valore corrente del derivato attualizzando al tasso privo di rischio il valore finale atteso.

4.2.1 Simulazione dell'evoluzione del prezzo del sottostante

Il prezzo dell'attività sottostante S è modellato utilizzando un'equazione differenziale stocastica (SDE) sotto la misura neutrale verso il rischio, già definita nell'equazione (3.15) del capitolo precedente:

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

dove:

- μ è il tasso di crescita atteso del prezzo dell'azione.

- σ è la volatilità del prezzo dell'azione.
- W_t è un moto browniano standard.

4.3 Dipendenza Markoviana e Suddivisione degli Intervalli

La simulazione Monte Carlo si basa sulla suddivisione della vita del derivato in tanti piccoli incrementi di tempo Δt . Ogni incremento è determinato esclusivamente dal valore precedente, poiché si tratta di un processo markoviano. Questo significa che il valore futuro di S dipende solo dal suo valore attuale e non dalla sequenza di eventi che lo ha preceduto.

4.3.1 Costruzione dei Sentieri Temporali

L'approssimazione discretizzata della (3.1) consente di determinare il valore di S al tempo t in base al valore di S al tempo zero. Il valore al tempo $\Delta t + 1$ è basato sul valore al tempo Δt , e così via. Ogni simulazione comporta la costruzione di un intero sentiero temporale per S sulla base di N estrazioni casuali da una distribuzione normale.

$$S(t + \Delta t) - S(t) = \mu S(t) \Delta t + \sigma S(t) \varepsilon \sqrt{\Delta t} \quad (4.1)$$

In pratica, sulla base della non negatività dei prezzi, è di solito più accurato simulare $\ln(S)$ piuttosto che S . Utilizzando il lemma di Itô, il processo seguito da $\ln(S)$ è:

$$d \ln(S) = \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dW_t \quad (4.2)$$

Discretizzando questa equazione, otteniamo:

$$\ln \left(\frac{S(t + \Delta t)}{S(t)} \right) = \left(\mu - \frac{\sigma^2}{2} \right) \Delta t + \sigma \epsilon \sqrt{\Delta t} \quad (4.3)$$

Così,

$$S(t + \Delta t) = S(t) \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) \Delta t + \sigma \epsilon \sqrt{\Delta t} \right) \quad (4.4)$$

Il vantaggio di lavorare con $\ln(S)$ sta nel fatto che questa trasformazione di S segue un moto Browniano Geometrico. Pertanto, si giunge all'equazione:

$$\ln\left(\frac{S(T)}{S(0)}\right) = \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\epsilon \quad (4.5)$$

cui vale per qualsiasi T . Ne segue che

$$S(T) = S(0) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\epsilon\right) \quad (4.6)$$

4.3.2 Calcolo del Payoff

Si simulano più percorsi (un grande numero N) del prezzo dell'attività S da $t = 0$ a $t = T$. Per ogni percorso, si calcola il payoff finale dell'opzione al tempo T . Come precedentemente discusso, il payoff finale dipende dal tipo di opzione:

- Per un'opzione call: $\text{Payoff}_{\text{call}} = \max(S(T) - K, 0)$
- Per un'opzione put: $\text{Payoff}_{\text{put}} = \max(K - S(T), 0)$

Il payoff atteso è stimato mediando i payoff di tutti i percorsi simulati:

$$\text{Payoff medio}_{\text{call}} = \frac{1}{N} \sum_{i=1}^N \max(S_i(T) - K, 0) \quad (4.7)$$

$$\text{Payoff medio}_{\text{put}} = \frac{1}{N} \sum_{i=1}^N \max(K - S_i(T), 0) \quad (4.8)$$

4.3.3 Misura Neutrale al Rischio

Nel contesto della misura neutrale al rischio, il rendimento atteso dell'asset sottostante è impostato al tasso privo di rischio r anziché al tasso di rendimento atteso del mercato μ . Di conseguenza, in sostituzione alla (4.6), per simulare i percorsi del prezzo dell'asset sotto la misura neutrale al rischio, utilizziamo la seguente formula [2]:

$$S(T) = S(0) \exp\left(\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\epsilon\right) \quad (4.9)$$

dove:

- $S(T)$ è il prezzo dell'asset al tempo T ,
- $S(0)$ è il prezzo iniziale dell'asset al tempo $t = 0$,
- r è il tasso privo di rischio,
- σ è la volatilità dell'asset,
- ϵ è una variabile casuale normalmente distribuita con media zero e varianza uno,
- T è l'orizzonte temporale.

Nella misura neutrale al rischio, il drift del processo di prezzo è dato da:

$$r - \frac{\sigma^2}{2} \quad (4.10)$$

Questo significa che, sotto la misura neutrale al rischio, il rendimento atteso dell'asset è pari al tasso privo di rischio r , e non al tasso di rendimento atteso del mercato μ . La parte stocastica della formula, $\sigma\sqrt{T}\epsilon$, rappresenta l'effetto della volatilità sull'andamento del prezzo dell'asset.

4.3.4 Attualizzazione del Payoff

Il prezzo attuale dell'opzione si ottiene attualizzando il payoff medio al tasso privo di rischio r :

$$\text{Prezzo dell'Opzione Call} = e^{-rT} \cdot \text{Payoff medio}_{\text{call}} \quad (4.11)$$

$$\text{Prezzo dell'Opzione Put} = e^{-rT} \cdot \text{Payoff medio}_{\text{put}} \quad (4.12)$$

4.4 Ottimizzazione del Numero di Simulazioni

Il numero delle simulazioni da effettuare dipende dall'accuratezza richiesta. Se si effettuano N simulazioni indipendenti, si calcola di solito la deviazione standard, oltre che la media, dei valori correnti del derivato ottenuti nelle N simulazioni. Nel contesto fornito, μ è una variabile. Rappresenta la stima della media del valore corrente del derivato calcolata sulle N simulazioni Monte Carlo. Poiché è una stima basata su un campione (le simulazioni), può variare a seconda del campione di simulazioni utilizzato. L'errore standard della stima è:

$$\frac{\sigma}{\sqrt{N}} \quad (4.13)$$

Pertanto, l'intervallo di confidenza al 95% per il prezzo f del derivato è:

$$\mu - 1.96 \frac{\sigma}{\sqrt{N}} < f < \mu + 1.96 \frac{\sigma}{\sqrt{N}} \quad (4.14)$$

Come si vede, l'incertezza circa il valore del derivato è inversamente proporzionale alla radice quadrata del numero delle simulazioni. Per raddoppiare l'accuratezza della stima, dobbiamo quadruplicare il numero delle simulazioni; per decuplicare l'accuratezza dobbiamo centuplicare il numero delle simulazioni; e così via.

I campioni richiesti, ϵ_i , si ottengono nel modo seguente:

$$\epsilon_1 = \alpha_{11}x_1 \quad (4.15)$$

$$\epsilon_2 = \alpha_{21}x_1 + \alpha_{22}x_2 \quad (4.16)$$

$$\epsilon_3 = \alpha_{31}x_1 + \alpha_{32}x_2 + \alpha_{33}x_3 \quad (4.17)$$

e così via.

Questa procedura è nota come "scomposizione di Cholesky".

4.5 Generazione di Numeri Casuali e Quasi-Casuali

Uno dei pilastri del **Metodo Monte Carlo** è la generazione di numeri casuali, che devono essere distribuiti in modo uniforme e indipendente per garantire l'affidabilità delle simulazioni. Nella pratica, si ricorre a sequenze pseudo-casuali, generate attraverso algoritmi come il metodo delle congruenze lineari. Tuttavia, questi metodi possono presentare limitazioni nella distribuzione uniforme nello spazio multidimensionale, problema che può essere parzialmente mitigato dall'impiego di numeri quasi-casuali. Le sequenze quasi-casuali, come la sequenza di Halton o quella di Sobol, sono progettate per migliorare l'uniformità della copertura spaziale, riducendo la *discrepanza*, ovvero la deviazione rispetto alla distribuzione uniforme ideale. Queste sequenze permettono di ottenere simulazioni più efficienti, soprattutto in contesti di alta dimensionalità, tipici della finanza.

4.6 Tecniche di Riduzione della Varianza

Il **Metodo Monte Carlo** può richiedere un elevato numero di simulazioni per ottenere stime accurate, il che comporta un significativo onere computazionale. Per mitigare questo problema, vengono utilizzate diverse tecniche di **riduzione della varianza**. Tra le più comuni vi sono:

- **Il metodo delle variabili di controllo:** Si utilizza una variabile ausiliaria, la cui speranza matematica è nota, per ridurre la varianza della stima.
- **Il campionamento stratificato:** Consiste nel dividere lo spazio di simulazione in sottospazi (strati) e nel campionare in modo più uniforme all'interno di ciascuno di essi.
- **Il metodo di antithetic variates:** Utilizza la simulazione di coppie di variabili casuali, con valori speculari, al fine di ridurre la varianza complessiva delle stime.
- **Importance Sampling:** Modifica la distribuzione di probabilità per concentrare le simulazioni nelle regioni più significative, migliorando l'efficienza delle stime.

Queste tecniche consentono di migliorare l'efficienza delle simulazioni, riducendo il numero di iterazioni necessarie per ottenere una precisione desiderata.

Capitolo 5

Extra Gradient Boost (XGBoost)

XGBoost (eXtreme Gradient Boosting) è un potente algoritmo di machine learning sviluppato per affrontare problemi di regressione, classificazione e ranking. Appartenente alla famiglia dei modelli di *ensemble learning*, il modello XGBoost si distingue per l'efficienza computazionale, la gestione ottimizzata dei dati mancanti e la capacità di analizzare l'importanza delle variabili esplicative. Grazie alla sua flessibilità e robustezza, è diventato uno degli strumenti più utilizzati per la previsione dei prezzi di titoli azionari e non solo.

5.1 L'Ensemble Learning

Il concetto di ensemble learning è la base del modello XGBoost. Questa tecnica combina le previsioni di più modelli di base per migliorare la precisione complessiva del modello finale. L'idea è che, aggregando i risultati di vari modelli secondo diversi schemi, si possa ottenere un unico modello finale più robusto e accurato rispetto a quelli individuali.

5.1.1 Bagging

Il *bagging* (Bootstrap Aggregating) è una tecnica di ensemble che riduce la varianza del modello combinando previsioni di modelli che vengono addestrati in parallelo su campioni casuali del dataset. Un esempio noto di bagging è il Random Forest, dove molti alberi decisionali vengono addestrati in parallelo e le loro previsioni sono mediate per ottenere il risultato finale.

5.1.2 Boosting

Il *boosting* è un'altra tecnica di ensemble, ma a differenza del bagging, costruisce i modelli base in sequenza. Ogni nuovo modello si concentra sugli errori commessi dai modelli precedenti, migliorando , secondo uno schema sequenziale, la performance complessiva. Il gradient boosting, di cui XGBoost è un'implementazione avanzata, è una delle tecniche più popolari di boosting.

5.2 Gradient Boosting

Il gradient boosting è il cuore dell'algoritmo XGBoost e si basa sull'idea di ottimizzare un modello in modo iterativo aggiungendo nuovi modelli per ottimizzare i residui del modello corrente. Dunque, si costruisce il modello finale secondo una sequenza incrementale.

Il processo inizia con la definizione di un modello iniziale F_0 , spesso rappresentato dalla media dei valori target. Questo modello fornisce una stima di base, di cui si calcolano i residui, ossia la differenza tra i valori target osservati e quelli predetti da F_0 . Questi rappresentano gli errori che il modello $h_1(x)$ cercherà di stimare.

In XGBoost, $h_1(x)$ è generalmente un albero di regressione che stima i residui basandosi su una selezione dei regressori più significativi tra le variabili selezionate x .

Il risultato è il modello aggiornato $F_1(x)$ che combina il modello iniziale $F_0(x)$ con le correzioni fornite da $h_1(x)$:

$$F_1(x) = F_0(x) + h_1(x)$$

Questo processo continua in una sequenza iterativa, dove ogni nuovo modello aggiuntivo contribuisce a ridurre gli errori complessivi.

5.3 Inizializzazione del Modello con la Funzione di Perdita

L'inizializzazione di un modello di regressione in XGBoost si basa sull'ottimizzazione della funzione di perdita, tipicamente l'errore quadratico medio (MSE). Il problema si traduce in minimizzare l'MSE tra le previsioni del modello e i valori osservati. La funzione di perdita è formulata come:

$$L(y_i, F_0(x_i)) = (y_i - F_0(x_i))^2$$

Si cerca di minimizzare la somma degli errori quadrati, trovando la funzione $F_0(x)$ che ottimizza le previsioni:

$$\operatorname{argmin}_{F_0} \sum_{i=1}^n (y_i - F_0(x_i))^2$$

La funzione $F_0(x)$ che minimizza l'MSE si ottiene calcolando la derivata parziale rispetto a $F_0(x)$ e risolvendo:

$$\frac{\partial}{\partial F_0(x)} \left[\sum_{i=1}^n (y_i - F_0(x_i))^2 \right] = 0$$

Il valore ottimale per $F_0(x)$ è la media dei valori target osservati, poiché la media minimizza la somma degli errori al quadrato.

Questa fase preliminare di costruzione del modello crea la base del processo di minimizzazione della funzione di perdita tramite il metodo della discesa del gradiente.

5.3.1 Discesa del Gradiente in XGBoost

Gradiente: il gradiente rappresenta la derivata della funzione di perdita rispetto alle predizioni del modello. In XGBoost, il gradiente indica quanto cambia la funzione di perdita rispetto ad una piccola variazione nelle predizioni.

Dopo ogni iterazione, XGBoost calcola il gradiente della funzione di perdita rispetto alle predizioni attuali del modello. Questo gradiente rappresenta quanto il modello sta "sbagliando" in ogni punto. Così, XGBoost costruisce un nuovo albero decisionale che cerca di correggere questi errori. Questo albero è addestrato per prevedere il gradiente, ossia la differenza tra le predizioni attuali e i valori reali. Il nuovo albero viene aggiunto al modello corrente. Le stime attuali vengono aggiornate sommando le stime dell'albero appena costruito a quelle esistenti, ma moltiplicate per un piccolo fattore (il tasso di apprendimento, indicato con η). Questo passaggio avvicina le predizioni ai valori reali, riducendo la funzione di perdita. Questo processo viene ripetuto iterativamente, con XGBoost che costruisce nuovi alberi che correggono gli errori residui dei precedenti. Ogni nuovo albero cerca di ridurre ulteriormente la funzione di perdita, migliorando così la precisione del modello.

5.4 Iperparametri di XGBoost

La scelta degli iperparametri in XGBoost è cruciale per ottimizzare le prestazioni del modello. Di seguito viene fornita una panoramica dei principali iperparametri:

- **nrounds:** Numero di iterazioni del boosting. Un numero elevato può migliorare le prestazioni ma aumenta il rischio di overfitting. È uno dei parametri più importanti da calibrare.
- **max depth:** Profondità massima degli alberi. Controlla la complessità del modello. Alberi più profondi possono catturare relazioni più complesse, ma possono anche causare overfitting.
- **min child weight:** Peso minimo di una foglia. Determina la somma minima dei pesi (ossia il numero di osservazioni) in una foglia dell'albero. Un valore maggiore rende il modello più conservativo.
- **eta:** Tasso di apprendimento, spesso indicato come shrinkage. Valori più bassi rallentano l'apprendimento, richiedendo più iterazioni ma migliorando la generalizzazione.
- **gamma:** Riduzione minima della perdita necessaria per effettuare una suddivisione in un nodo. Un valore maggiore rende il modello più conservativo e può prevenire la creazione di nodi irrilevanti.
- **subsample:** Frazione di dati utilizzati per costruire ogni albero. Utilizzare solo una parte del dataset può aiutare a prevenire l'overfitting, mantenendo il modello più robusto.
- **col sample by tree:** Frazione di variabili utilizzate per costruire ogni albero. Ridurre questa frazione può aiutare a prevenire l'overfitting, migliorando la robustezza del modello.

5.5 Cross-Validation: Il Metodo dei 5 Folds

La cross-validation è una tecnica essenziale per valutare l'affidabilità di un modello, specialmente quando si dispone di un dataset limitato. La 5-fold cross-validation è uno dei metodi più comuni e funziona come segue:

1. **Suddivisione del Dataset:** il dataset viene diviso in 5 parti uguali (folds).

2. Addestramento e Validazione: in ogni iterazione, 4 folds vengono utilizzati per addestrare il modello e 1 fold viene utilizzato per la validazione. Questo processo viene ripetuto 5 volte, ruotando il fold di validazione ogni volta.

3. Calcolo delle Prestazioni: si calcolano le metriche di prestazione per ciascuna iterazione (ad esempio, accuracy, F1-score, RMSE) e si ottiene la media per avere una stima complessiva delle prestazioni del modello.

Capitolo 6

Dati e Analisi Esplorativa

6.1 Dati Originali

Prima di approfondire i dati utilizzati per l’analisi, è fondamentale delineare il contesto in cui verrà implementata. In particolare, è importante sottolineare l’elevata sensibilità dei dati trattati, che rende il loro reperimento universalmente complesso. Poiché si tratta di dati relativi a contratti finanziari, sono generalmente considerati *dati sensibili* e, come tali, inaccessibili a investitori privati non professionisti. La situazione non è migliore per ricercatori e studenti.

Il sito del New York Stock Exchange richiede la compilazione di specifici form per poter avanzare una richiesta di accesso ai dati, ma presuppone che l’utente sia un investitore professionista registrato sulla piattaforma, con un abbonamento attivo. Situazione analoga si verifica per la piattaforma dati del Nasdaq¹. Un’opzione alternativa è il portale della Princeton University, che offre un repository ben fornito, ma riservato esclusivamente a dottorandi di università partner.²

Altre fonti di dati includono la Options Clearing Corporation (OCC), Bar-chart, Kibot, e Interactive Brokers. Tuttavia, piattaforme come Yahoo Finance e Quandl, che offrono ampie serie storiche sui prezzi dei titoli azionari, non consentono il download di serie storiche sulle opzioni.

Il dataset *Flex Trades* contiene informazioni sulle transazioni di contratti d’opzione *flex* negoziate nel Cboe Global Market, tutte registrate il 20 giugno 2024, a partire dalle ore 09:46 fino alle 16:43 dello stesso giorno.

Il file, denominato `flex_trades_eod_2024-06-20.csv.zip`, è reperibile nella sezione DataShop del sito ufficiale del Chicago Board Options Exchange

¹<https://data.nasdaq.com>

²<https://dss.princeton.edu/catalog>

(Cboe) sotto la sezione Flex Reports ³. Questo campione comprende 1098 osservazioni, con l'intera serie storica di luglio 2024 disponibile a pagamento. Il dataset utilizza l'expertise di Cboe Hanweck, una divisione specializzata in dati avanzati e analisi dei mercati finanziari, per calcolare variabili come:

- Prezzo teorico proprietario: calcolato con modelli avanzati di pricing. Questi sono interni alla sezione e non vengono fornite informazioni circa le strutture.
- volatilità: una stima della volatilità implicita del contratto.
- Greche: misure di sensitività del contratto di opzione (Delta, Gamma, Vega, Theta, Rho).

Le variabili del dataset includono:

1. **trade_datetime**: Data e ora dell'operazione.
2. **underlying_symbol**: Simbolo del sottostante.
3. **root**: Simbolo di trading dell'opzione.
4. **expiration**: Data di scadenza dell'opzione.
5. **strike**: Prezzo di esercizio dell'opzione.
6. **option_type**: Tipo di opzione (Call o Put).
7. **settlement**: Tipo di liquidazione (AM/PM). Il prezzo di settlement (liquidazione) per le opzioni AM è calcolato utilizzando i prezzi di apertura degli asset sottostanti nel giorno di scadenza. Questo significa che il prezzo finale è determinato al mattino (da cui "AM"). Il prezzo di settlement per le opzioni PM è determinato utilizzando i prezzi di chiusura degli asset sottostanti nel giorno di scadenza. Il prezzo finale è quindi stabilito alla fine della giornata di trading (da cui "PM").
8. **exercise**: Stile di esercizio dell'opzione.
9. **delivery**: Tipo di consegna (fisica o in contanti).
10. **multiplier**: Moltiplicatore sul contratto.
11. **trade_price**: Prezzo di negoziazione dell'opzione.
12. **trade_size**: Quantità negoziata.

³<https://datashop.cboe.com/flex-reports>

13. **theo_price**: Prezzo teorico dell'opzione.
14. **theo_iv**: Volatilità implicita teorica.
15. **trade_iv**: Volatilità implicita basata sul prezzo di negoziazione.
16. **trade_delta**: Delta al momento della negoziazione.
17. **trade_gamma**: Gamma al momento della negoziazione.
18. **trade_vega**: Vega al momento della negoziazione.
19. **trade_theta**: Theta al momento della negoziazione.
20. **trade_rho**: Rho al momento della negoziazione.
21. **open_interest**: Interesse aperto all'inizio della data di mercato corrente.
22. **trade_count**: Numero di negoziazioni.
23. **underlying_bid**: Miglior prezzo di offerta sull'ultima quotazione del sottostante.
24. **underlying_ask**: Miglior prezzo di domanda sull'ultima quotazione del sottostante.
25. **underlying_trade_price**: Ultimo prezzo di negoziazione del sottostante.

6.2 Calcolo della Volatilità Storica degli Asset Sottostanti

Come già detto, la volatilità storica è una misura chiave nella valutazione del rischio finanziario. Tuttavia, questa variabile risulta il grande assente tra gli elementi presenti nel dataset originale. Poiché rappresenta sia un'informazione chiave relativa all'asset sottostante, sia un elemento propedeutico all'implementazione dei modelli dapprima presentati. Questa verrà calcolata per ogni simbolo associato agli asset sottostanti presenti nel dataset.

6.2.1 Definizione delle finestre Temporali

La scelta delle finestre temporali per il calcolo della volatilità è cruciale, poiché influenza direttamente sulla stima della stessa. Di seguito, si presentano alcuni punti da considerare riguardo alla scelta delle finestre temporali e al loro impatto:

Volatilità a Breve Termine Se si utilizza una finestra temporale breve (ad esempio, 1 anno), la volatilità stimata rifletterà principalmente i movimenti di prezzo recenti. Questo può essere utile se il mercato ha subito un cambiamento significativo di recente, ma potrebbe non rappresentare accuratamente la volatilità di lungo periodo.

Volatilità a Medio/Lungo Termine Finestre più lunghe (2-3 anni) tendono a catturare una volatilità più stabile e meno sensibile a eventi recenti di mercato. Tuttavia, possono essere meno reattive ai cambiamenti rapidi nelle condizioni di mercato, rendendole meno adatte per valutazioni che richiedono una maggiore sensibilità ai movimenti recenti.

In mercati stabili, finestre lunghe possono fornire una stima più affidabile della volatilità. In mercati volatili o in transizione, finestre brevi potrebbero offrire una misura più accurata della volatilità attuale.

Uno degli obiettivi principali di questa analisi consiste proprio nel determinare quale finestra temporale risulta ottimale per prevedere i prezzi teorici e di scambio delle diverse tipologie di opzioni.

6.2.2 Selezione dei Dati

Il primo passo è stato selezionare i simboli univocamente presenti nel dataset, escludendo quelli che rappresentano indici (simboli contenenti il carattere `~`). Questo filtro è necessario perché la successiva analisi di valutazione dei prezzi si concentrerà sulle sole opzioni scritte su titoli azionari e non sugli indici. Successivamente, i dati sono stati filtrati per includere solo i simboli selezionati. Sono state convertite le date di scadenza (`expiration`) e le date di trading (`trade datetime`) in formato data, e calcolato il numero di giorni che intercorre tra queste per definire il numero di giorni alla scadenza del contratto. Inoltre, il dataset è stato filtrato per mantenere solo le opzioni di tipo europeo (`exercise == "E"`).

Il risultato è un dataset filtrato, *dati rid* contenente 990 osservazioni.

6.2.3 Calcolo della Volatilità Annuale

La volatilità annuale è stata calcolata per ciascun simbolo e per ogni finestra temporale utilizzando i dati storici dei prezzi di chiusura aggiustata. Il calcolo è stato effettuato attraverso una funzione che recupera i dati da Yahoo Finance, calcola il rendimento giornaliero sulla serie dei prezzi aggiustati e determina la deviazione standard, moltiplicata per la radice quadrata del numero di giorni di trading annuali.

Sulla base di quanto esposto prima, si è, inoltre, ripetuta la procedura su tre finestre temporali differenti, che coprono periodi variabili, al fine di catturare eventuali cambiamenti nella volatilità degli asset nel tempo. Le finestre temporali ricoprono rispettivamente 1, 2 e 3 anni dal giorno dell'avvenuta transazione registrata nel dataset.

6.2.4 Applicazione della Funzione

La funzione è stata applicata a ciascun simbolo e per ogni finestra temporale. I risultati ottenuti sono stati organizzati in un dataframe per poter essere integrati con i dati già esistenti. Infine, le colonne relative alla volatilità storica sono state unite al dataset principale `dati rid`, sotto la dicitura di `volatilità annuale 1`, `volatilità annuale 2` e `volatilità annuale 3`. Questo processo ha permesso di aggiungere al dataset originale le colonne della volatilità storica annuale per ciascun asset sottostante, calcolata sulle tre diverse finestre temporali. Questi dati saranno successivamente utilizzati nei modelli di pricing per la valutazione delle opzioni.

6.3 Analisi Outliers con Local Outlier Factor

L'algoritmo LOF calcola un punteggio che misura la deviazione della densità locale di un dato punto rispetto ai suoi vicini. Una spiegazione più dettagliata si trova consultando il paper originale [8]. I passaggi fondamentali sono: caricamento dei dati, preparazione dei subset, applicazione dell'algoritmo, separazione di inliers e outliers, combinazione dei risultati e mescolamento righe funzionale alla 5 folds cross validation.

La descrizione dettagliata del codice R è disponibile in Appendice, sotto la dicitura di "Listing Codice R LOF".

6.4 Analisi Esplorativa

Applicando l'algoritmo Local Outlier Factor (LOF), il numero di osservazioni nel dataset è stato ridotto da 990 a 927. In dettaglio, inizialmente il dataset comprendeva 538 opzioni Call e 432 opzioni Put. Dopo l'applicazione dell'algoritmo LOF, che ha identificato e rimosso le osservazioni considerate anomale, sono rimaste 530 opzioni Call e 397 opzioni Put. Questo processo di rimozione degli outlier ha quindi comportato una diminuzione totale di 63 osservazioni, migliorando la qualità e l'affidabilità del dataset per le successive analisi.

Non ci sono opzioni con prezzo uguale a 0. Tuttavia, 202 opzioni presentano un prezzo inferiore a \$1 (79 opzioni Call; 123 opzioni Put).

Il prezzo medio delle opzioni Call è pari a \$133.561. Il 25% dei valori è inferiore a \$4.598. La mediana è pari a \$19.055, quindi significativamente inferiore alla media, indicando un'asimmetria positiva. Questa è confermata dalla differenza che intercorre tra il primo ed il terzo quartile , tabella 6.1. Il massimo prezzo registrato per una Call è pari a \$5547.650.

Così come per le Call, anche il valore minimo della distribuzione dei prezzi di scambio delle opzioni Put è \$0.01. Il 25% dei dati è inferiore a \$0.41. Questo suggerisce che una parte significativa dei dati si concentra in valori molto bassi. La mediana è pari a \$3.46 , mentre la media (\$57.86) è notevolmente più alta 6.2. Anche per la distribuzione dei prezzi delle opzioni Put ci sono alcuni valori molto alti che stanno elevando la media. Anche in questo caso, in figura 6.2 si nota una distribuzione asimmetrica con una coda lunga a destra. Il massimo prezzo tra le opzioni del dataset, corrisponde ad una opzione Put ed è pari a \$49415.90. Questo è riconducibile al sottostante *SPDR S&P 500 ETF Trust (SPY)* .

Statistiche	Min.	1 Qu.	Mediana	Media	3Qu.	Max.
Value	4.598	19.055	133.561	93.297	547.650	

Tabella 6.1: Distribuzione prezzi di scambio delle Call

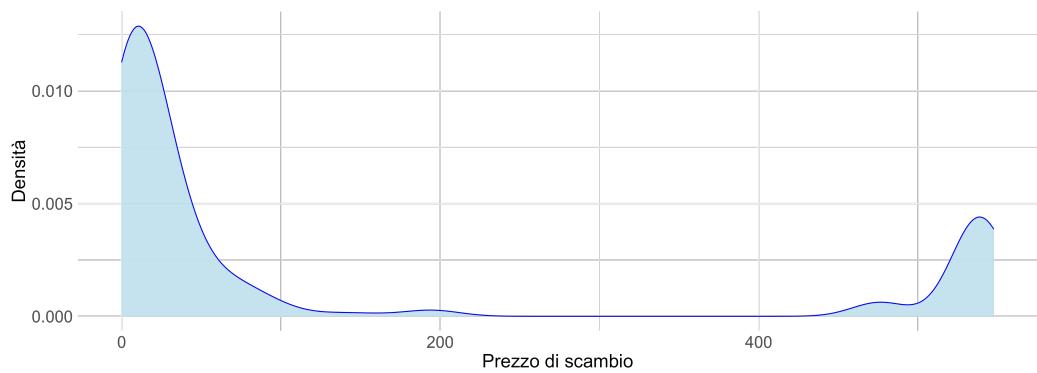


Figura 6.1: Density Plot Prezzi di Scambio Call

Statistiche	Min.	1 Qu.	Mediana	Media	3Qu.	Max.
Value	0.01	0.41	3.46	57.86	9.00	9415.90

Tabella 6.2: Distribuzione prezzi di scambio delle Put

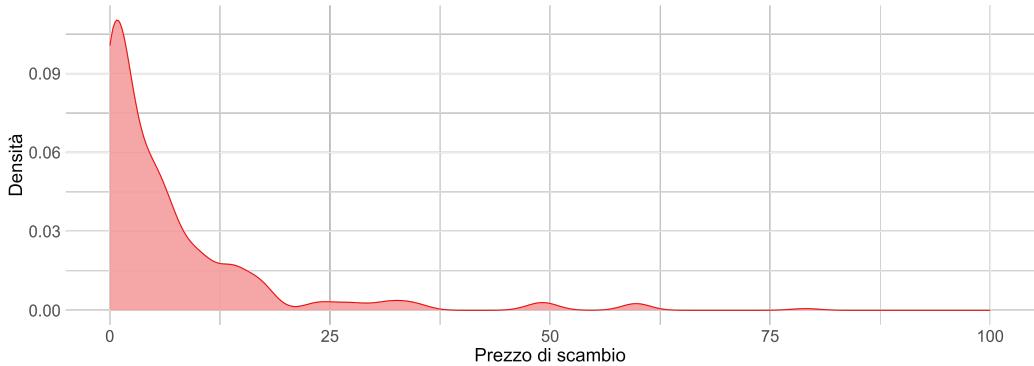


Figura 6.2: Density Plot Prezzi di Scambio Put

In relazione ai giorni che mancano all'esercizio, l'analisi esplorativa conferma quanto suggerito dal Capitolo 2 ed in tabella 2.1. Infatti, in figura 6.3 si nota come vi è incertezza circa l'andamento dei prezzi delle opzioni europee rispetto ai giorni che mancano alla maturità. La relazione non è più chiara se confrontiamo le singole tipologie (figura 6.4 e figura 6.5). Per le opzioni Put sembra esserci un andamento parabolico con concavità verso l'alto, per cui i prezzi più alti sembrano corrispondere contro intuitivamente ai contratti Put con scadenza imminente e scadenza più lontana. Questa osservazione potrebbe indicare che gli investitori attribuiscono un valore maggiore alle opzioni Put che scadono molto presto, probabilmente per proteggersi da movimenti imminenti del mercato, e anche a quelle che scadono dopo molto tempo, potenzialmente per via dell'incertezza a lungo termine. Dinamiche coerenti con il ruolo di strumento di copertura che le Put, ma in generale le opzioni, possono assolvere in una strategia di gestione del rischio applicata ad un portafoglio titoli.

Rispetto al prezzo d'esercizio, invece, si nota come i prezzi per le Call diminuiscono e aumentano per le Put. Rispettivamente in figura 6.6 e figura 6.7. Evidente è la relazione negativa tra il prezzo delle opzioni Put e il prezzo delle attività sottostanti 6.9. Infatti, i valori più alti di Put corrispondono tutti per prezzi del sottostante più bassi. In figura 6.8 si apprezza come la maggior parte dei punti si concentra nel quadrante inferiore sinistro del grafico, sug-

gerendo che la maggior parte delle opzioni Call ha un prezzo relativamente basso quando il prezzo di scambio dell'attività sottostante è basso. Nonostante nelle aree superiori a destra del grafico ci sono alcune opzioni Call con prezzi più alti in corrispondenza di prezzo di scambio dell'attività sottostante più alti. La concentrazione globale dei punti suggerisce una relazione positiva tra il prezzo di scambio dell'attività sottostante e il prezzo delle Call.

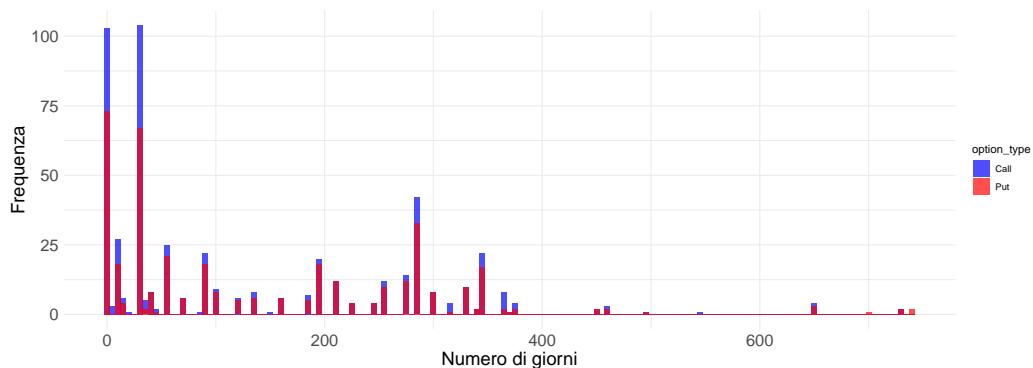


Figura 6.3: Distribuzione dei Prezzi di Scambio rispetto ai giorni alla maturità

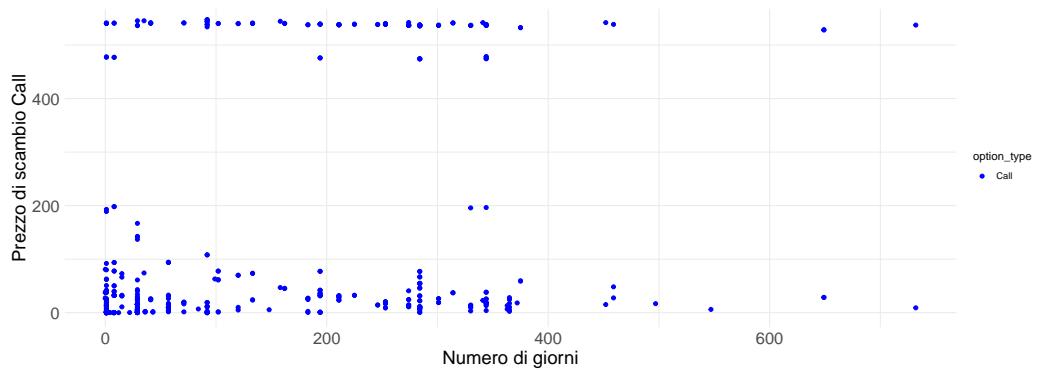


Figura 6.4: Scatter Plot Prezzi di scambio Call rispetto alla maturità

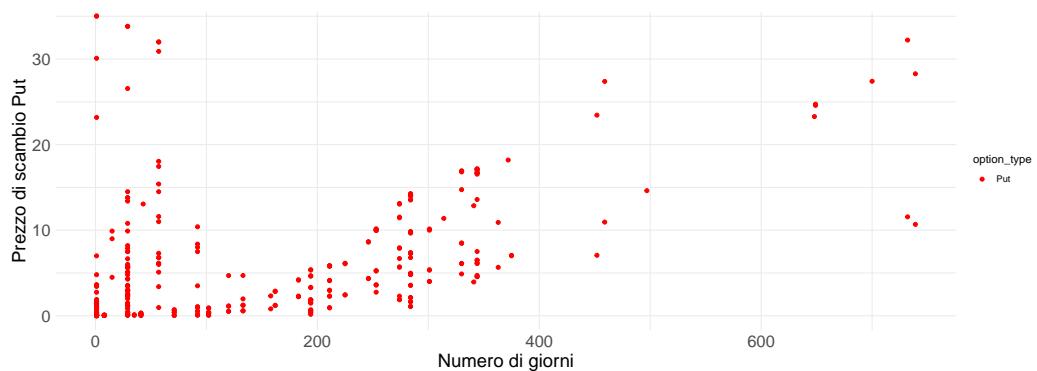


Figura 6.5: Scatter Plot Prezzi di scambio Put rispetto alla maturità

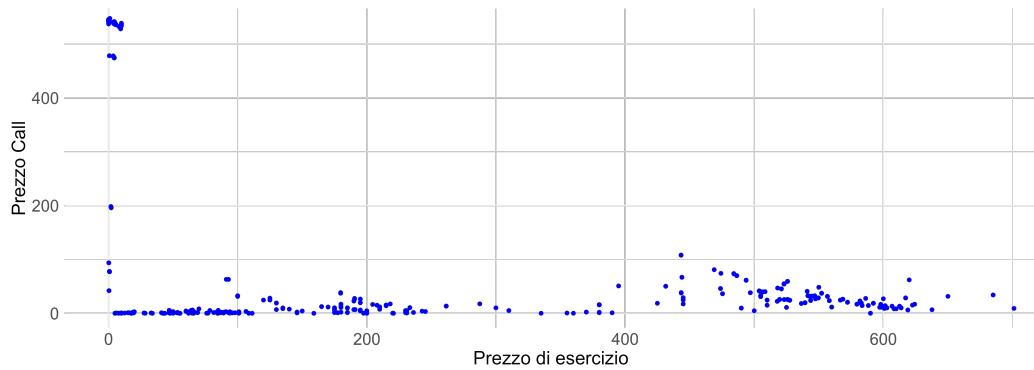


Figura 6.6: ScatterPlot Prezzi di scambio Call al rispetto prezzo d'esercizio

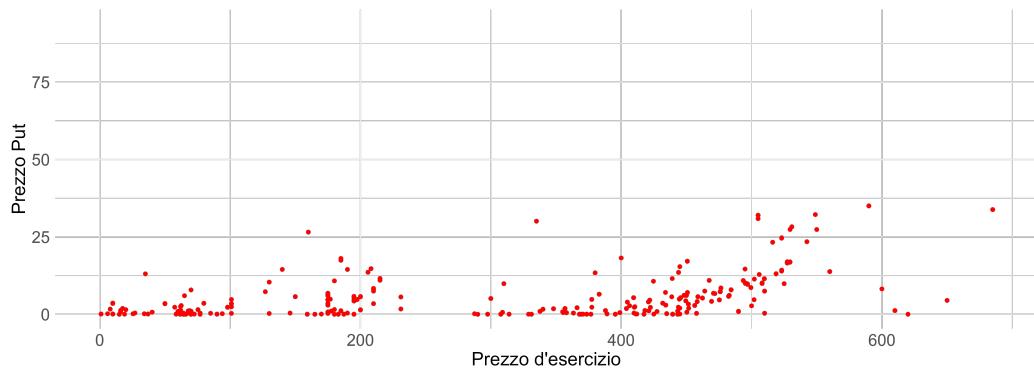


Figura 6.7: ScatterPlot Prezzi di scambio Put al rispetto prezzo d'esercizio

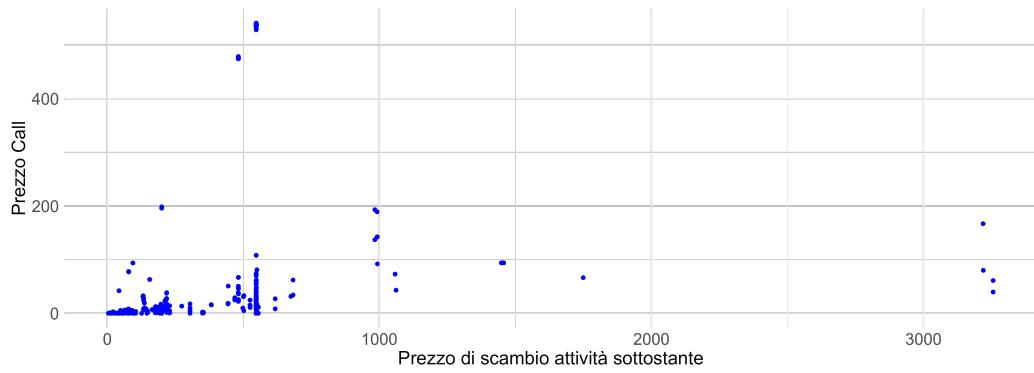


Figura 6.8: Prezzo di scambio Call rispetto prezzo attività sottostante

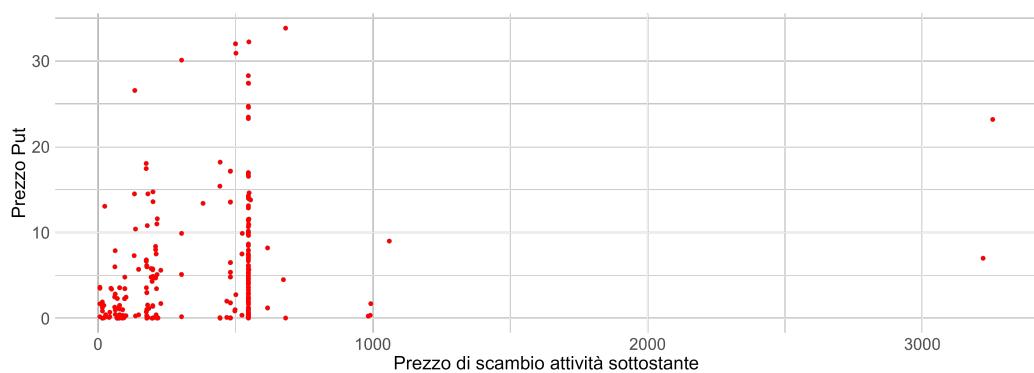


Figura 6.9: Prezzo di scambio Put rispetto prezzo attività sottostante

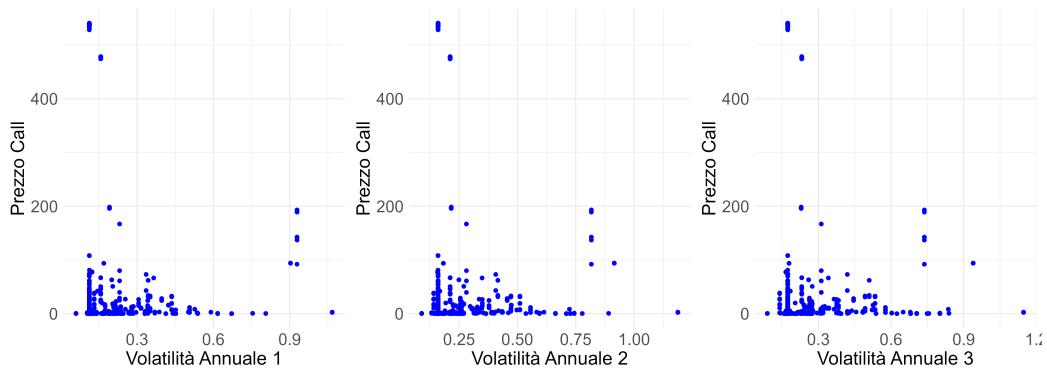


Figura 6.10: Prezzi di scambio Call in funzione delle 3 volatilità annuali calcolate

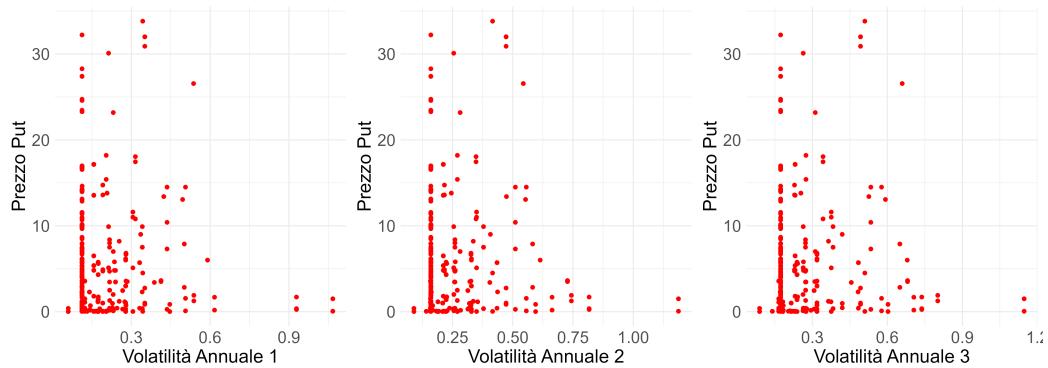


Figura 6.11: Prezzi di scambio Put in funzione delle 3 volatilità annuali calcolate

Nei grafici delle opzioni Call 6.10 e Put 6.11, si osserva la distribuzione dei prezzi di scambio rispetto a tre diverse misurazioni della volatilità annuale (calcolate rispettivamente su serie di rendimenti di 1 anno; 2 anni; 3 anni). Si nota che la maggior parte dei prezzi delle Put è concentrata a livelli di volatilità relativamente bassi, generalmente al di sotto di 0,3-0,4. A bassi livelli di volatilità (inferiori a 0,3), i prezzi delle Put mostrano una distribuzione ampia, con molti valori compresi tra \$0.1 e \$10, ma con alcuni prezzi che raggiungono valori più alti (fino a \$40). Quindi, prezzi bassi in corrispondenza di livelli di volatilità bassi. Tuttavia, si riscontra un andamento parabolico con concavità verso il basso. Infatti, fino a livello di volatilità 0,4 si hanno prezzi bassi in corrispondenza di livelli di volatilità bassi. Poi tendono ad aumentare all'aumentare di questa, ma oltre questo livello la relazione tende ad invertirsi. Inoltre, sembra che la dispersione dei prezzi aumenti nel caso della volatilità storica annuale 3.

Anche i prezzi delle Call sono concentrati a livelli di volatilità bassi (sotto 0,3-0,4). Nella fascia di bassa volatilità, i prezzi delle Call sono variabili, con alcuni valori che raggiungono anche \$400 o più. La maggior parte si trova nella fascia inferiore a \$200 in corrispondenza di livelli di volatilità non oltre 0,4. Con l'aumentare della volatilità, la dispersione dei prezzi delle Call si riduce notevolmente. Ma, anche per le Call non c'è una chiara tendenza di aumento dei prezzi con l'aumento della volatilità. Sembra piuttosto che si riducano.

In conclusione, i grafici mostrano pattern preliminari simili per ciascuna delle tre volatilità annuali. Tuttavia, non si traggono conclusioni definitive su quanto la scelta della volatilità influenzi i prezzi di scambio. In entrambi i casi, infatti, non sembra esserci una relazione lineare positiva, evidente ed immediata tra volatilità e prezzo delle opzioni. Questi risultati, piuttosto, suggeriscono che la volatilità influisce sui prezzi delle opzioni in modi complessi e meno immediato da quanto suggerito tabella 2.1. Lo scopo dell'analisi comparativa successiva, si concentrerà proprio sull'influenza delle diverse misure di volatilità nei modelli di previsione dei prezzi (Black-Scholes, Monte Carlo, XGBoost). Tale confronto sarà fondamentale per determinare l'effettivo impatto delle stime di volatilità annuali nella valutazione delle opzioni.

Analizzando l'andamento delle volatilità storiche annuali su un campione di azioni sottostanti, figura 6.12, graficamente emerge un aumento dei livelli della volatilità storica per la serie calcolata sul periodo di 3 anni. Per entrambe le opzioni Call e Put, i grafici 6.13 e 6.14 mostrano che la volatilità storica annuale tende ad aumentare in prossimità della scadenza, con variazioni più marcate per la volatilità annuale a 3 anni. A questa corrispondono anche i valori più alti. Questo comportamento è normale, poiché l'avvicinarsi della scadenza amplifica la sensibilità delle opzioni ai movimenti del prezzo del sottostante e agli eventi di mercato, rendendo la volatilità più variabile e elevata. Sebbene le Call e le Put rispondano simmetricamente ai cambiamenti nella volatilità, in alcuni casi possono verificarsi differenze dovute a squilibri di domanda e offerta o a aspettative di mercato asimmetriche. Tuttavia, l'aumento generale della volatilità vicino alla scadenza è un fenomeno comune e riflette l'incertezza crescente riguardo ai movimenti futuri del sottostante.

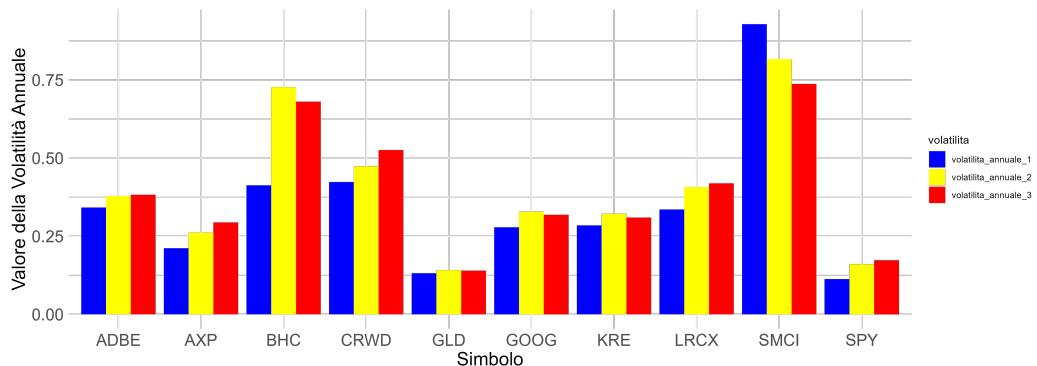
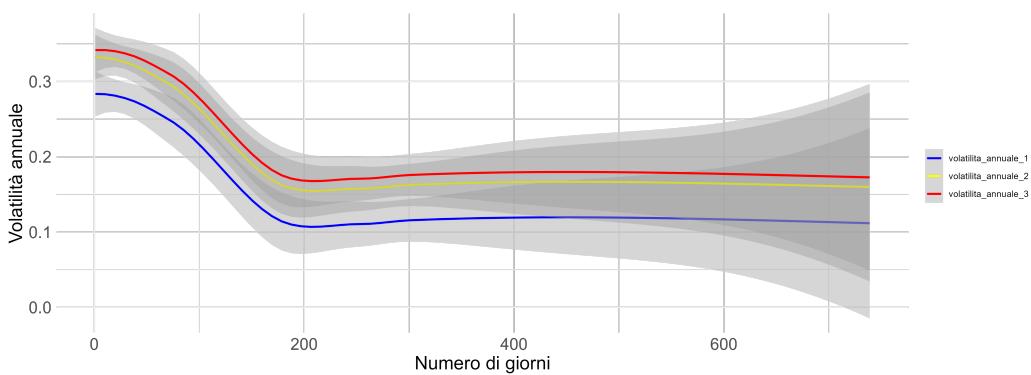
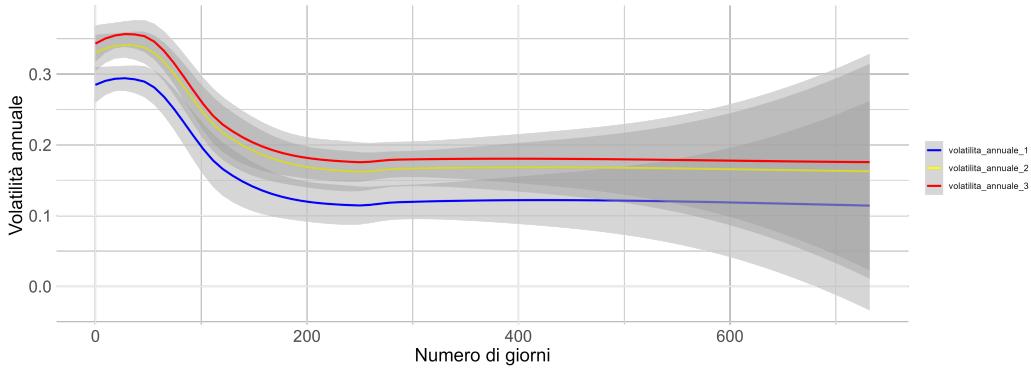


Figura 6.12: Volatilità storica 1,2 e 3 anni su un campione di asset sottostanti



Infine, analizziamo i prezzi delle opzioni rispetto alla volatilità implicita teorica. La volatilità implicita è un concetto fondamentale nel pricing delle opzioni, in quanto fornisce indicazioni sulle aspettative del mercato in merito alla volatilità futura. A differenza della volatilità storica, che si basa sui movimenti passati dei prezzi, la volatilità implicita è orientata al futuro. Riflette le aspettative del mercato sulla volatilità futura di un titolo, di un settore o del mercato nel suo complesso. In entrambi i casi, figure 6.15 e 6.16, si riscontra una concentrazione di punti con volatilità implicita bassa. Cioè, prezzi bassi in corrispondenza di livelli di volatilità implicita bassi. Questo sembra poi crescere e rimanere costante al crescere dei livelli di volatilità implicita, prefigurando la presenza di tre gruppi di prezzi. Inoltre, il prezzo

delle Call tende ad essere più disperso su valori di volatilità implicita bassi rispetto alle Put. Infatti, alcune Call hanno prezzi molto alti, specialmente con volatilità implicita vicina a zero.

L'andamento osservato nei grafici, con la presenza di tre gruppi distinti di prezzi delle opzioni in funzione della volatilità implicita, potrebbe sembrare insolito, ma può verificarsi in determinate condizioni di mercato. Questo fenomeno potrebbe essere dovuto a una segmentazione del mercato, dove gli investitori si concentrano su specifiche fasce di volatilità implicita considerate più "sicure" o "rischiose," portando così a una stabilizzazione dei prezzi in quei livelli. Inoltre, alcuni trader o istituzioni potrebbero adottare strategie come il market making o la copertura (hedging) che mantengono i prezzi delle opzioni stabili in determinate fasce di volatilità, indipendentemente dall'aumento della volatilità implicita. Infine, eventi di mercato come annunci aziendali, dati economici o scadenze di opzioni durante il periodo di osservazione potrebbero aver influenzato i prezzi in modo non lineare rispetto alla volatilità implicita.

Nonostante si siano evidenziate le differenze tra volatilità implicita e volatilità storica, entrambe mostrano andamenti complessi e non lineari rispetto ai prezzi di mercato, rendendo difficile individuare relazioni semplici e universali. Questo sottolinea come lo studio delle dinamiche tra volatilità e prezzi delle opzioni rappresenti ancora un campo di ricerca complesso e in evoluzione, con molte relazioni strutturali che restano da esplorare.

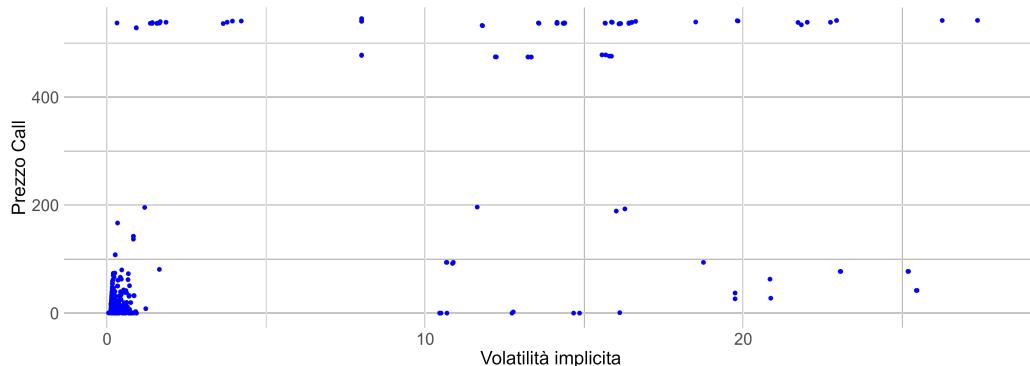


Figura 6.15: Prezzi opzioni Call rispetto alla volatilità implicita teorica

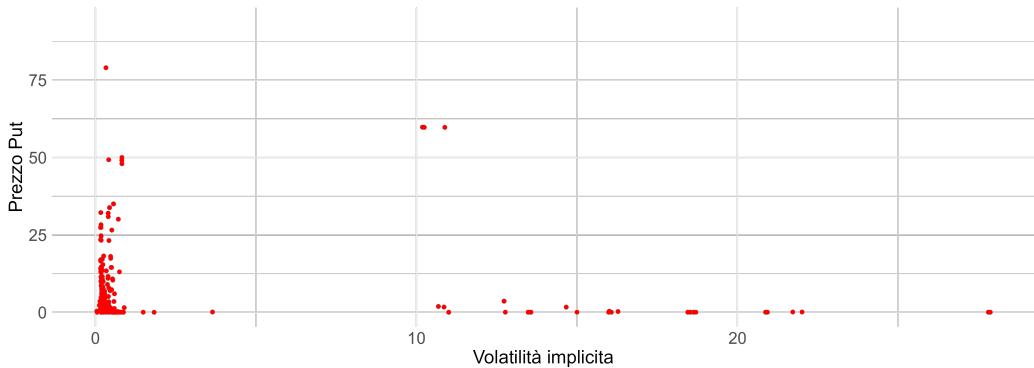


Figura 6.16: Prezzi opzioni Put rispetto alla volatilità implicita teorica

6.4.1 Sintesi Analisi Esplorativa

Tabella 6.3: Sintesi delle caratteristiche e relazioni delle opzioni

Caratteristica	Call	Put
Numero di opzioni	538	432
Prezzo inferiore a \$1	83	133
Prezzo medio	\$135.93	\$118.86
Mediana	\$18.79	\$9.56
Minimo prezzo	\$0.01	\$0.01
Massimo prezzo	\$547.65	\$9415.90
Prezzo 25% quantile	\$4.38	\$0.44
Relazione con giorni alla maturità	Relazione incerta	Relazione incerta
Relazione con prezzo d'esercizio	Negativa	Positiva
Relazione con prezzo del sottostante	Positiva	Negativa
Relazione con volatilità storica	Alta dispersione a bassi livelli, diminuisce con l'aumento della volatilità	Alta dispersione a bassi livelli, andamento parabolico
Relazione con volatilità implicita	Dispersione a bassi livelli	Più concentrato a bassi livelli, andamento parabolico

Le relazioni tra i prezzi delle opzioni e le misure di volatilità storica e implicita sono notevolmente complesse e non lineari. La volatilità storica, calcolata sulla base dei movimenti passati dei prezzi, e la volatilità implicita, che riflette le aspettative del mercato riguardo alla volatilità futura, influenzano i prezzi delle opzioni in modi intricati. L'alta dispersione dei prezzi è indicativa della variabilità nei comportamenti di mercato e delle diverse strategie adottate dagli investitori.

La doppia funzione delle opzioni come strumento di copertura (salvaguardia) e speculazione finanziaria spiega in parte la dispersione nei prezzi. Le opzioni sono utilizzate sia per proteggersi da movimenti avversi del mercato sia per speculare sui movimenti futuri dei prezzi, il che può portare a diverse interpretazioni per la medesima tipologia di opzioni. Infatti, se è immediato associare prezzi più alti a livelli di volatilità più alta per via della funzione speculativa che queste hanno nel contesto della gestione portafoglio; non è da escludere, che il mercato dopo una certa soglia di volatilità (e dunque di rischio) non ripone più fiducia nello strumento.

L'effetto dei giorni alla maturità sui prezzi delle opzioni è caratterizzato da un alto grado di incertezza. Le opzioni possono mostrare comportamenti molto variabili man mano che si avvicinano alla scadenza. Questo fenomeno è spesso collegato alla volatilità implicita e alla variazione del valore temporale delle opzioni. Un aspetto rilevante è che la volatilità storica tende ad aumentare avvicinandosi alla scadenza. Questo incremento della volatilità può essere attribuito all'aumento dell'incertezza e alla speculazione che caratterizzano i giorni finali di vita dell'opzione.

La relazione tra prezzo di scambio e prezzo d'esercizio, invece, tende ad essere piuttosto nitida coerentemente con quella che è la definizione di payoff e di stato di profittevolezza delle opzioni.

Capitolo 7

Analisi Comparativa Modelli di Pricing

Alla luce delle osservazioni e delle riflessioni precedentemente esposte, procederemo ora a un'analisi comparativa dei prezzi delle opzioni utilizzando tre approcci distinti. In particolare, stimeremo i prezzi delle opzioni con il modello di Black-Scholes, il metodo Monte Carlo e il modello di machine learning XGBoost. Successivamente, confronteremo questi modelli sulla base della loro performance previsiva, utilizzando come criterio di valutazione l'errore quadratico medio (RMSE). Questa analisi ci permetterà di valutare l'accuratezza e l'efficacia di ciascun approccio nel prevedere i prezzi delle opzioni, offrendo una visione più approfondita sulle potenzialità e i limiti dei diversi metodi di stima.

7.1 Modello Black-Scholes

In questa sezione, viene implementato il modello di Black-Scholes per stimare i prezzi delle opzioni Call e Put utilizzando tre diverse misure di volatilità storica annuale calcolata su periodi di 1, 2 e 3 anni. Successivamente, si procede alla valutazione delle stime rispetto ai prezzi di scambio osservati e teorici calcolando il Root Mean Square Error (RMSE) per ciascun set di volatilità.

Il modello di Black-Scholes è stato implementato utilizzando l'omonima funzione in R, il cui codice viene riportato in Appendice nella sezione "*Listings Codice R: Modello BLACK SCHOLES*". La funzione prende in input il prezzo del sottostante (S), il prezzo di esercizio (X), il numero di giorni alla scadenza (T_{days}), la volatilità storica annuale (σ), il tasso di interesse privo di rischio (r), e il tipo di opzione ('c' per Call e 'p' per Put).

Abbiamo calcolato i prezzi delle opzioni utilizzando tre diverse misure di volatilità storica annuale e li abbiamo aggiunti al dataset principale come nuove colonne.

Tabella 7.1: RMSE dei Prezzi delle Opzioni utilizzando diverse Volatilità Storiche

Tipo	Sigma	RMSE Prezzi di Scambio	RMSE Prezzi Teorici
Call	Sigma 1	8.486074	4.652387
Call	Sigma 2	10.821921	8.165605
Call	Sigma 3	11.883068	9.529278
Put	Sigma 1	5142.183342	60.818201
Put	Sigma 2	5142.181526	60.661057
Put	Sigma 3	5142.181825	60.680896

Nota: la dicitura Sigma1, Sigma2 e Sigma3 corrisponde alla volatilità storica annuale (σ) calcolata rispettivamente su finestre temporali di 1 anno; 2 anni e 3 anni.

I risultati riportati in tabella 7.1 mostrano il confronto tra gli errori quadrati medi (RMSE) dei prezzi delle opzioni Call e Put calcolati utilizzando le tre diverse volatilità storiche annuali. Sono stati confrontati i RMSE rispetto ai prezzi di mercato osservati e ai prezzi teorici.

Call: Il modello di Black-Scholes utilizzando la prima volatilità storica (Sigma 1) fornisce i risultati migliori per le opzioni Call, con un RMSE di 8.486074 rispetto ai prezzi di mercato e di 4.652387 rispetto ai prezzi teorici. Questo indica che Sigma 1 potrebbe essere la volatilità storica più adatta per stimare il prezzo delle opzioni Call, almeno rispetto alle altre due. Infatti, man mano che si utilizzano volatilità calcolate su periodi più ampi (Sigma 2 e Sigma 3), l'RMSE aumenta, raggiungendo il massimo con Sigma 3. Questo suggerisce che, per le Call, la volatilità storica più alta porta a stime meno accurate, con un aumento dell'errore sia rispetto ai prezzi di scambio sia rispetto ai prezzi teorici. Le considerazioni trovano riscontro grafico nel confronto tra valori stimati ed osservati: figure 7.1 e 7.2. In quest'ultima, si nota che la densità dei prezzi di scambio stimati con volatilità storica annuale calcolata su un anno di rendimenti del sottostante (Sigma1) è quella che meglio approssima i prezzi osservati. Tuttavia, le differenze non sono così prorompenti, tali da definire "giusta" o "sbagliata" una previsione rispetto all'altra. Notiamo, infatti, che il confronto si traduce nello scorgere il livello di accuratezza più

alto per definire la previsione preferibile. In contesti come la finanza, minime differenze possono ripercuotersi su cifre importanti di perdite e profitti per via dell'effetto della leva finanziaria.

Put: Per le opzioni Put, i risultati sono significativamente diversi. Gli RMSE rispetto ai prezzi di mercato sono notevolmente più alti, oltre 5000, indicando una forte discrepanza tra i prezzi stimati dal modello e i prezzi di mercato reali. L'RMSE rispetto ai prezzi teorici è molto più basso, intorno a 60 per tutte e tre le volatilità. Ma ,comunque, più alto rispetto ai valori registrati per i prezzi teorici delle Call. I dati indicano che il modello più accurato risulta essere quello calcolato con la volatilità storica calcolata su due anni di rendimenti del sottostante (Σ^2). Tuttavia, la differenza dell'ordine di grandezza del RMSE rispetto ai valori registrati per le opzioni Call suggerisce che potrebbe esserci un errore sistematico nella stima dei prezzi delle opzioni Put, che non viene catturato correttamente dal modello teorico, causando un aumento consistente dell'RMSE rispetto ai prezzi osservati. L'ausilio grafico della figura 7.3 permette di notare la presenza di un valore stimato oltre i \$10000 in corrispondenza di un valore osservato di circa \$60. Si tratta di una Put scritta sul sottostante MicroStrategy Incorporated, una società che opera nella produzione software. Questo è un caso estremamente particolare. Infatti, in corrispondenza di un prezzo teorico pari a \$103722.94, di un prezzo del sottostante pari a \$1451.74 ed un prezzo d'esercizio \$105175.00, il modello stima un prezzo di scambio dell'opzione pari a \$102517. L'anomalia è che si concretizza una transazione per cui il prezzo di scambio è pari a \$59.75. Questo rappresenta l'unico caso per l'intero dataset per cui si ha una differenza così netta tra prezzi di scambio e prezzi teorici, ma, soprattutto, tra prezzo opzione e prezzo sottostante. I valori relativamente alle rispettive distribuzioni, però, non sono anomali. Soprattutto, il prezzo di scambio rientra nel range di valori della distribuzione. Quindi si tratta di una situazione di mercato particolare, isolata. Per cui, si ripete la procedura di stima al netto di questa osservazione. In tabella 7.2 si riesce ad apprezzare un netto miglioramento dell'accuratezza delle stime, indice dell'influenza della specifica transazione. Rimangono invariate le considerazioni sulla volatilità storica. Infatti, per le opzioni Put, i risultati inferiori di RMSE si hanno in corrispondenza della volatilità storica calcolata su 2 anni (Σ^2)

In figura 7.4 si conferma la concentrazione dei prezzi di scambio delle Put intorno a valori sotto \$1 e l'assenza di differenze tra le stime con diversa volatilità.

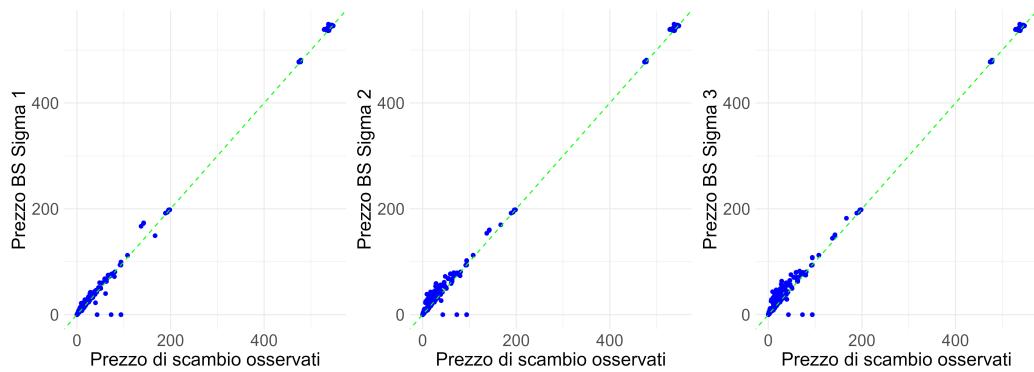


Figura 7.1: Confronto prezzi osservati - prezzi stimati per Call

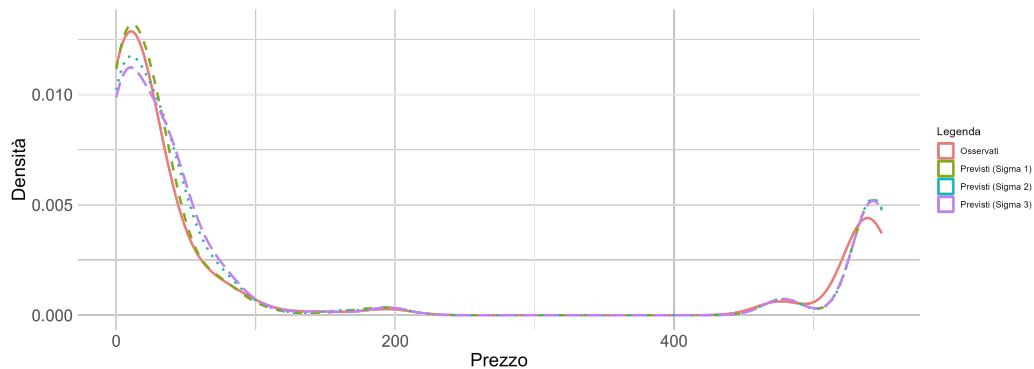


Figura 7.2: Confronto densità prezzi osservati - prezzi stimati Call

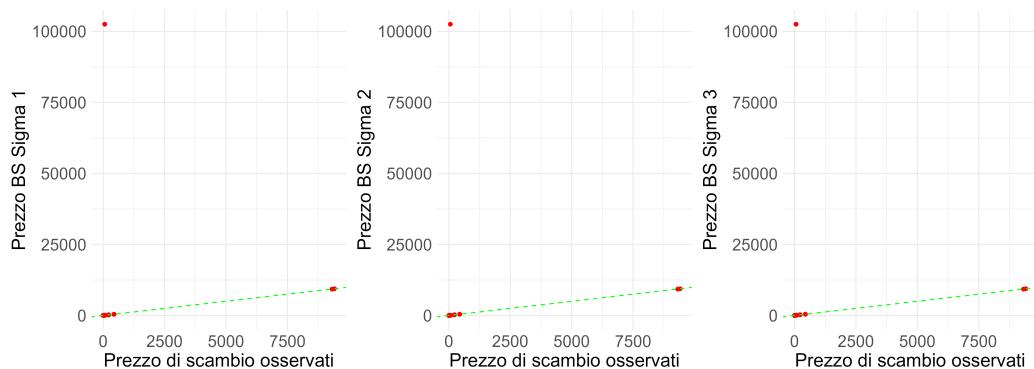


Figura 7.3: Confronto prezzi osservati - prezzi stimati per Put

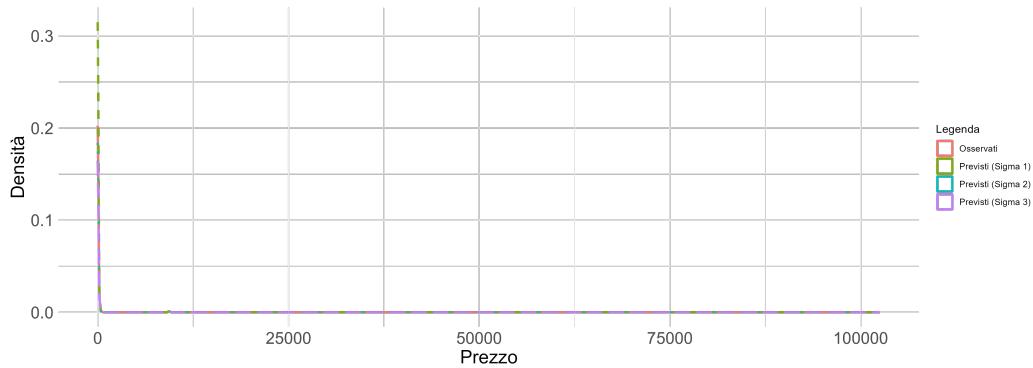


Figura 7.4: Confronto densità prezzi osservati - prezzi stimati Put

Tabella 7.2: RMSE dei Prezzi delle Opzioni utilizzando diverse Volatilità Storiche aggiornata

Tipo	Sigma	RMSE Prezzi di Scambio	RMSE Prezzi Teorici
Call	Sigma 1	8.486074	4.652387
Call	Sigma 2	10.821921	8.165605
Call	Sigma 3	11.883068	9.529278
Put	Sigma 1	5.793401	5.959212
Put	Sigma 2	3.852626	4.046522
Put	Sigma 3	4.233741	4.334481

7.2 Metodo Monte Carlo

In questa sezione, si applica il Metodo Monte Carlo per stimare i prezzi delle opzioni utilizzando le tre diverse misure di volatilità storica annuale. Si prosegue il confronto calcolando il Root Mean Square Error (RMSE) per ciascun set di volatilità.

In appendice, nella sezione *"Listings Codice R: Metodo Montecarlo"*, è presente il codice R relativo all'applicazione del metodo. La funzione prende in input le stesse variabili considerate nel modello Black-Scholes. Occorre precisare che il codice è riferito ai risultati ottenuti con un numero di simulazioni impostato a 50000 ma seguirà il confronto con i risultati ottenuti rispettivamente con 5000 e 10000 simulazioni.

Tabella 7.3: RMSE dei Prezzi delle Opzioni utilizzando diverse Volatilità Storiche con 5000 simulazioni

Tipologia	Sigma	RMSE Prezzi di Scambio	RMSE Prezzi Teorici
Call	Sigma 1	8.507745	4.690424
Call	Sigma 2	10.851639	8.208169
Call	Sigma 3	11.868192	9.515948
Put	Sigma 1	5141.865833	61.139816
Put	Sigma 2	5141.739771	61.099854
Put	Sigma 3	5141.971514	60.893463

Tabella 7.4: RMSE dei Prezzi delle Opzioni utilizzando diverse Volatilità Storiche con 10000 simulazioni

Tipologia	Sigma	RMSE Prezzi di Scambio	RMSE Prezzi Teorici
Call	Sigma 1	8.537991	4.740874
Call	Sigma 2	10.795442	8.136496
Call	Sigma 3	11.866619	9.510944
Put	Sigma 1	5142.532465	60.470324
Put	Sigma 2	5141.419413	61.427437
Put	Sigma 3	5141.978449	60.883311

Tabella 7.5: RMSE dei Prezzi delle Opzioni utilizzando diverse Volatilità Storiche con 50000 simulazioni

Tipologia	Sigma	RMSE Prezzi di Scambio	RMSE Prezzi Teorici
Call	Sigma 1	8.489399	4.660243
Call	Sigma 2	10.840557	8.187646
Call	Sigma 3	11.881579	9.526007
Put	Sigma 1	5142.364774	60.640273
Put	Sigma 2	5142.033689	60.807891
Put	Sigma 3	5142.326137	60.539544

Precisione e Stabilità I valori di RMSE per i prezzi di scambio e teorici, per entrambe le tipologie di opzioni, tendono a stabilizzarsi con l'aumento del numero di simulazioni, suggerendo che un numero maggiore di simulazioni riduce la variabilità dei risultati e migliora l'affidabilità delle stime.

Call: Le opzioni Call mostrano un RMSE per entrambi i prezzi molto più contenuto rispetto ai risultati ottenuti per le opzioni Put. Inoltre, per i prezzi

di scambio, questo è superiore rispetto ai prezzi teorici. Questo risultato indica che il modello di simulazione si adatta meglio ai dati sui prezzi teorici delle opzioni di tipologia Call rispetto ai prezzi di scambio effettivi. La migliore previsione per la tipologia Call è riconducibile alla combinazione Sigma1 - numero simulazioni 50000. Questo vale sia per i prezzi di scambio che per i prezzi teorici. Graficamente, figure 7.5 e 7.6, si rappresenta il confronto tra prezzi di scambio osservati e previsti. Le differenze delle stime sono minime rispetto alle tre diverse volatilità storiche. Ma in valore assoluto, le stime con Sigma1 sono le più accurate.

Put: I valori di RMSE per i prezzi di scambio sono molto elevati e simili tra i diversi valori di volatilità storica (Sigma), suggerendo che il modello di simulazione Monte Carlo potrebbe avere difficoltà a catturare correttamente il prezzo di scambio delle opzioni Put. I valori di RMSE per i prezzi teorici sono molto inferiori e non presentano considerevoli differenze al variare di Sigma. Tuttavia, questi sono molto maggiori rispetto ai valori riscontrati sulle Call. L'analisi grafica delle figure 7.7 e 7.8 conducono alle medesime riflessioni già effettuate sulle stime dei prezzi delle opzioni Put, ottenute con il Modello Black Scholes. Relativamente al confronto rispetto alle tre diverse volatilità storiche non si apprezzano differenze significative. Piuttosto, ci si interroga circa l'influenza di quel valore oltre i \$100000. Quindi, si ripete la procedura con 50000 simulazioni, al netto della transazione avvenuta per l'opzione scritta su MicroStrategy Incorporated.

In tabella 7.6, si riporta il confronto delle stime per le opzioni Put aggiornato senza la suddetta transazione. Si osserva un netto miglioramento della performance predittiva, benché le considerazioni circa l'influenza della volatilità storica restino simili. Infatti, la migliore previsione si ha in corrispondenza di Sigma2, sia per i prezzi di scambio che per i prezzi teorici.

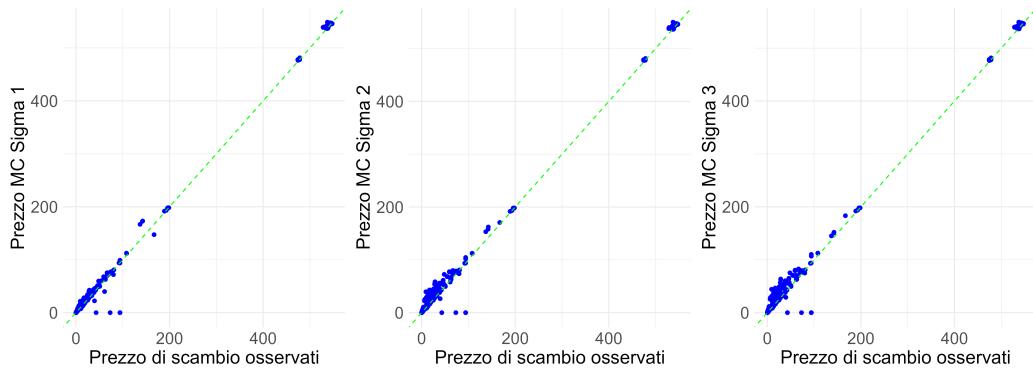


Figura 7.5: Confronto prezzi osservati e prezzi previsti Call con Metodo Monte Carlo nsim=50000

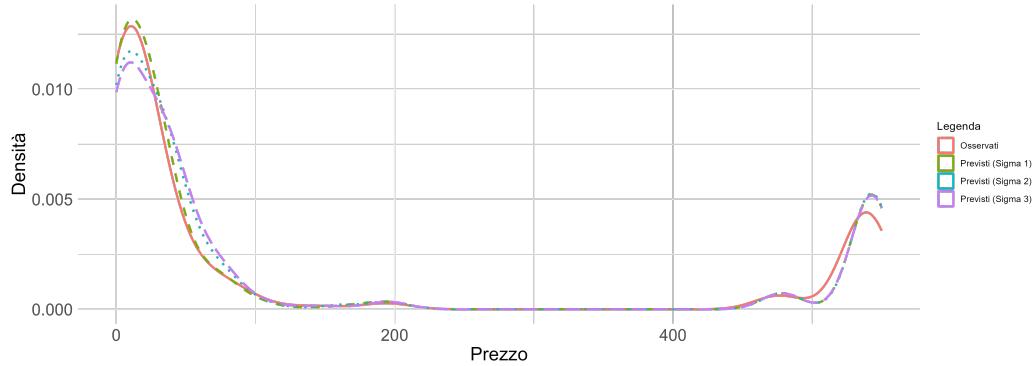


Figura 7.6: Confronto densità prezzi osservati e prezzi previsti Call com Metodo Monte Carlo nsim=50000

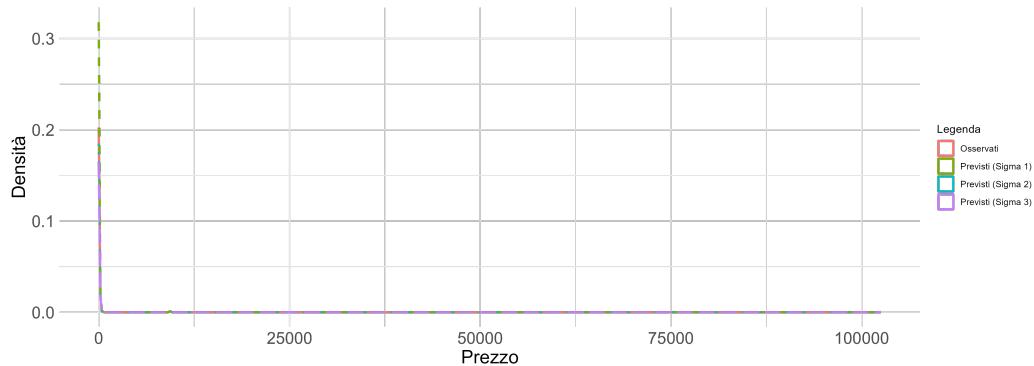


Figura 7.8: Confronto densità prezzi osservati e prezzi previsti Put com Metodo Monte Carlo nsim=50000

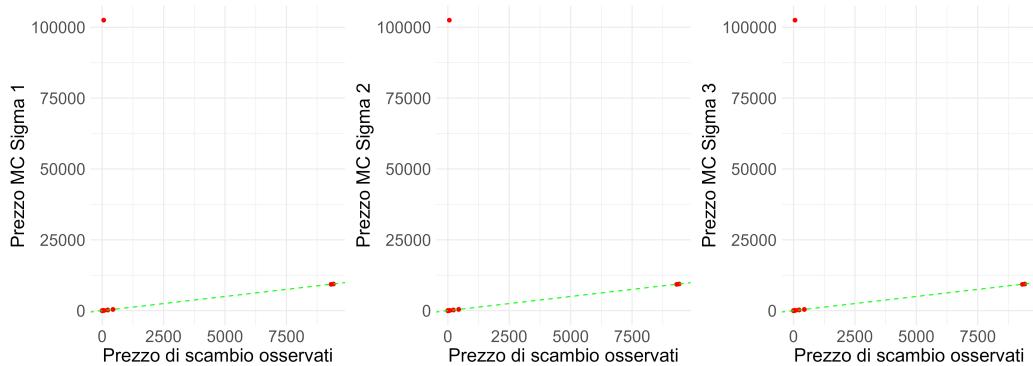


Figura 7.7: Confronto prezzi osservati e prezzi previsti Put con Metodo Monte Carlo nsim=50000

Tabella 7.6: RMSE su Prezzi Empirici e Teorici per Diversi Valori di Sigma con Monte Carlo per Call e Put aggiornata

Tipologia	Sigma	RMSE Prezzi di Scambio	RMSE Prezzi Teorici
Call	Sigma 1	8.489399	4.660243
Call	Sigma 2	10.840557	8.187646
Call	Sigma 3	11.881579	9.526007
Put	Sigma 1	5.820740	5.985682
Put	Sigma 2	3.830828	4.036580
Put	Sigma 3	4.269459	4.370194

7.3 XGBoost

In questa sezione, riportiamo il confronto dei valori di RMSE quali risultati ottenuti utilizzando l'algoritmo di machine learning XGBOOST per diverse tipologie di opzioni (Call e Put) e prezzi (teorici e di scambio). Sono state utilizzate tre diverse misure di volatilità storica, calcolate su 1 anno (Sigma1), 2 anni (Sigma2) 3 anni (Sigma3), della serie storica del rendimento del titolo sottostante. La tabella seguente riassume i risultati con le Sigma ordinate in modo crescente.

Tabella 7.7: Confronto dei valori di RMSE per opzioni Call e Put con diverse volatilità storiche.

Tipologia	Prezzi	Sigma 1	Sigma 2	Sigma 3
Call	Teorici	6.88184	7.10407	6.88629
Call	Scambio	3.43618	3.43640	3.44714
Put	Teorici	1443.24967	1443.30648	1443.21717
Put	Scambio	8.77617	8.71587	9.07329

Come si può osservare in tabella 7.7, gli errori per le opzioni Call e Put sui prezzi teorici sono più elevati rispetto a quelli sui prezzi di scambio. Questo suggerisce che il modello ha una maggiore difficoltà a predire i prezzi teorici rispetto a quelli osservati. La performance sui prezzi teorici per le Call è significativamente migliore rispetto a quella per le Put. Questo pattern è costante su tutte le misure di volatilità. L'RMSE varia leggermente con l'aumentare dell'orizzonte temporale della volatilità storica. Tuttavia, non sembra esserci un cambiamento drastico, suggerendo che la scelta tra volatilità a 1 anno, 2 anni o 3 anni potrebbe non essere un fattore critico nella performance del modello. L'uso dei prezzi di scambio e un periodo di volatilità più breve sembrano generalmente migliorare la precisione delle previsioni (minore RMSE).

Call: l'RMSE più basso per le Call si ottiene nello scenario con Sigma1. Questo è pari 6.88184 per i prezzi teorici, mentre sui prezzi di scambio è pari a 3.43618. Le previsioni sui prezzi teorici includono più variabili (specificamente `underlying_ask`, `trade_size`, `trade_iv`) rispetto a quelle sui prezzi di scambio.

- **Prezzi Teorici:** *Features selezionate: 14 variabili*
`trade_delta`, `trade_gamma`, `trade_vega`, `trade_theta`, `trade_rho`, `theo_iv`,

strike, n_giorni, option_type, volatilita_annuale_1, underlying_bid, underlying_ask, trade_size, trade_iv

- **Prezzi di Scambio:** *Features selezionate: 11 variabili*
 trade_delta, trade_gamma, trade_vega, trade_theta, trade_rho, theo_iv, strike, n_giorni, option_type, volatilita_annuale_1, underlying_bid

Differenze:

- Nella previsione dei prezzi di scambio mancano le variabili `underlying_ask`, `trade_size`, e `trade_iv` rispetto ai prezzi teorici. In questa l'underlying bid è la variabile più importante. Seguita dalla greca trade vega, dalla volatilità annuale 1 (Sigma1) e dalla greca trade theta.

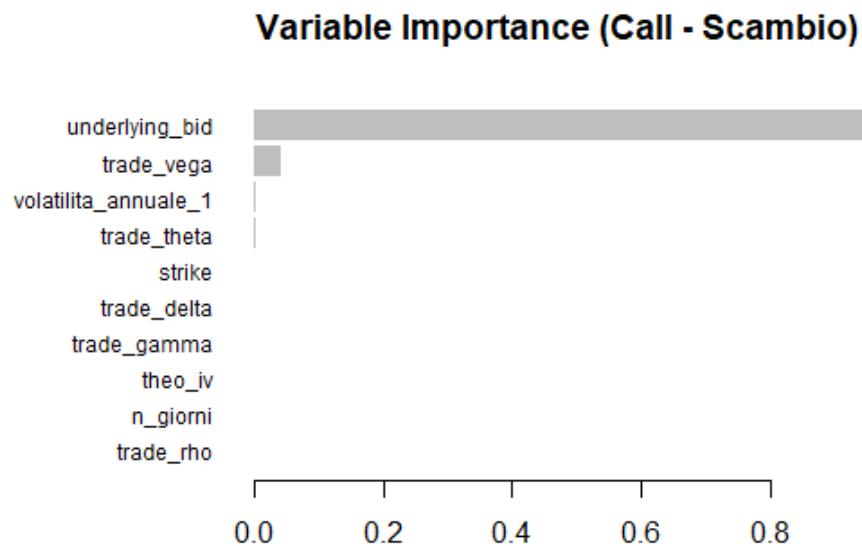


Figura 7.9: Grafico Importanza Variabili Call

Put: I valori di RMSE per i prezzi teorici delle opzioni Put sono significativamente più alti rispetto ai prezzi di scambio. Per questi il miglior risultato si ha in corrispondenza dei prezzi calcolati tenendo conto della volatilità storica riferita a due anni (σ_2) pari a 8.71587. Invece, per i prezzi teorici il

valore minimo è in corrispondenza di Sigma3, pari a 1443.21717.

In riferimento all'importanza delle variabili selezionate, la differenza principale è l'inclusione della variabile `underlying_ask` nelle previsioni dei prezzi di scambio, che non appare nelle previsioni sui prezzi teorici.

- **Prezzi Teorici:** *Features selezionate: 12 variabili* trade_delta, trade_gamma, trade_vega, trade_theta, trade_rho, theo_iv, strike, n_giorni, option_type, volatilita_annuale_1, open_interest, trade_iv
- **Prezzi di Scambio:** *Features selezionate: 13 variabili* trade_delta, trade_gamma, trade_vega, trade_theta, trade_rho, theo_iv, strike, n_giorni, option_type, volatilita_annuale_1, underlying_ask, trade_iv, open_interest

Differenze:

- Nella previsione dei prezzi di scambio è presente la variabile `underlying_ask`, che non è utilizzata per la previsione dei prezzi teorici. L'`underlying ask`, cioè il prezzo più basso a cui un venditore è disposto a vendere l'attività sottostante in un mercato finanziario, risulta essere la variabile più importante. Seguita dal prezzo d'esercizio e dalla greca `trade vega`.

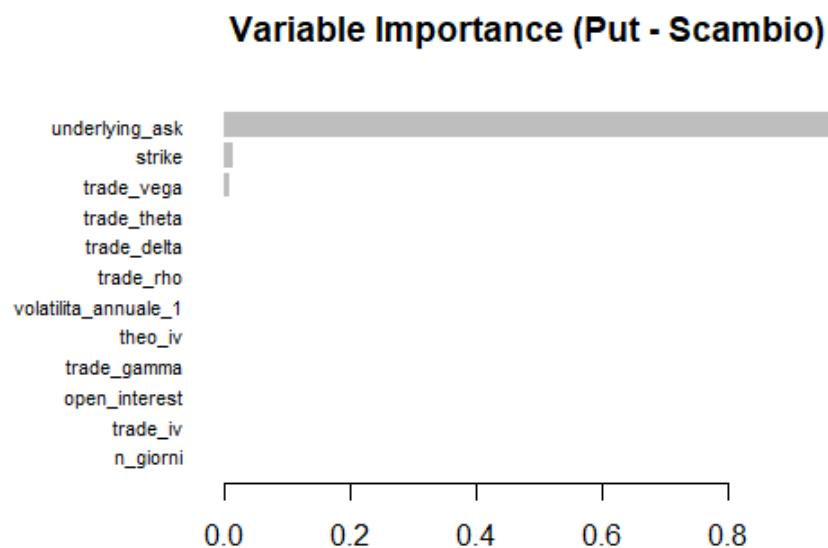


Figura 7.10: Grafico Importanza Variabili Put

Sintesi:

Queste differenze suggeriscono che le variabili legate ai prezzi di mercato (come `underlying_ask` e `underlying_bid`) sono più rilevanti nelle previsioni dei prezzi di scambio rispetto a quelle dei prezzi teorici.

Il modello XGBoost sembra essere relativamente robusto per le previsioni sui prezzi di scambio senza considerazioni a priori sui casi particolari di mercato, con variazioni limitate del RMSE al variare della volatilità storica considerata (Σ). Infatti, sia per le Call e sia per le Put la previsione dei prezzi di scambio risulta la migliore in prima battuta. A questo si aggiunge la possibilità di poter estrapolare un’ulteriore informazione: l’importanza delle variabili nel determinare la stima dei prezzi.

La previsione dei prezzi teorici delle opzioni Put sembra essere problematica, con valori di RMSE molto alti che potrebbero indicare un problema con il modello o con i dati. Il fatto che il RMSE per i prezzi teorici delle opzioni Put sia significativamente più alto rispetto ai prezzi di scambio potrebbe indicare che il modello ha imparato molto bene le specificità dei dati di training, ma non riesce a generalizzare ai dati di test. In realtà, solo grazie al confronto dei due approcci precedenti riusciamo a notare il caso di mercato specifico riferito all’opzione Put scritta sull’azione sottostante MicroStrategy Incorporated. Pertanto, si è ripetuta la procedura di stima al netto di questa transazione, ottenendo un RMSE 16.20847 sui prezzi teorici calcolati con la volatilità storica su 1 anno di rendimenti (Σ_1); RMSE pari a 16.249170 su prezzi teorici calcolati con volatilità storica riferita a due anni (Σ_2) e RMSE pari a 16.242458 su prezzi teorici calcolati con volatilità storica riferita a 3 anni (Σ_3). Sicuramente un ordine di grandezza più coerente con quello delle altre stime, ma ben lontano dalla performance sulla stima dei prezzi teorici dei due modelli applicati precedentemente.

Il miglioramento della performance predittiva per i prezzi di scambio senza considerazioni a priori circa la transazione ”problematica”, soprattutto per la tipologia Put, è apprezzabile graficamente dal confronto delle densità delle stime dei prezzi di scambio (figure 7.11 e 7.13) e dal confronto tra i valori osservati e i valori previsti per i prezzi di scambio (figure 7.12 e 7.14).

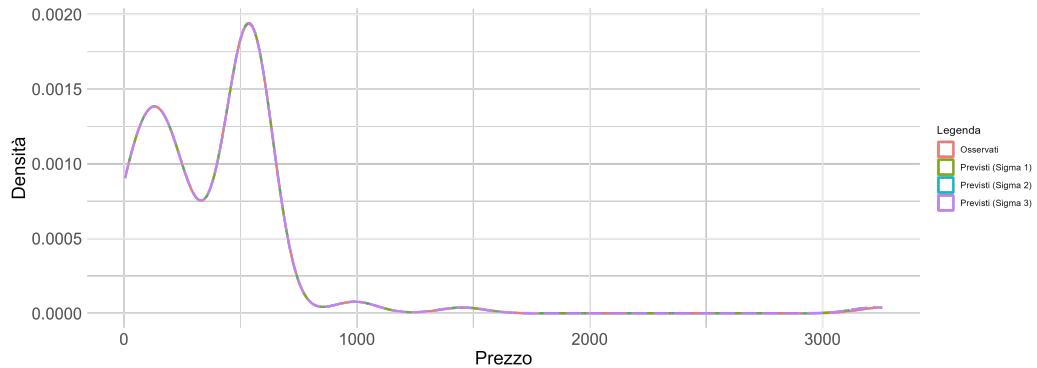


Figura 7.11: Density Plot XGBoost Prezzi di Scambio Call

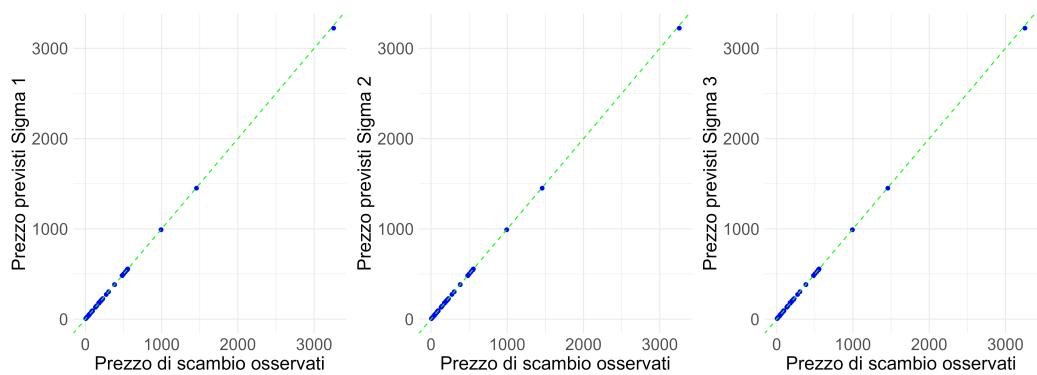


Figura 7.12: Confronto prezzi osservati e previsti XGB - Call

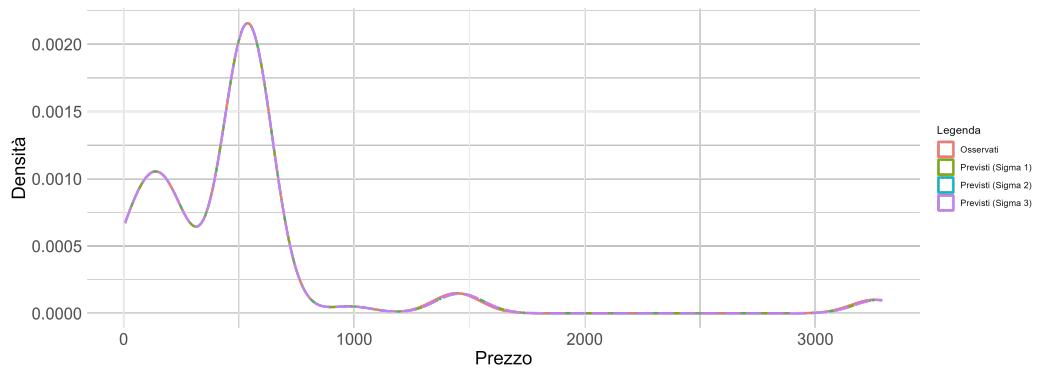


Figura 7.13: Density Plot XGBoost Prezzi di Scambio Put

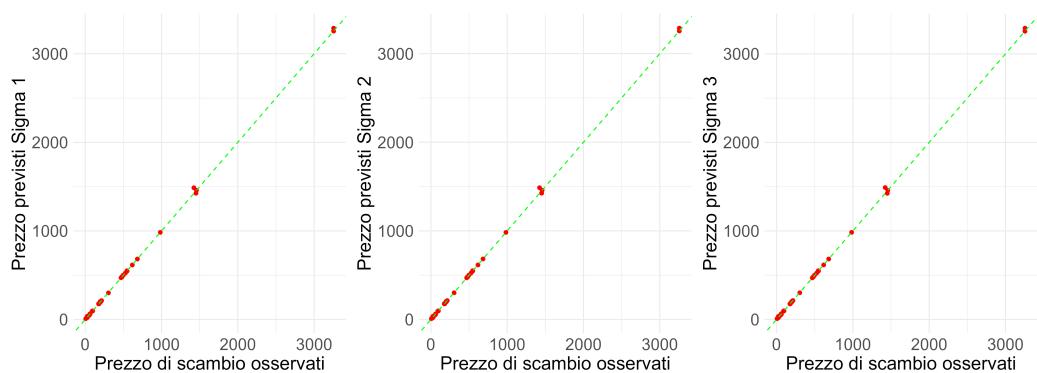


Figura 7.14: Confronto prezzi osservati e previsti XGB - Put

Ottimizzazione del Modello con Random Grid Search .

Descriviamo i passaggi eseguiti per ottimizzare il modello XGBoost utilizzando la strategia di Random Grid Search combinata alla 5 folds Cross validation, con l'obiettivo di ridurre l'overfitting e migliorare la generalizzazione del modello.

Il primo step riguarda la creazione di una griglia di iperparametri:

- `max_depth`: Profondità massima degli alberi.
- `eta`: Tasso di apprendimento.
- `gamma`: Parametro di regolarizzazione per la riduzione delle foglie.
- `colsample_bytree`: Percentuale di colonne da campionare per ogni albero.
- `min_child_weight`: Peso minimo per una foglia dell'albero.
- `sub_sample`: Percentuale di dati da campionare per ogni albero.

Questa griglia ha permesso di esplorare diverse combinazioni di parametri per trovare quella ottimale in grado di offrire il miglior compromesso tra RMSE e R^2 . Durante questo processo, è stata implementata la 5 folds cross-validation per valutare le prestazioni del modello per ciascuna combinazione di iperparametri. Questo approccio ha permesso di identificare i migliori iperparametri per addestrare il modello XGBoost finale. Per evitare l'overfitting durante l'addestramento, è stato utilizzato l'*early stopping* (interruzione precoce) con lo scopo di fermare l'addestramento quando il RMSE sui dati di validazione non migliorava dopo 10 iterazioni.

I parametri ottimali ottenuti dalla procedura ed utilizzati per il modello finale sono i seguenti:

- `objective = "reg:squarederror"`
- `eval_metric = "rmse"`
- `max_depth = 5`
- `eta = 0.2`
- `gamma = 2`
- `colsample_bytree = 1`
- `min_child_weight = 1`

- `subsample` = 0.8

Una volta addestrato il modello finale, ho calcolato il RMSE per le previsioni sui dati di test, separatamente per le opzioni call e put, sia per i prezzi di scambio che per i prezzi teorici. Infine, ho ripetuto la procedura includendo la volatilità annuale considerata ed escludendo le altre due da non considerare. Si noti come a seguito del caricamento dei dati, sono state rese numeriche tutte le variabili che non lo fossero. Infatti, la funzione `xgb.DMatrix` non accetta altro tipo. Per cui, sono state necessarie le seguenti operazioni:

- Conversione della variabile `option_type` in formato binario, dove 1 rappresenta le opzioni put e 0 le opzioni call.
- Conversione della variabile `delivery` in formato binario, dove 1 rappresenta la consegna *cash* e 0 *physical*.
- Conversione della variabile `n_giorni` in formato intero.
- Rimozione di colonne non rilevanti per l'analisi (`underlying_symbol`, `expiration`, `exercise`, `root`, `trade_condition_id`, `trade_datetime`, `settlement`, `multiplier`, `exchange_sequence_number`, `exchange_id`, `volatilita_annuale_2`, `volatilita_annuale_3`).

I dati sono stati quindi suddivisi in due insiemi separati: `X_call` e `X_put`, contenenti rispettivamente i dati relativi alle opzioni call e put.

Successivamente, vengono definite le variabili target per le previsioni dei prezzi di scambio e teorici:

- **Prezzi di scambio:** `Y_call_scambio` e `Y_put_scambio` rappresentano rispettivamente i prezzi di scambio per le opzioni call e put.
- **Prezzi teorici:** `Y_call_theo` e `Y_put_theo` rappresentano rispettivamente i prezzi teorici per le opzioni call e put.

Le colonne target sono state rimosse dai rispettivi dataset `X_call` e `X_put` per creare i dataset `X_call_scambio`, `X_put_scambio`, `X_call_theo`, e `X_put_theo`.

Le variabili sono state suddivise in due categorie:

Variabili opzionali: `delivery`, `open_interest`, `underlying_bid`,
`underlying_ask`, `trade_size`, `trade_iv`.

Variabili non opzionali: `trade_delta`, `trade_gamma`, `trade_vega`, `trade_theta`,
`trade_rho`, `theo_iv`, `strike`, `n_giorni`, `option_type`, `volatilita_annuale_1`.

Il codice relativo all'implementazione del modello è completamente esposto in Appendice, sezione *Listings Codice R: XGBoost*

Capitolo 8

Conclusioni

8.1 Introduzione ai risultati

L'analisi dei risultati riportati in tabella 8.1 offre una panoramica approfondita sulle prestazioni dei modelli di pricing delle opzioni Call e Put, con particolare attenzione alla scelta della volatilità storica. I dati evidenziano come i modelli considerati reagiscano in modo diverso alle variazioni della volatilità storica, con implicazioni per l'accuratezza delle previsioni dei prezzi delle opzioni.

Per le opzioni Call, la volatilità storica calcolata su un anno (Sigma 1) ha prodotto le stime più accurate sia rispetto ai prezzi di mercato osservati che a quelli teorici.

In contrasto, i risultati sui prezzi teorici delle opzioni Put mostrano una discrepanza significativa rispetto alle Call, suggerendo un disallineamento tra i prezzi stimati dal modello e quelli reali. Secondo una prima analisi, XGBoost è il miglior modello per la previsione dei prezzi di scambio di questa tipologia di opzioni. Inoltre, volatilità storiche calcolate su archi temporali più ampi portano a stime più precise dei prezzi delle Put.

Tabella 8.1: Combinazioni Ottimali di Modelli e Volatilità per i Prezzi delle Opzioni

Tipo di Opzione	Tipo di Prezzo	Sigma	Miglior Modello	RMSE
Call	Prezzi Teorici	1	Black-Scholes	4.652387
Call	Prezzi di Scambio	1	XGBoost	3.43618
Put	Prezzi Teorici	3	Monte Carlo	60.539544
Put	Prezzi di Scambio	2	XGBoost	8.71587

Un elemento centrale emerso dall’analisi è la presenza di un caso particolare di mercato, non semplicemente un outlier, che ha significativamente influenzato le stime. Questo caso, relativo a un’opzione Put scritta su azioni MicroStrategy Incorporated, ha mostrato come circostanze di mercato specifiche possano portare a notevoli discrepanze tra prezzi stimati e osservati, senza indicare un errore sistematico nei dati, ma piuttosto una situazione di mercato unica che richiede un’analisi più dettagliata. Infatti, ripetendo l’analisi comparativa al netto di questo caso specifico, cambia la valutazione relativa alla determinazione dei prezzi di scambio e teorici delle opzioni Put. In tabella 8.2 sono riportati i risultati aggiornati. Si nota, come la rimozione della transazione specifica favorisca la selezione del Metodo Monte Carlo quale miglior modello per la previsione delle opzioni Put. Sia per i prezzi di scambio, che per quelli teorici.

Tabella 8.2: Combinazioni Ottimali di Modelli e Volatilità per i Prezzi delle Opzioni senza la Put scritta su MicroStrategy Incorporated

Tipo di Opzione	Tipo di Prezzo	Sigma	Miglior Modello	RMSE
Call	Prezzi Teorici	1	Black-Scholes	4.652387
Call	Prezzi di Scambio	1	XGBoost	3.43618
Put	Prezzi Teorici	3	Monte Carlo	3.830828
Put	Prezzi di Scambio	2	MonteCarlo	4.036580

8.2 Modello di Black-Scholes

Opzioni Call: Il modello di Black-Scholes, utilizzando la volatilità storica di un anno (Sigma 1), dimostra la migliore performance con valori di RMSE più bassi per predire i prezzi teorici delle Call. Questo suggerisce che Sigma 1 è la scelta preferibile per le previsioni delle Call.

Opzioni Put: Il modello di Black-Scholes presenta valori di RMSE molto elevati per i prezzi di mercato delle Put, indicando una chiara difficoltà nel prevedere accuratamente questi prezzi. Sebbene le previsioni basate sui prezzi teorici siano relativamente migliori, esse non raggiungono ancora livelli ottimali. Questo ha suggerito che le Put sono particolarmente sensibili a specifici scenari di mercato non catturati dalla volatilità storica. Infatti, il vantaggio assoluto di questo modello si è concretizzato nella possibilità di riuscire ad inquadrare una situazione “limite”, rispetto alla quale è stata ripetuta la procedura di stima, al netto della stessa, per poi valutarne nuovamente i risultati.

8.3 Metodo Monte Carlo

Opzioni Call: L'uso del metodo Monte Carlo mostra miglioramenti dell'RMSE con l'aumentare del numero di simulazioni. Con 50.000 simulazioni, si ottengono migliori previsioni, se confrontate con i valori ottenuti con 5000 e 10000 simulazioni. Tuttavia, i risultati per la previsione dei prezzi delle opzioni Call vengono superati dagli altri approcci. Vale sia per i prezzi teorici sia per i prezzi di scambio.

Opzioni Put: Per le Put, inizialmente l'RMSE per i prezzi di scambio rimane molto elevato e simile tra le diverse volatilità storiche, suggerendo che il metodo Monte Carlo riscontra difficoltà a catturare accuratamente i prezzi di scambio delle Put. Tuttavia, alla stregua del modello Black Scholes, si rivela molto performante per la stima dei prezzi teorici. A seguito della rimozione della transazione problematica, con questo metodo si perviene alle migliori stime per i prezzi delle opzioni Put.

8.4 Modello di Machine Learning XGBoost

In questo contesto, il modello XGBoost si è rivelato particolarmente efficace nella previsione dei prezzi di scambio, anche in presenza di situazioni di mercato particolari. La sua robustezza consente di ottenere stime affidabili e, soprattutto, fornisce informazioni preziose sulle variabili più influenti nella determinazione dei prezzi. Tuttavia, il modello XGBoost mostra carenze nella determinazione dei prezzi teorici, specialmente in presenza di casi particolari di mercato. Infatti, mentre il modello XGBoost rappresenta un valido strumento per la previsione dei prezzi di scambio, la sua efficacia nella stima dei prezzi teorici è limitata, evidenziando la necessità di una maggiore attenzione nell'interpretazione dei risultati, soprattutto in contesti di mercato complessi o atipici.

Opzioni Call: XGBoost mostra una performance robusta nella previsione dei prezzi di scambio delle Call. La migliore dei tre approcci. Le variazioni dell'RMSE al variare della volatilità storica, suggeriscono che la performance ottima corrisponde a Sigma 1.

Opzioni Put: Il modello XGBoost ha difficoltà significative nella previsione dei prezzi teorici delle Put, con RMSE molto elevati. Viceversa, sui prezzi di scambio riesce a pervenire alla migliore prestazione. Interessante è notare come per le opzioni Put, i risultati ottimali sui prezzi di scambio si hanno in corrispondenza della volatilità calcolata su 2 anni di rendimenti; mentre, sui prezzi teorici addirittura su 3 anni.

Importanza delle Variabili nel Modello XGBoost

Call Per i prezzi di scambio, le variabili più importanti sono l'underlying bid, la volatilità storica calcolata su 1 anno e le greche (trade vega e trade theta). Per i prezzi teorici, variabili aggiuntive come l'underlying ask e il trade size sono rilevanti. Questo suggerisce che le variabili legate ai prezzi di mercato sono fondamentali per ottenere previsioni più accurate.

Put Per i prezzi di scambio delle Put, l'underlying ask è cruciale. Per i prezzi teorici, l'inclusione di variabili come l'open interest e il trade size non migliora significativamente le previsioni rispetto ai prezzi di scambio.

8.5 Sintesi

L'analisi condotta sui modelli di Black-Scholes, Monte Carlo e XGBoost evidenzia come la scelta del modello e della volatilità storica influiscano sulla precisione delle previsioni dei prezzi delle opzioni Call e Put. In particolare, il modello di Black-Scholes si è dimostrato particolarmente efficace nel prevedere i prezzi teorici delle Call, soprattutto utilizzando una volatilità storica calcolata su un anno (Sigma 1). Questo suggerisce che nel caso di dati storici limitati, l'uso di finestre più brevi può essere migliore per la previsione dei prezzi delle opzioni Call. Inoltre, il metodo Monte Carlo due modelli teorici (Black Schoels e Monte Carlo) hanno permesso di inquadrare un caso di mercato "particolare" che comprometteva le conclusioni sulle stime dei prezzi teorici.

Il metodo Monte Carlo, mostra miglioramenti con l'aumento del numero di simulazioni. Per le Put, è il miglior Metodo per la stima sia dei prezzi teorici che dei prezzi di scambio.

Il modello XGBoost, invece, emerge come il più robusto per la previsione dei prezzi di scambio delle Call con Sigma 1 evidenziando l'importanza delle variabili di mercato, come l'underlying bid per le Call e l'underlying ask per le Put.

Questa analisi mette in luce come variazioni anche minime nelle ipotesi modellistiche o nella scelta delle variabili possano portare a differenze sostanziali nei risultati, un aspetto cruciale in finanza, dove piccoli margini di errore possono avere impatti significativi su decisioni di investimento e gestione del rischio. La riflessione si estende oltre la gestione del rischio di portafoglio, poiché le tecniche di pricing delle opzioni trovano applicazione anche nel calcolo della probabilità di default (Probability of Default, PD) in contesti di credit risk. La capacità di modellare correttamente la distribuzione dei prezzi di mercato e di identificare i fattori critici di rischio può essere decisiva non

solo per la gestione di portafogli titoli, ma anche per la valutazione del rischio di credito, dove un'accurata stima della PD è fondamentale per la solidità e la sostenibilità finanziaria.

8.6 Estensione dell'Analisi a un Contesto Out of Sample

L'analisi presentata in questo studio si basa su dati di una singola giornata di transazioni, fornendo importanti indicazioni sull'efficacia dei modelli di pricing delle opzioni, come Black-Scholes, Monte Carlo e XGBoost, in relazione a diverse volatilità storiche. Tuttavia, per migliorare la robustezza e l'affidabilità delle conclusioni, sarebbe opportuno estendere questa analisi a un contesto *out of sample*, utilizzando serie storiche annuali sui prezzi delle opzioni. Un'analisi *out of sample* permetterebbe di valutare la capacità predittiva dei modelli su periodi di tempo più estesi e variabili, testando la loro performance in condizioni di mercato diverse, che includono eventi significativi, variazioni della volatilità e cambiamenti nelle dinamiche del sottostante. Questo approccio potrebbe consentire di osservare se i modelli mantengono una performance costante nel tempo o se mostrano segni di deterioramento in specifici periodi o condizioni di mercato. Inoltre, con un periodo di osservazione più lungo, sarebbe possibile identificare e studiare meglio eventuali situazioni di mercato particolari, simili a quelle osservate nel nostro studio con la Put su MicroStrategy Incorporated. Tale analisi permetterebbe di affinare ulteriormente i modelli o introdurre meccanismi per gestire queste situazioni in modo più efficace. L'utilizzo di serie storiche annuali consentirebbe anche di confrontare continuamente le prestazioni delle diverse volatilità storiche (Σ_1 , Σ_2 , Σ_3) e di verificare se le osservazioni fatte in questo studio rimangono valide nel tempo.

Infine, estendere l'analisi a un contesto *out of sample* migliorerebbe la comprensione dell'efficacia dei modelli di pricing, fornendo una base più solida per il loro utilizzo in ambito pratico. Tale estensione permetterebbe di rafforzare la validità delle conclusioni tratte, dimostrando la capacità dei modelli di adattarsi alle dinamiche reali del mercato finanziario.

Bibliografia

- [1] F. Black, M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ., 81 (3) (1973), pp. 637-654.
- [2] Boyle, P., *Options: A Monte Carlo Approach*, Journal of Financial Economics, Vol. 4, pp. 323-338, 1977.
- [3] Kahneman, D., & Tversky, A. (1979). *Prospect Theory: An Analysis of Decision under Risk*. Econometrica, 47(2), 263-291.
- [4] Thaler, R. H. (1980). *Toward a Positive Theory of Consumer Choice*. Journal of Economic Behavior & Organization, 1(1), 39-60.
- [5] S.L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Rev. Financ. Stud., 6 (2) (1993), pp. 327-343.
- [6] Boyle, P., Broadie, M., Glasserman, P., *Monte Carlo Methods for Security Pricing*, Journal of Economic Dynamics and Control, Vol. 2, pp. 1267-1321, 1997.
- [7] G. Bakshi, C. Cao, Z. Chen, *Empirical performance of alternative option pricing models*, J. Finance, 52 (5) (1997), pp. 2003-2049.
- [8] Breunig, Markus M. and Kriegel, Hans-Peter and Ng, Raymond T. and Sander, Jörg *LOF: Identifying Density-Based Local Outliers* , (2000) ,pp 93-104,
- [9] Wilmott, P., *Introduces Quantitative Finance*, Wiley, Chichester, 2007.
- [10] Hull, J. C., *Opzioni, futures e altri derivati*, Pearson Italia, Milano, 2015.
- [11] Y.C. Chang, K.H. Chang, G.J. Wu, *Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions*, Appl. Soft Comput., 73 (2018), pp. 914-920.

- [12] C.F. Ivaşcu, *Option pricing using machine learning*, Expert Syst. Appl., 163 (2021), Article 113799.
- [13] Pham Hoang Vuong, Trinh Tan, *Stock-Price Forecasting Based on XG-Boost and LSTM* Information Science Faculty, Sai Gon University, Ho Chi Minh City, 700000, Vietnam, 04 May 2021
- [14] Q. Luo, Z. Jia, H. Li, Y. Wu, *Analysis of parametric and non-parametric option pricing models*, School of Mathematics, Hefei University of Technology, Hefei 230009, PR China, Version of Record 4 November 2022.
- [15] Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J., *Dive into Deep Learning*, Release 1.0.0-alpha0, July 16, 2022.

Appendice

.1 Listings Codice R: Importazione e processamento dati

Listing 1: Selezione dei simboli unici

```
1 simboli <- unique(dati$underlying_symbol) %>%
2   str_subset(., pattern = "\\"^", negate = TRUE)
```

Listing 2: Filtraggio e preprocessing dei dati

```
1 dati_rid <- dati %>%
2   filter(underlying_symbol %in% simboli) %>%
3   mutate(expiration = as.Date(expiration,
4     format = "%d/%m/%Y")) %>%
5   mutate(trade_datetime = as.Date(trade_datetime)) %>%
6   mutate(n_giorni = expiration - trade_datetime) %>%
7   filter(exercise == "E")
```

Listing 3: Definizione delle finestre temporali

```
1 date_ranges <- list(
2   c(as.Date("2023-06-20"), as.Date("2024-06-20")),
3   c(as.Date("2022-06-20"), as.Date("2024-06-20")),
4   c(as.Date("2021-06-20"), as.Date("2024-06-20"))
5 )
```

Listing 4: Funzione per il calcolo della volatilità annuale

```
1
2 calcola_volatilita <- function(symbol, date_range) {
3   getSymbols(symbol, src = "yahoo",
4   from = date_range[1], to = date_range[2],
5   auto.assign = FALSE) %>%
6   Ad() %>%
7   log() %>%
8   diff() %>%
9   sd(na.rm = TRUE) * sqrt(252)
10 }
```

Listing 5: Applicazione della funzione e creazione del dataframe

```
1
2 volatilita_list <- lapply(simboli, FUN = function(i) {
3   lapply(date_ranges, FUN = function(range) {
4     calcola_volatilita(i, range)}
```

```

5   }) %>%
6   set_names(paste0("volatilita_annuale_",
7     seq_along(date_ranges)))
8 } %>%
9 set_names(simboli)
10
11 support <- bind_rows(lapply(names(volatilita_list),
12
13 function(symbol) {
14   data.frame(
15     simbolo = symbol,
16     t(volatilita_list[[symbol]]))
17   )
18 }))
```

Listing 6: Integrazione delle volatilità al dataset

```

1 dati_rid <- dati_rid %>%
2   left_join(support,
3   by = c("underlying_symbol" = "simbolo"))
4
5 colonne_volatilita <- paste0("volatilita_annuale_",
6
7 seq_along(date_ranges))
8
9 dati_rid <- dati_rid %>%
10   mutate(across(all_of(colonne_volatilita), as.numeric)
11   )
```

.2 Listings Codice R: Analisi Outliers - LOF

```

1 library(Rlof)
2 # Suddividiamo i dati in call e put
3 data_calls <- data[data$option_type == 'c', ]
4 data_puts <- data[data$option_type == 'p', ]
5
6 data_calls <- data_calls[, !(names(data_calls)
7 %in% c('option_type'))]
8 data_puts <- data_puts[, !(names(data_puts)
9 %in% c('option_type'))]
10
11 # Utilizzare solo colonne numeriche
12 data_calls_numeric <- data_calls[, sapply(data_calls,
13   is.numeric)]
13 data_puts_numeric <- data_puts[, sapply(data_puts, is.
14   numeric)]
```

```

14
15 # Definiamo il numero di vicini per LOF
16 k_neighbors <- 20
17
18 call_LOF <- lof(data_calls_numeric, k = k_neighbors)
19 put_LOF <- lof(data_puts_numeric, k = k_neighbors)
20
21 # Ispezioniamo la distribuzione di LOF
22 summary(call_LOF)
23 summary(put_LOF)
24
25 # Definiamo una soglia per la classificazione degli
26 # outlier
27 threshold <- quantile(call_LOF, 0.95)
28
29 # Funzione per separare inliers e outliers in base ai
30 # punteggi LOF
31 inlier_outlier_sep <- function(LOF, dataset, threshold)
32 {
33 # Converti i risultati LOF in un dataframe
34 LOF_df <- data.frame(LOF)
35
36 # Identifica gli inliers (LOF sotto la soglia)
37 # e gli outliers (LOF sopra la soglia)
38 inlier_index_vals <- which(LOF_df[, 1] <= threshold)
39 outlier_index_vals <- which(LOF_df[, 1] > threshold)
40
41 # Verifica se ci sono inliers o outliers
42 print(length(inlier_index_vals)) # Numero di inliers
43 print(length(outlier_index_vals)) # Numero di
44 # outliers
45
46 # Sottogruppo del dataset basato sugli indici di
47 # inliers e outliers
48 inlier <- dataset[inlier_index_vals, ]
49 outlier <- dataset[outlier_index_vals, ]
50
51 return(list(inlier = inlier, outlier = outlier))
52 }
53
54 # Separa inliers e outliers per calls e puts
55 call_inlier_outlier <-
56 inlier_outlier_sep(call_LOF, data_calls,
57 threshold)
58 put_inlier_outlier <-
59 inlier_outlier_sep(put_LOF, data_puts,
60 threshold)
61
62 # Estraiamo inliers e outliers dalle liste

```

```

58 call_inlier <- call_inlier_outlier$inlier
59 call_outlier <- call_inlier_outlier$outlier
60 put_inlier <- put_inlier_outlier$inlier
61 put_outlier <- put_inlier_outlier$outlier
62
63 # Aggiungiamo la colonna 'option_type' agli inliers
64 call_inlier$option_type <- 'c'
65 put_inlier$option_type <- 'p'
66
67 # Combiniamo gli inliers di call e put in un unico
#      dataframe
68 final_train_data <- rbind(call_inlier, put_inlier)
69
70 # Mescoliamo le righe del dataframe finale
71 set.seed(123)
72 final_train_data <- final_train_data[sample(nrow(final_
    train_data)),]
73
74 print(nrow(final_train_data)) # Numero combinato di
#      inliers
75 print(nrow(data))           # Numero di righe
#      originali

```

.3 Listings Codice R: Modello BLACK SCHOLES

Listing 7: Funzione per il calcolo dei prezzi delle opzioni con il modello di Black-Scholes

```

1 black_scholes <- function(S, X, T_days,
2                             sigma, r = 0.051, option_type) {
3
4     T <- T_days / 252
5     d1 <- (log(S / X) + (r + 0.5 * sigma^2) * T) / (
6         sigma * sqrt(T))
7     d2 <- d1 - sigma * sqrt(T)
8
9     if (option_type == "c") {
10        price <- S * pnorm(d1) - X * exp(-r * T) *
11            pnorm(d2)
12    } else if (option_type == "p") {
13        price <- X * exp(-r * T) * pnorm(-d2) - S *
14            pnorm(-d1)
15    } else {
16        stop("Invalid option type.
17             Use 'c' for call or 'p' for put.")
18    }
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
279
280
281
282
283
284
285
286
287
288
289
289
290
291
292
293
294
295
296
297
298
299
299
300
301
302
303
304
305
306
307
308
309
309
310
311
312
313
314
315
316
317
318
319
319
320
321
322
323
324
325
326
327
328
329
329
330
331
332
333
334
335
336
337
338
339
339
340
341
342
343
344
345
346
347
348
349
349
350
351
352
353
354
355
356
357
358
359
359
360
361
362
363
364
365
366
367
368
369
369
370
371
372
373
374
375
376
377
378
379
379
380
381
382
383
384
385
386
387
388
389
389
390
391
392
393
394
395
396
397
398
399
399
400
401
402
403
404
405
406
407
408
409
409
410
411
412
413
414
415
416
417
418
419
419
420
421
422
423
424
425
426
427
428
429
429
430
431
432
433
434
435
436
437
438
439
439
440
441
442
443
444
445
446
447
448
449
449
450
451
452
453
454
455
456
457
458
459
459
460
461
462
463
464
465
466
467
468
469
469
470
471
472
473
474
475
476
477
478
479
479
480
481
482
483
484
485
486
487
488
489
489
490
491
492
493
494
495
496
497
498
499
499
500
501
502
503
504
505
506
507
508
509
509
510
511
512
513
514
515
516
517
518
519
519
520
521
522
523
524
525
526
527
528
529
529
530
531
532
533
534
535
536
537
538
539
539
540
541
542
543
544
545
546
547
548
549
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
678
679
680
681
682
683
684
685
686
687
687
688
689
689
690
691
692
693
694
695
696
696
697
698
699
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
717
718
719
719
720
721
722
723
724
725
726
727
727
728
729
729
730
731
732
733
734
735
736
737
737
738
739
739
740
741
742
743
744
745
746
747
747
748
749
749
750
751
752
753
754
755
756
757
757
758
759
759
760
761
762
763
764
765
766
766
767
768
768
769
769
770
771
772
773
774
775
776
776
777
778
778
779
779
780
781
782
783
784
785
785
786
787
787
788
788
789
789
790
791
792
793
794
794
795
795
796
796
797
797
798
798
799
799
800
801
801
802
802
803
803
804
804
805
805
806
806
807
807
808
808
809
809
810
810
811
811
812
812
813
813
814
814
815
815
816
816
817
817
818
818
819
819
820
820
821
821
822
822
823
823
824
824
825
825
826
826
827
827
828
828
829
829
830
830
831
831
832
832
833
833
834
834
835
835
836
836
837
837
838
838
839
839
840
840
841
841
842
842
843
843
844
844
845
845
846
846
847
847
848
848
849
849
850
850
851
851
852
852
853
853
854
854
855
855
856
856
857
857
858
858
859
859
860
860
861
861
862
862
863
863
864
864
865
865
866
866
867
867
868
868
869
869
870
870
871
871
872
872
873
873
874
874
875
875
876
876
877
877
878
878
879
879
880
880
881
881
882
882
883
883
884
884
885
885
886
886
887
887
888
888
889
889
890
890
891
891
892
892
893
893
894
894
895
895
896
896
897
897
898
898
899
899
900
900
901
901
902
902
903
903
904
904
905
905
906
906
907
907
908
908
909
909
910
910
911
911
912
912
913
913
914
914
915
915
916
916
917
917
918
918
919
919
920
920
921
921
922
922
923
923
924
924
925
925
926
926
927
927
928
928
929
929
930
930
931
931
932
932
933
933
934
934
935
935
936
936
937
937
938
938
939
939
940
940
941
941
942
942
943
943
944
944
945
945
946
946
947
947
948
948
949
949
950
950
951
951
952
952
953
953
954
954
955
955
956
956
957
957
958
958
959
959
960
960
961
961
962
962
963
963
964
964
965
965
966
966
967
967
968
968
969
969
970
970
971
971
972
972
973
973
974
974
975
975
976
976
977
977
978
978
979
979
980
980
981
981
982
982
983
983
984
984
985
985
986
986
987
987
988
988
989
989
990
990
991
991
992
992
993
993
994
994
995
995
996
996
997
997
998
998
999
999
1000
1000
1001
1001
1002
1002
1003
1003
1004
1004
1005
1005
1006
1006
1007
1007
1008
1008
1009
1009
1010
1010
1011
1011
1012
1012
1013
1013
1014
1014
1015
1015
1016
1016
1017
1017
1018
1018
1019
1019
1020
1020
1021
1021
1022
1022
1023
1023
1024
1024
1025
1025
1026
1026
1027
1027
1028
1028
1029
1029
1030
1030
1031
1031
1032
1032
1033
1033
1034
1034
1035
1035
1036
1036
1037
1037
1038
1038
1039
1039
1040
1040
1041
1041
1042
1042
1043
1043
1044
1044
1045
1045
1046
1046
1047
1047
1048
1048
1049
1049
1050
1050
1051
1051
1052
1052
1053
1053
1054
1054
1055
1055
1056
1056
1057
1057
1058
1058
1059
1059
1060
1060
1061
1061
1062
1062
1063
1063
1064
1064
1065
1065
1066
1066
1067
1067
1068
1068
1069
1069
1070
1070
1071
1071
1072
1072
1073
1073
1074
1074
1075
1075
1076
1076
1077
1077
1078
1078
1079
1079
1080
1080
1081
1081
1082
1082
1083
1083
1084
1084
1085
1085
1086
1086
1087
1087
1088
1088
1089
1089
1090
1090
1091
1091
1092
1092
1093
1093
1094
1094
1095
1095
1096
1096
1097
1097
1098
1098
1099
1099
1100
1100
1101
1101
1102
1102
1103
1103
1104
1104
1105
1105
1106
1106
1107
1107
1108
1108
1109
1109
1110
1110
1111
1111
1112
1112
1113
1113
1114
1114
1115
1115
1116
1116
1117
1117
1118
1118
1119
1119
1120
1120
1121
1121
1122
1122
1123
1123
1124
1124
1125
1125
1126
1126
1127
1127
1128
1128
1129
1129
1130
1130
1131
1131
1132
1132
1133
1133
1134
1134
1135
1135
1136
1136
1137
1137
1138
1138
1139
1139
1140
1140
1141
1141
1142
1142
1143
1143
1144
1144
1145
1145
1146
1146
1147
1147
1148
1148
1149
1149
1150
1150
1151
1151
1152
1152
1153
1153
1154
1154
1155
1155
1156
1156
1157
1157
1158
1158
1159
1159
1160
1160
1161
1161
1162
1162
1163
1163
1164
1164
1165
1165
1166
1166
1167
1167
1168
1168
1169
1169
1170
1170
1171
1171
1172
1172
1173
1173
1174
1174
1175
1175
1176
1176
1177
1177
1178
1178
1179
1179
1180
1180
1181
1181
1182
1182
1183
1183
1184
1184
1185
1185
1186
1186
1187
1187
1188
1188
1189
1189
1190
1190
1191
1191
1192
1192
1193
1193
1194
1194
1195
1195
1196
1196
1197
1197
1198
1198
1199
1199
1200
1200
1201
1201
1202
1202
1203
1203
1204
1204
1205
1205
1206
1206
1207
1207
1208
1208
1209
1209
1210
1210
1211
1211
1212
1212
1213
1213
1214
1214
1215
1215
1216
1216
1217
1217
1218
1218
1219
1219
1220
1220
1221
1221
1222
1222
1223
1223
1224
1224
1225
1225
1226
1226
1227
1227
1228
1228
1229
1229
1230
1230
1231
1231
1232
1232
1233
1233
1234
1234
1235
1235
1236
1236
1237
1237
1238
1238
1239
1239
1240
1240
1241
1241
1242
1242
1243
1243
1244
1244
1245
1245
1246
1246
1247
1247
1248
1248
1249
1249
1250
1250
1251
1251
1252
1252
1253
1253
1254
1254
1255
1255
1256
1256
1257
1257
1258
1258
1259
1259
1260
1260
1261
1261
1262
1262
1263
1263
1264
1264
1265
1265
1266
1266
1267
1267
1268
1268
1269
1269
1270
1270
1271
1271
1272
1272
1273
1273
1274
1274
1275
1275
1276
1276
1277
1277
1278
1278
1279
1279
1280
1280
1281
1281
1282
1282
1283
1283
1284
1284
1285
1285
1286
1286
1287
1287
1288
1288
1289
1289
1290
1290
1291
1291
1292
1292
1293
1293
1294
1294
1295
1295
1296
1296
1297
1297
1298
1298
1299
1299
1300
1300
1301
1301
1302
1302
1303
1303
1304
1304
1305
1305
1306
1306
1307
1307
1308
1308
1309
1309
1310
1310
1311
1311
1312
1312
1313
1313
1314
1314
1315
1315
1316
1316
1317
1317
1318
1318
1319
1319
1320
1320
1321
1321
1322
1322
1323
1323
1324
1324
1325
1325
1326
1326
1327
1327
1328
1328
1329
1329
1330
1330
1331
1331
1332
1332
1333
1333
1334
1334
1335
1335
1336
1336
1337
1337
1338
1338
1339
1339
1340
1340
1341
1341
1342
1342
1343
1343
1344
1344
1345
1345
1346
1346
1347
1347
1348
1348
1349
1349
1350
1350
1351
1351
1352
1352
1353
1353
1354
1354
1355
1355
1356
1356
1357
1357
1358
1358
1359
1359
1360
1360
1361
1361
1362
1362
1363
1363
1364
1364
1365
1365
1366
1366
1367
1367
1368
1368
1369
1369
1370
1370
1371
1371
1372
1372
1373
1373
1374
1374
1375
1375
1376
1376
1377
1377
1378
1378
1379
1379
1380
1380
1381
1381
1382
1382
1383
1383
1384
1384
1385
1385
1386
1386
1387
1387
1388
1388
1389
1389
1390
1390
1391
1391
1392
1392
1393
1393
1394
1394
1395
1395
1396
1396
1397
1397
1398
1398
1399
1399
1400
1400
1401
1401
1402
1402
1403
1403
1404
1404
1405
1405
1406
1406
1407
1407
1408
1408
1409
1409
1410
1410
1411
1411
1412
1412
1413
1413
1414
1414
1415
1415
1416
1416
1417
1417
1418
1418
1419
1419
1420
1420
1421
1421
1422
1422
1423
1423
1424
1424
1425
1425
1426
1426
1427
1427
1428
1428
1429
1429
1430
1430
1431
1431
1432
1432
1433
1433
1434
1434
1435
1435
1436
1436
1437
1437
1438
1438
1439
1439
1440
1440
1441
1441
1442
1442
1443
1443
1444
1444
1445
1445
1446
1446
1447
1447
1448
1448
1449
1449
1450
1450
1451
1451
1452
1452
1453
1453
1454
1454
1455
1455
1456
1456
1457
1457
1458
1458
1459
1459
1460
1460
1461
1461
1462
1462
1463
1463
1464
1464
1465
1465
1466
1466
1467
1467
1468
1468
1469
1469
1470
1470
1471
1471
1472
1472
1473
1473
1474
1474
1475
1475
1476
1476
1477
1477
1478
1478
1479
1479
1480
1480
1481
1481
1482
1482
1483
1483
1484
1484
1485
1485
1486
1486
1487
1487
1488
1488
1489
1489
1490
1490
1491
1491
1492
1492
1493
1493
1494
1494
1495
1495
1496
1496
1497
1497
1498
1498
1499
1499
1500
1500
1501
1501
1502
1502
1503
1503
1504
1504
1505
1505
1506
1506
1507
1507
1508
1508
1509
1509
1510
1510
1511
1511
1512
1512
1513
1513
1514
1514
1515
1515
1516
1516
1517
1517
1518
1518
1519
1519
1520
1520
1521
1521
1522
1522
1523
1523
1524
1524
1525
1525
1526
1526
1527
1527
1528
1528
1529
1529
1530
1530
1531
1531
1532
1532
1533
1533
1534
1534
1535
1535
1536
1536
1537
1537
1538
1538
1539
1539
1540
1540
1541
1541
1542
1542
1543
1543
1544
1544
1545
1545
1546
1546
1547
1547
1548
1548
1549
1549
1550
1550
1551
1551
1552
1552
1553
1553
1554
1554
1555
1555
1556
1556
1557
1557
1558
1558
1559
1559
```

```

15     }
16     return(price)
17 }
```

Listing 8: Calcolo dei prezzi delle opzioni utilizzando tre misure di volatilità storica

```

1 volatilities <- list(
2   data$volatilita_annuale_1,
3   data$volatilita_annuale_2,
4   data$volatilita_annuale_3
5 )
6
7 for (i in seq_along(volatilities)) {
8   col_name <- paste0("bs_price_sigma", i)
9   data[[col_name]] <- mapply(black_scholes,
10                           S = data$underlying_
11                             trade_price,
12                           X = data$strike,
13                           T_days = data$n_giorni,
14                           sigma = volatilities[[i
15                             ]],
16                           MoreArgs = list(r =
17                             0.051),
18                           option_type = data$option_type)
19 }
```

.4 Listings Codice R: Metodo Montecarlo

Listing 9: Calcolo del prezzo delle opzioni tramite simulazione Monte Carlo

```

1 volatilities <- list(data$volatilita_annuale_1,
2   data$volatilita_annuale_2,
3   data$volatilita_annuale_3)
4
5 monte_carlo_option_pricing <- function(S, X, T_days,
6                                         sigma, r = 0.051,
7                                         option_type, n_sim =
8                                         50000) {
9   T <- T_days / 252
10  payoff <- numeric(n_sim)
11
12 for (i in 1:n_sim) {
13   S_T <- S * exp((r - 0.5 * sigma^2) * T +
```

```

14     sigma * sqrt(T) * rnorm(1))
15
16     if (option_type == "c") {
17       payoff[i] <- max(S_T - X, 0)
18     } else if (option_type == "p") {
19       payoff[i] <- max(X - S_T, 0)
20     } else {
21       stop("Invalid option type.
22           Use 'c' for call or 'p' for put.")
23     }
24   }
25
26   price <- mean(payoff) * exp(-r * T)
27   return(price)
28 }
29
30 for (i in seq_along(volatilities)) {
31   col_name <- paste0("mc_price_sigma", i)
32   data[[col_name]] <- mapply(monte_carlo_option_pricing
33     ,
34     S = data$underlying_trade_
35       price,
36     X = data$strike,
37     T_days = data$n_giorni,
38     sigma = volatilities[[i]],
39     MoreArgs = list(r = 0.051,
40     n_sim = 50000),
41     option_type = data$option_
42       type)
43 }
```

.5 Listings Codice R: XGBoost

Listing 10: Implementazione del Modello XGBoost

```
1 # Carica i dati
2 load("DATASET_DEFINITIVO.RData")
3
4 # Preprocessamento dei dati
5 X <- final_train_data %>%
6   mutate(option_type = if_else(option_type == "p", 1,
7     0)) %>%
8   mutate(delivery = if_else(delivery == "C", 1, 0)) %>%
9   mutate(n_giorni = as.integer(n_giorni)) %>%
10  select(-c("underlying_symbol", "expiration", "
11    exercise", "root",
12      "trade_condition_id", "trade_datetime", "
13        settlement",
14        "multiplier", "exchange_sequence_number", "
15          exchange_id",
16        "volatilita_annuale_2", "volatilita_annuale
17          _3"))
18
19 X_call <- X %>% filter(option_type == 0)
20 X_put <- X %>% filter(option_type == 1)
21
22 # Target per prezzi di scambio
23 Y_call_scambio <- X_call$underlying_trade_price
24 Y_put_scambio <- X_put$underlying_trade_price
25
26 # Target per prezzi teorici
27 Y_call_theo <- X_call$theo_price
28 Y_put_theo <- X_put$theo_price
29
30 # Rimuove le colonne target dai dataset
31 X_call_scambio <- X_call %>% select(-underlying_trade_
32   price)
33 X_put_scambio <- X_put %>% select(-underlying_trade_
34   price)
35
36 X_call_theo <- X_call %>% select(-theo_price)
37 X_put_theo <- X_put %>% select(-theo_price)
38
39 # Definizione delle variabili opzionali e non opzionali
40 features_optional <- c('delivery', 'open_interest', '
41   underlying_bid',
42       "underlying_ask", "trade_size",
43       "trade_iv")
44
45 features_non_optional <- c('trade_delta', 'trade_gamma'
46   , 'trade_vega',
```

```

36           'trade_theta', "trade_rho",
37           "theo_iv", "strike",
38           "n_giorni", "option_type", "
39           volatilita_annuale_1")
40
41 # Creazione di tutte le combinazioni di variabili
42 # opzionali
43 feature_subsets <- list()
44 for (n in 0:length(features_optional)) {
45   comb <- combn(features_optional, n, simplify = FALSE)
46   for (x in comb) {
47     feature_subsets <- append(feature_subsets, list(c(
48       features_non_optional, x)))
49   }
50 }
51
52 # Funzione per creare il modello XGBoost e calcolare
53 # RMSE
54 train_and_evaluate <- function(X, Y, feature_subset) {
55   set.seed(42)
56   trainIndex <- createDataPartition(Y, p = .7, list =
57     FALSE, times = 1)
58   X_train <- X[trainIndex, feature_subset, drop = FALSE
59   ]
60   X_test <- X[-trainIndex, feature_subset, drop = FALSE
61   ]
62   Y_train <- Y[trainIndex]
63   Y_test <- Y[-trainIndex]
64
65   dtrain <- xgb.DMatrix(data = as.matrix(X_train),
66                         label = Y_train)
67   dtest <- xgb.DMatrix(data = as.matrix(X_test), label
68                         = Y_test)
69
70   params <- list(
71     objective = "reg:squarederror",
72     eval_metric = "rmse",
73     max_depth = 5,
74     eta = 0.2,
75     gamma = 2,
76     colsample_bytree = 1,
77     min_child_weight = 1,
78     subsample = 0.8
79   )
80
81   xgbcv <- xgb.cv(params = params, data = dtrain,
82                     nrounds = 100,
83                     nfold = 5, showsd = TRUE, stratified
84                     = TRUE,

```

```

73         print_every_n = 10, early_stopping_
74             rounds = 10,
75             maximize = FALSE)
76
77 bst <- xgb.train(params = params, data = dtrain,
78   nrounds = xgbcv$best_iteration)
79
80 pred <- predict(bst, newdata = dtest)
81 RMSE <- sqrt(mean((Y_test - pred)^2))
82 return(RMSE)
83 }
84
85 # Esegue la Random Grid Search sui due insiemi di
86 # variabili
87 set.seed(42)
88 random_search_results <- lapply(feature_subsets,
89   function(fs) {
90     rmse_scambio_call <- train_and_evaluate(X_call_
91       scambio, Y_call_scambio, fs)
92     rmse_scambio_put <- train_and_evaluate(X_put_scambio,
93       Y_put_scambio, fs)
94     rmse_theo_call <- train_and_evaluate(X_call_theo, Y_
95       call_theo, fs)
96     rmse_theo_put <- train_and_evaluate(X_put_theo, Y_put
97       _theo, fs)
98     list(features = fs, rmse_scambio_call = rmse_scambio_
99       call,
100       rmse_scambio_put = rmse_scambio_put, rmse_theo_
101         call = rmse_theo_call,
102       rmse_theo_put = rmse_theo_put)
103   })
104
105 # Analisi dei risultati per selezionare il modello
106 # migliore
107 best_model <- random_search_results[[which.min(sapply(
108   random_search_results,
109   function(x) x$rmse_scambio_call))]]
110 print(best_model)

```