

Analisi del prezzo dell'oro

in Relazione con alcuni commodities e indici

Andrea Moressa,1074124

Amin Borqal,1073928

Loris Iacoban,1074130

INTRODUZIONE:

Anno accademico:
2021-2022

Il dataset utilizzato per il progetto è stato preso dal sito Kaggle:
<https://www.kaggle.com/code/somyaagarwal69/gold-forecasting-with-regression-arma/data?select=GoldUP.csv>

Il lavoro svolto consiste nel cercare di interpretare, attraverso modelli di regressione, l'andamento del prezzo dell'oro e la sua relazione con i regressori presenti nel dataset.

1. Parametri

Gold_Price : prezzo dell'oro in quel determinato mese

Crude_Oil : prezzo del petrolio in quel determinato mese

Interest_Rate o tasso di interesse : è l'importo che un finanziatore addebita a un mutuatario ed è una percentuale del capitale, l'importo prestato

USD_INR : tasso di cambio dei dollari in rupie

Sensex : si riferisce all'indice di riferimento della BSE (Bombay Stock Exchange) in India; è composto da 30 dei titoli più grandi e scambiati più attivamente e fornisce un indicatore dell'economia indiana

CPI o Consumer price index : dall'inglese, indice dei prezzi al consumo, è un indice che viene calcolato per mezzo di una media ponderata dei prezzi relativi ad un insieme di beni e servizi in un determinato periodo di tempo. Tale insieme è rappresentativo delle abitudini di spesa del consumatore medio. Il CPI è importante in quanto, misurando le variazioni dei prezzi, segnala l'aumento dell'inflazione

USD_Index : l'indice del dollaro; è una misura del valore del dollaro rispetto a un insieme di valute estere

2. Quesiti a cui rispondere

1. Come si comporta il prezzo dell'oro (**Gold_Price**) basandosi sugli altri attributi osservati?
2. Esiste multicollinearità nel modello selezionato al punto precedente (2.1)?
3. Nel modello selezionato al punto (2.1) i residui sono normali? Hanno media pari a zero?
4. Esiste una relazione tra il prezzo del petrolio (**Crude_Oil**) e l'indice del dollaro (**USD_Index**)?
5. È possibile utilizzare il metodo WLS per il modello selezionato al primo punto (2.1)?

2.1. Come si comporta il prezzo dell'oro (Gold_Price) basandosi sugli altri attributi osservati?

Per rispondere a questa domanda abbiamo analizzato singolarmente come gli attributi si comportassero con **Gold_Price**.

Regressori	R2	MSE	pValue
Crude_Oil	0.438	8.77e+03	1.78e-31
Interest_Rate	0.0586	1.13e+04	1.58 10 ⁴
USD_INR	0.732	6.05e+03	1.05e-69
Sensex	0.805	5.15e+03	3.17e-86
CPI	0.92	3.31e+03	7.83e-132
USD_Index	0.0176	1.16e+04	0.0402

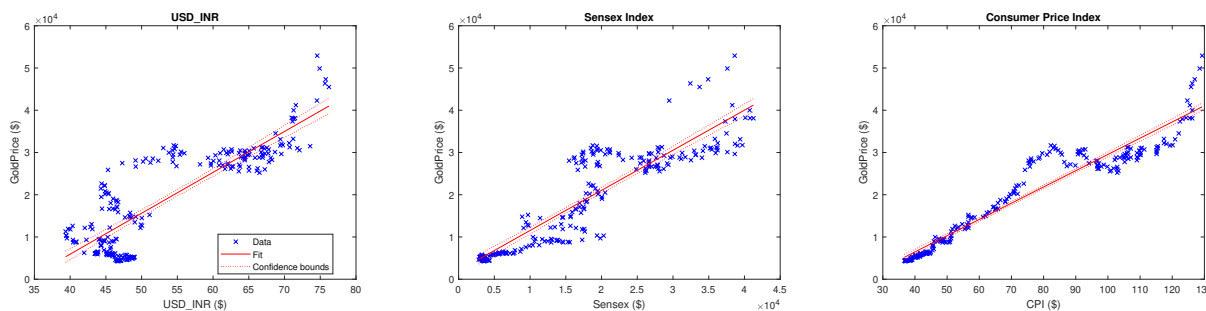
Attraverso la tecnica Stepwise backward selezioniamo il modello migliore che ha come regressori (**USD_INR**, **Sensex**, **CPI**), abbiamo poi applicato un modello di regressione lineare multipla.

Regressori	R2	MSE	pValue
CPI + Sensex + USD_INR	0.953	2.53e+03	3.65e-156
CPI + Sensex	0.927	3.17e+03	8.96e-135
CPI + USD_INR	0.929	3.13e+03	3.85e-136
USD_INR + Sensex	0.841	4.67e+03	4.31e-95

Dopo l'analisi del coefficiente di determinazione multipla e l'errore quadratico medio, abbiamo deciso di concentrarci sul modello **CPI + Sensex + USD_INR** in quanto il valore di R2 risulta essere il più alto. Ecco la tabella finale dei parametri $\hat{\beta}$, con statistiche test :

	Regressore	Valore stimato	SE	Statistica T	pValue
β_0	Intercetta	2437.2	1372.7	1.7755	0.077111
β_1	CPI	830.46	34.92	23.782	1.741e-64
β_2	Sensex	-0.69783	0.062574	-11.152	1.8518e-23
β_3	USD_INR	-583.4	50.383	-11.579	7.8805e-25

Dato che i parametri indipendenti sono tre non è stato possibile il grafico dell'iperpiano di regressione. Sarebbero stati in quattro dimensioni e di conseguenza non rappresentabili. Andiamo quindi a mostrare i grafici parziali per avere un'idea dell'andamento :

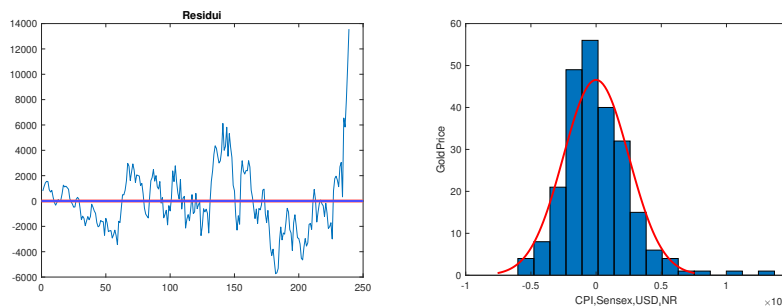


2.2. Esiste multicollinearità nel modello selezionato al punto precedente (2.1)?

La multicollinearità sorge quando c'è un'elevata correlazione tra due o più variabili esplicative, che porta lo stimatore dei minimi quadrati a non dare risultati attendibili. Se c'è una correlazione perfetta tra le due variabili, la matrice $(X'X)$ diventa singolare, ha determinante uguale a 0 circa e perciò non esiste la matrice inversa. Abbiamo quindi verificato che fosse maggiore di 0 per assicurarci che non vi fosse multicollinearità.

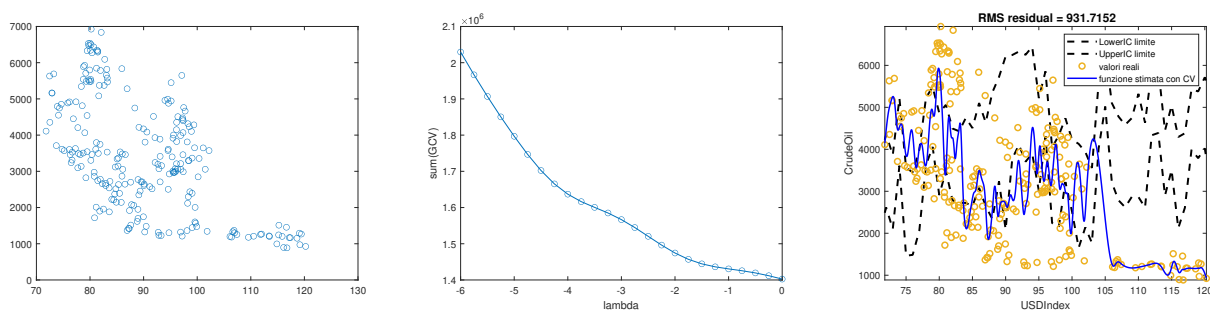
2.3. Nel modello selezionato al punto (2.1) i residui sono normali? Hanno media pari a zero?

I residui sono la differenza tra i valori osservati e stimati in un'analisi di regressione. L'iperpiano di regressione deve trovarsi lungo il centro dei punti dati. Pertanto la somma dei residui deve essere zero. Abbiamo verificato che i residui avessero una distribuzione normale e media uguale a zero.



2.4. Esiste una relazione tra il prezzo del petrolio (Crude_Oil) e l'indice del dollaro (USD_Index)?

Come si può vedere dal primo grafico, abbiamo una correlazione non lineare tra **Crude_Oil** e **USD_Index** con dati abbastanza dispersi. Per modellizzare questa relazione abbiamo utilizzato una regressione polinomiale ovvero la regressione spline. Abbiamo utilizzato l'algoritmo di crossvalidazione con k-fold per lo smoothing della funzione, una gestione ottimale del numero di nodi e per trovare il miglior parametro di liscio. Abbiamo Calcolato λ e la matrice di penalizzazione Rmat (derivante dalla penalizzazione) per penalizzare il modello, l'ordine è uguale a 6 e nbasis=243. Il lambda ottimale, da noi ottenuto, vale 0.01. Lo smoothing ottenuto tramite crossvalidazione e la funzione stimata le vediamo nei seguenti grafici :



2.5. Potevamo utilizzare i minimi quadrati ponderati per il modello selezionato al primo punto (2.1)?

Il modello è migliorabile utilizzando i minimi quadrati ponderati in cui la matrice di varianza-covarianza degli errori è la matrice dei pesi Ω . Il peso impone che le osservazioni che si trovano in zone dove l'errore è più disperso attorno all'iperpiano di regressione contino meno nel definire i coefficienti. Mediante ciò diventa possibile avere un iperpiano di regressione che "spiega" maggiormente i dati. I parametri $\beta_0 \beta_1 \beta_2 \beta_3$ che avevamo ottenuto con OLS sono 2472.2, 830.46, -0.69 e -583.3989. I nuovi ottenuti con WLS sono 2433.65, 830.78, -0.671 e -591.76.

3. Conclusioni

Attraverso tecniche di regressione lineare multipla che abbiamo verificato la forte correlazione che c'è tra il **Gold_price** e gli indici/commodities **CPI**, **Sensex**, **USD_INR**. Abbiamo deciso inoltre di applicare la regressione spline con tecniche di crossvalidazione tra il **Gold_price** e **USD_Index** per evitare overfitting e trovare un modello che "spiegasse" il più possibile una correlazione non lineare. Infine abbiamo cercato di migliorare il modello di regressione lineare multipla cercando un iperpiano di regressione che "spiegasse più accuratamente" i dati con l'utilizzo dei minimi quadrati ponderati.