Matrici random: analisi dello S&P500

Francesco Zeno Costanzo

7 novembre 2022

Introduzione

Lo scopo della trattazione è discutere alcuni aspetti di analisi del mercato, utilizzando le componenti dello S&P 500, tramite la teoria delle matrici random. I dati da analizzare sono stati presi dal sito: https://it.finance.yahoo.com

Introduzione

Sono stati presi N = 482 indici dello S&P500 in un lasso di tempo che corre dal 03/10/2017 al 03/10/2022 quindi un L = 1259. Si calcolano i ritorni ad un giorno definiti come:

$$G_i(t) = \log(S_i(t + \Delta t)) - \log(S_i(t))$$

dove S_i è il prezzo dell'i-esimo titolo. Data la differente volatilità dei titoli si definisce il ritorno normalizzato come:

$$g_i(t) = \frac{G_i(t) - \langle G_i \rangle}{\sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}}$$

La correlazione tra i titoli sarà data da:

$$C_{ij} = \langle g_i g_j \rangle$$

Quello che si vuole fare è confrontare alcune proprietà della matrice di correlazione con quelle di una matrice random, nella fattispecie autovalori e autovettori. Prendiamo una matrice A $(N \times L)$ le cui entrate sono variabili gaussiane a media zero e varianza uno e scorrelate fra di loro; si definisce la matrice random come:

$$R = \frac{1}{L} A A^T$$

che risulta essere una matrice $(N \times N)$ di cui calcoliamo autovalori e autovettori.

Matrici random

La distribuzione attesa per gli autovalori di tale matrice, nel limite di N ed L tendenti ad infinito, è:

$$P(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

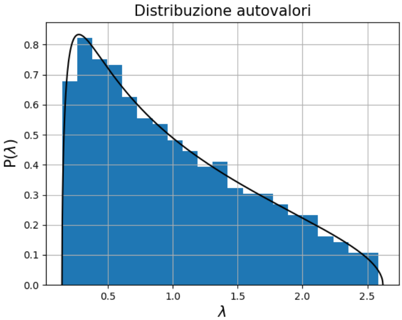
dove:

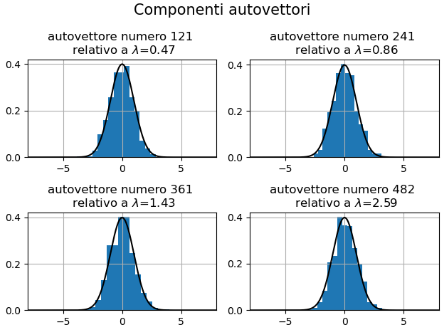
- $Q = \frac{L}{N}$
- $\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}$

Mentre la distribuzione delle componenti degli autovettori è data da:

$$\rho(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$$

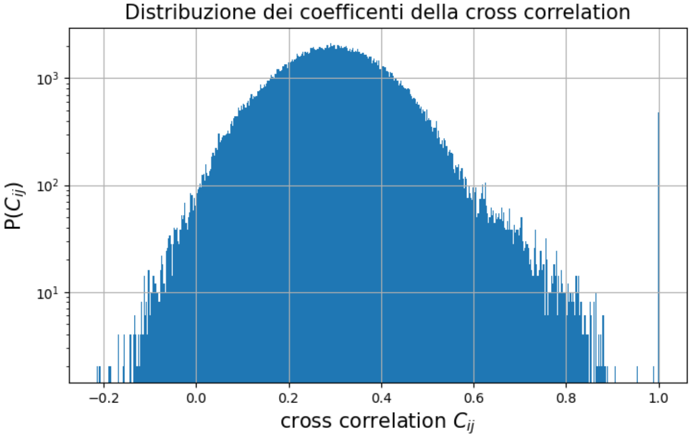
Nei seguenti grafici vediamo le due distribuzioni, calcolate su una vera matrice random con gli stessi N ed L.





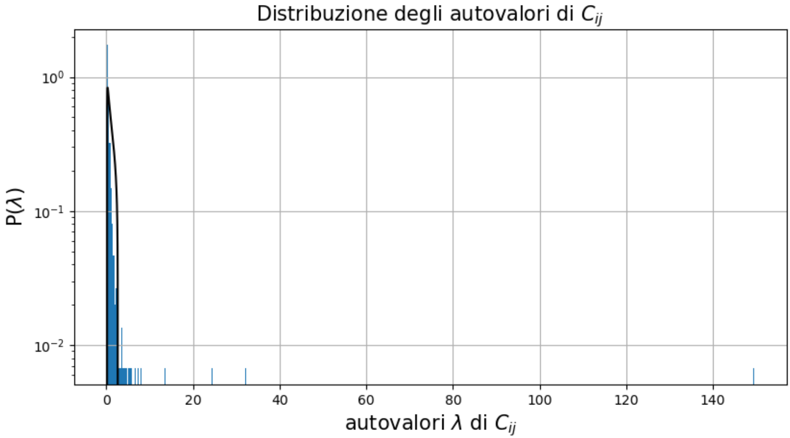
Analisi Mercato: correlazioni

Calcolando lo spettro degli autovalori per la matrice di correlazione fra i titoli vediamo che le cose non sono come prevede la teoria. Innanzi tutto vediamo dall'istogramma delle correlazioni come esso sia sbilanciato per valori positivi ad indicare una correlazione del mercato.

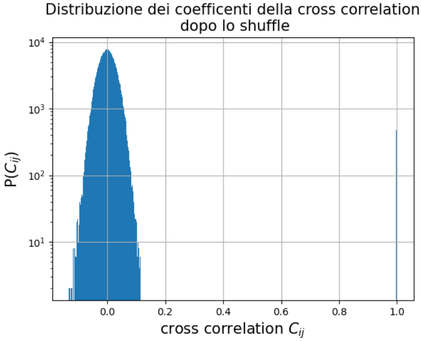


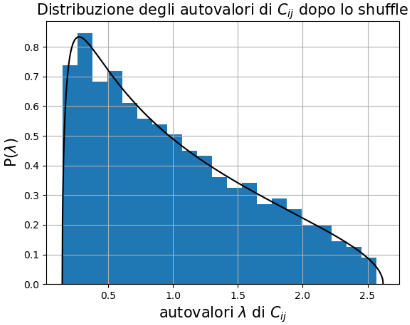
Analisi Mercato: correlazioni

Mentre per gli autovalori vediamo che alcuni sono fuori statistica.



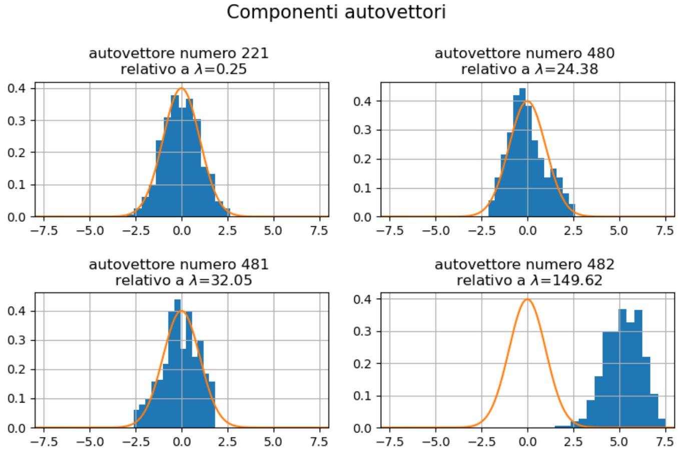
Per essere sicuri che gli autovalori fuori statistica portino effettivamente informazione si possono mischiare le serie temporali in modo da distruggere la correlazione:





Analisi Mercato: correlazioni

Se ora si considerano gli autovettori associati agli autovalori mostrati, si può vedere che quelli all'interno della regione predetta dalle matrici random sembrano essere gaussiani, mentre gli altri presentano deviazioni.



Analisi Mercato: andamento globale

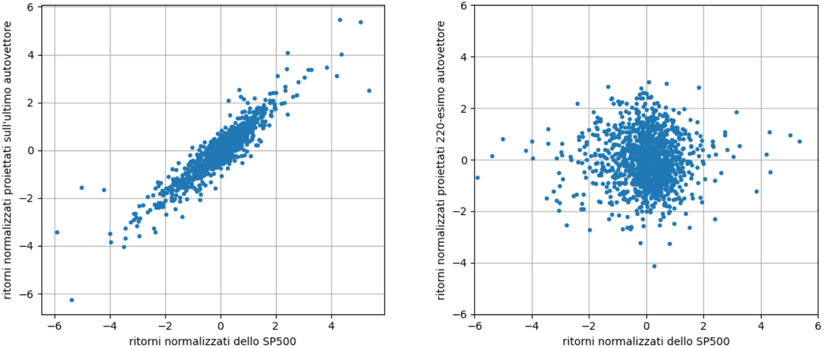
L'ultimo autovettore in particolare sembra essere più uniforme degli altri, il che può suggerire che molti titoli partecipino. Si può quindi interpretarlo come un andamento globale del mercato. Per vederlo meglio definiamo:

$$G^N(t) = \sum_{i=0}^{N} v_i^N G_i(t)$$

Ovvero la proiezione dei ritorni sull'ultimo autovettore, che per costruzione è il ritorno del portfolio definito da v^N.

Analisi Mercato: andamento globale

Confrontando G^N con G_{SP} si notano comportamenti simili:



Mentre usando un autovettore riferito ad un autovalore dentro la statistica non si notano dipendenze

Analisi Mercato: andamento globale

Calcolando le correlazioni tra G_{SP} e G^N e tra G_{SP} e G^{220} si ha:

- $\langle G_{SP} G^N \rangle = 0.923$
- $\langle G_{SP} G^{220} \rangle = 0.005$

Si può quindi interpretare v^N come ciò che quantifica l'influenza del mercato su tutti i titoli.

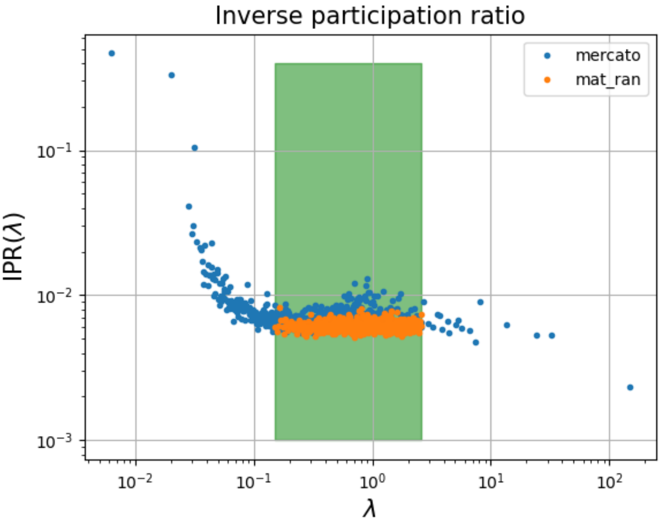
Analisi Mercato: andamento globale, IPR

Per capire quale sia il numero di componenti maggiormente significative nella distribuzione delle componenti degli autovettori si definisce:

$$I^k = \sum_{i=0}^{N} (v_i^k)^4$$

chiamato 'inverse participation ratio', il cui inverso ci dà una stima del numero dei partecipanti al k-esimo autovettore

Analisi Mercato: andamento globale, IPR



in verde la regione compresa tra λ_- e λ_+

Analisi Mercato: andamento globale, IPR

Si può quindi vedere che per l'autovalore più grande si ha il minimo di I^k e quindi un numero di partecipanti di circa 426 titoli.
Per il massimo di I^k invece, ovvero per l'autovalore più piccolo, si ha che il numero di partecipanti è circa 2 titoli.

Si può ora andare a vedere a quali titoli corrispondano le maggiori componenti degli autovettori, va prima però eliminato l'effetto globale che il mercato sembra avere.

Analisi Mercato: rimozione andamento globale

Per rimuovere l'andamento del mercato si può pensare ad un modello del tipo:

$$G_i(t) = \alpha_i + \beta_i M(t) + \epsilon_i(t)$$

dove:

- $M(t)$ è comune a tutto il mercato e approssimato con G^N
- $\epsilon_i(t)$ parte random con $\langle \epsilon_i \rangle = 0$
- α_i e β_i delle costanti

e calcolare poi una matrice $Cr_{ij} = \langle \epsilon_i \epsilon_j \rangle$

Analisi Mercato: rimozione andamento globale

Si esegue un fit di ogni ritorno secondo il modello:

$$G_i(t) = \beta_i G^N + c_i$$

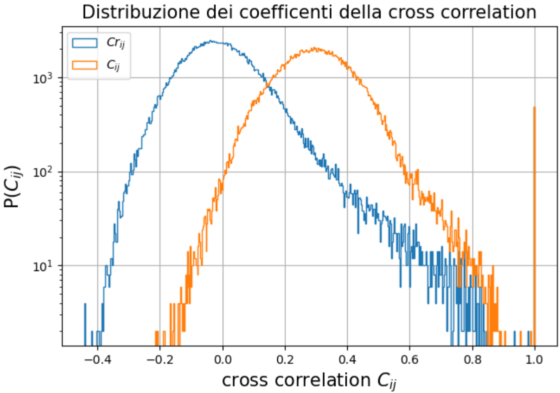
sapendo i β_i si stimano gli α_i usando il fatto che $\langle \epsilon_i \rangle = 0$:

$$\alpha_i = \langle G_i \rangle - \beta_i \langle G^N \rangle$$

per cui:

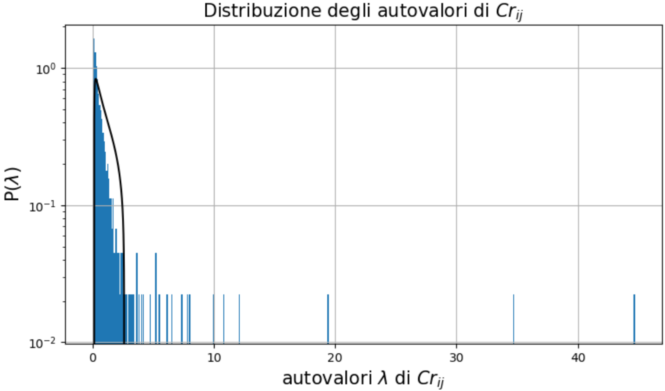
$$\epsilon_i(t) = G_i(t) - \alpha_i - \beta_i G^N$$

Analisi Mercato: rimozione andamento globale



Ora La distribuzione è maggiormente centrata in zero:
$\mu(C_{ij}) = 0.3$; $\mu(Cr_{ij}) = 3 \times 10^{-4}$.
Si potrebbe dire che buona parte del mercato è stata tolta.

Se infatti si va guardare gli autovalori si nota che ancora alcuni sono fuori statistica, benché il loro valore si sia comunque ridotto:



Analisi Mercato: Titoli

Si possono quindi andare a vedere le componenti dei relativi autovettori e trovare i titoli che hanno più rilevanza.

Autovettore: v^{481}

Symbol	Name	Sector
NVDA	Nvidia	Information Technology
ADBE	Adobe	Information Technology
ANET	Arista Networks	Information Technology
NRG	NRG Energy	Utilities
DRI	Darden Restaurants	Consumer Discretionary

Autovettore: v^{480}

Symbol	Name	Sector
O	Realty Income Corporation	Real Estate
VTR	Ventas	Real Estate
WELL	Welltower	Real Estate
PCAR	Paccar	Industrials
SRE	Sempra Energy	Utilities

Analisi Mercato: Titoli

Autovettore: v^{479}

Symbol	Name	Sector
MAS	Masco	Industrials
FBHS	Fortune Brands Home & Security	Industrials
SWK	Stanley Black & Decker	Industrials
WHR	Whirlpool Corporation	Consumer Discretionary
IT	Gartner	Information Technology

Autovettore: v^{478}

Symbol	Name	Sector
CVX	Chevron Corporation	Energy
KMI	Kinder Morgan	Energy
WMB	Williams Companies	Energy
COP	ConocoPhillips	Energy
FBHS	Fortune Brands Home & Security	Industrials

Analisi Mercato: Titoli

Autovettore: v^{477}

Symbol	Name	Sector
PNC	PNC Financial Services	Financials
JPM	JPMorgan Chase	Financials
SPGI	S&P Global	Financials
XOM	ExxonMobil	Energy
MRK	Merck & Co.	Health Care

Autovettore: v^{476}

Symbol	Name	Sector
HD	Home Depot	Consumer Discretionary
TRV	The Travelers Companies	Financials
AVGO	Broadcom	Information Technology
LRCX	Lam Research	Information Technology
AVGO	Broadcom	Information Technology

Ottimizzazione portfolio

Volendo creare un portfolio su questa analisi si può definire il guadagno del portfolio come:

$$\Phi = \sum_i w_i G_i$$

dove i G_i sono i ritorni dell'indice i-esimo e i w_i la frazione di denaro investita nel titolo i-esimo, per il quale vale una relazione di normalizzazione $\sum_i w_i = 1$.
Il rischio del portfolio può essere calcolato come:

$$\Omega^2 = \sum_i \sum_j w_i w_j C_{ij} \sigma_i \sigma_j$$

Minimizzando il rischio a Φ fissato, come primo vincolo, e imponendo la normalizzazione dei w_i, come secondo vincolo, si ottiene il portfolio ottimale.

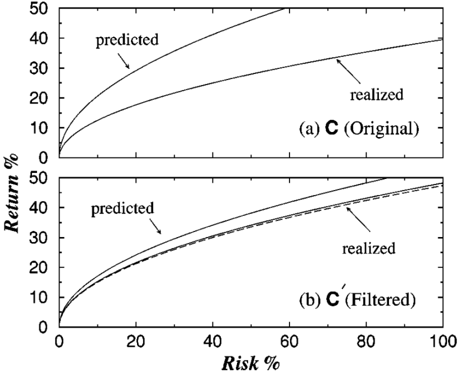
Poiché si è visto che le informazioni stanno negli ultimi autovalori e autovettori, per migliorare la predizione si può procedere diversamente sostituendo la matrice C_{ij} con una matrice filtrata $C^{'}_{ij}$:

$$C^{'} = \Lambda^{-1} C \Lambda$$

Dove Λ è tale che: $\Lambda_{ii} = (0, ..., 0, \lambda_p, ..., \lambda_N)$ ovvero contenente solo gli autovalori che deviano dalla distribuzione. Si impone inoltre che $C^{'}_{ii} = 1$ in modo da preservare la traccia.

Ottimizzazione portfolio

Di seguito è riportato il grafico presente nell'articolo di Plerou et al.



Nella figura in alto la curva 'predicted' è ottenuta usando i G_i degli indici del 1995 e la C del 1994 mentre quella 'realized' è ottenuta usando gli stessi G_i e la C del 1995. Nella figura in basso invece la curva 'predicted' è stata costruita usando i G_i degli indici del 1995 e la $C^{'}$ del 1994, e le altre sono ottenute con le stesse G_i usando la $C^{'}$ del 1995 (linea continua) e la C del 1995 (linea tratteggiata)

Bibliografia

Bibliografia:

Random matrix approach to cross correlation in financial data,
V. Plerou et al., Physical Review E, 65, 066126 (2002).