



Università degli Studi di Milano Bicocca

Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di laurea di Data Science

FINANCIAL MARKET ANALYTICS

Report Esame

Luca Poli 852027 l.poli6@campus.unimib.it

Sintesi

Introduzione	2
Capitolo 1: Strategie univariate	3
1.1 Descrizione della metodologia e dei fattori utilizzati.	3
1.1.2 Fattori e metriche utilizzati.....	3
1.2 Analisi complessiva dei risultati per tutte le strategie univariate	7
1.3 Analisi delle 3 migliori strategie	8
Capitolo 2: Strategie multivariate	10
2.1 Screening sequenziale	10
2.2 Screening simultaneo con ZScore.....	11
2.3 Analisi complessiva dei risultati per tutte le strategie multivariate	12
Conclusione	14
Appendice: Dettagli tecnici.....	15

Introduzione

In questo progetto di analisi dei mercati finanziari, svilupperò un modello di fattori (noto anche come modello di rischio) basato sul dataset fornito "Euro.xls". Il dataset è composto da 40 fattori per circa 800 titoli.

Un modello fattoriale è un modello statistico basato sull'ipotesi che le variazioni sistematiche dei prezzi siano generate principalmente da determinati fattori. Poiché il set di dati contiene variabili basate sui fondamentali delle società, costruirò un modello fondamentale (in contrapposizione a un modello economico basato su variabili macroeconomiche).

La relazione sarà suddivisa in due capitoli principali. Nel primo capitolo analizzerò i dati forniti in modo più dettagliato ed eseguirò strategie univariate, in cui viene considerato un solo fattore alla volta, seguite da un'analisi dei risultati. Nel secondo capitolo, presenterò diversi tipi di strategie multivariate che utilizzano più fattori. Inoltre, proporrò la mia soluzione per risolvere il problema della correlazione tra le strategie univariate e analizzerò i risultati.

Capitolo 1: Strategie univariate

In questo capitolo, fornirò una descrizione più dettagliata dei dati forniti, della metodologia utilizzata per l'elaborazione e l'analisi dei dati. Successivamente, eseguirò strategie univariate su 10 fattori (per identificare i migliori) e presenterò i risultati.

1.1 Descrizione della metodologia e fattori utilizzati.

1.1.1 Metodologia

Il set di dati è composto da 40 fattori per 797 titoli azionari. Tuttavia, alcuni di questi titoli hanno dati parziali, ovvero mancano alcuni fattori o valori di prezzo per date specifiche. Pertanto, ho effettuato dei controlli per escludere tali fattori dal portafoglio.

Dopo un'iniziale elaborazione e pulizia dei dati (descritta nell'Appendice), ho spostato e trasformato i fattori per eliminare la distorsione da look-ahead e per assicurarmi che i valori più alti dei fattori siano migliori. In particolare, ho spostato i fattori relativi ai rendimenti e al prezzo di 1 mese e quelli relativi ai fondamentali della società di 3 mesi. Poi, nel processo di trasformazione, i fattori sono stati classificati in tre gruppi in base alla loro desiderabilità:

- Fattori che sono migliori quando sono più alti, che sono stati lasciati invariati.
- Fattori che sono migliori se più bassi, per i quali è stato utilizzato l'inverso del fattore.
- Fattori che sono migliori quando sono inferiori a un certo valore, per il quale è stata utilizzata la differenza tra il fattore e il valore target.

Le strategie univariate sono implementate come segue:

- Per ogni data, classifico i titoli disponibili in base a un determinato fattore (prendendo i migliori trasformati) e prendo i 30 titoli migliori da includere nel portafoglio.
- Quindi calcolo i rendimenti del portafoglio per ogni data.
- Dopo aver estratto lo stato del portafoglio in ogni periodo, lo utilizzo per correggere i rendimenti con le commissioni.
- Calcolo quindi le metriche e gli indici per ogni strategia e li confronto con il benchmark (composto da 800 titoli equamente ponderati) per valutare la performance della strategia.

1.1.2 Fattori e metriche utilizzati.

Per l'implementazione di strategie univariate, inizialmente ho scelto un numero elevato di fattori (combinando sia quelli tecnici che quelli fondamentali). I risultati ottenuti mi hanno aiutato a determinare un insieme più ridotto di fattori da utilizzare nelle strategie multivariate. I fattori analizzati sono i seguenti:

- RSI_14D: Oscillatore del momentum che misura la velocità e la variazione dei movimenti di prezzo.
- PE_RATIO: metrica di valutazione che indica la disponibilità degli investitori a pagare per gli utili.
- PX_TO_BOOK_RATIO: confronta il valore di mercato di una società con il suo valore contabile, indicando la valutazione relativa.
- BEST_EPS: il più alto valore di utile per azione riportato per un'azione.
- WACC_COST_EQUITY: misura del costo del capitale di un'azienda da fonti di capitale proprio.
- MOV_AVG_30D: Media mobile a 30 giorni che attenua le fluttuazioni a breve termine.
- EBITDA_MARGIN: rapporto tra gli utili prima degli interessi, delle imposte, degli ammortamenti e delle svalutazioni e il fatturato totale, che indica la redditività.
- DEBITO NETTO PER AZIONE: Misura della leva finanziaria di una società, calcolata come debito netto diviso per il numero di azioni in circolazione.
- NORMALIZED_ACCRUALS_CF_METHOD: rettifiche apportate agli utili dichiarati per riflettere la reale performance economica dell'azienda.
- VOLATILITÀ_90D: Misura dell'oscillazione del prezzo di un titolo negli ultimi 90 giorni, che indica il rischio di mercato.

Una delle metriche più importanti utilizzate per valutare la performance delle strategie è l'information ratio. L'information ratio è una misura del rendimento di un portafoglio superiore a quello di un benchmark, corretto per il rischio aggiuntivo assunto. Si calcola come segue:

$$IR = \frac{R_P - R_B}{\sigma_P}$$

Dove:

- IR è il rapporto di informazione.
- R_P è il rendimento del portafoglio.
- R_B è il rendimento del benchmark.

- σ_p è l'errore di inseguimento.

Ho valutato ogni strategia utilizzando una serie di parametri e indici:

- Rapporto informativo: Il rapporto più importante che contribuisce a determinare l'eccesso di rendimento (rispetto al benchmark) ottenuto selezionando quel fattore.
- Rendimenti: Rendimenti del portafoglio per ogni periodo, compresi i seguenti calcoli:
 - Media: Rendimento medio per ogni periodo.
 - STD: deviazione standard, utile come misura del rischio.
 - STD negativo: deviazione standard dei rendimenti negativi.
 - Totale: rendimento totale del portafoglio nell'intero periodo.
- Alfa: Eccesso di rendimento del portafoglio rispetto al benchmark, con i seguenti calcoli:
 - Media: Media degli alfa per ogni periodo.
 - Totale: alfa totale generato dal portafoglio.
- Rendimento corretto per il rischio: Rendimento corretto per il rischio, calcolato come Rendimento / Rischio.
- Rapporto di Sharpe: Sharpe ratio, calcolato come il rendimento corretto per il rischio al netto del tasso privo di rischio. Esprime il rischio complessivo.
- Beta: Il beta del portafoglio rispetto al benchmark esprime il rischio sistematico.
- Treynor Ratio: calcolato come il rendimento al netto del tasso privo di rischio, corretto per il beta del portafoglio. Esprime il rischio sistematico.
- Sortino Ratio: come lo Sharpe ratio, ma considera solo il rischio di ribasso (cioè una deviazione standard negativa).

Nota: le strategie sono valutate con e senza commissioni (0,4% per un'operazione di vendita).

1.1.3 Benchmark

Il benchmark è costituito da 800 titoli equamente ponderati. Viene utilizzato per confrontare le performance delle strategie e per valutare i risultati.

Di seguito sono riportate le statistiche e i rendimenti del benchmark:

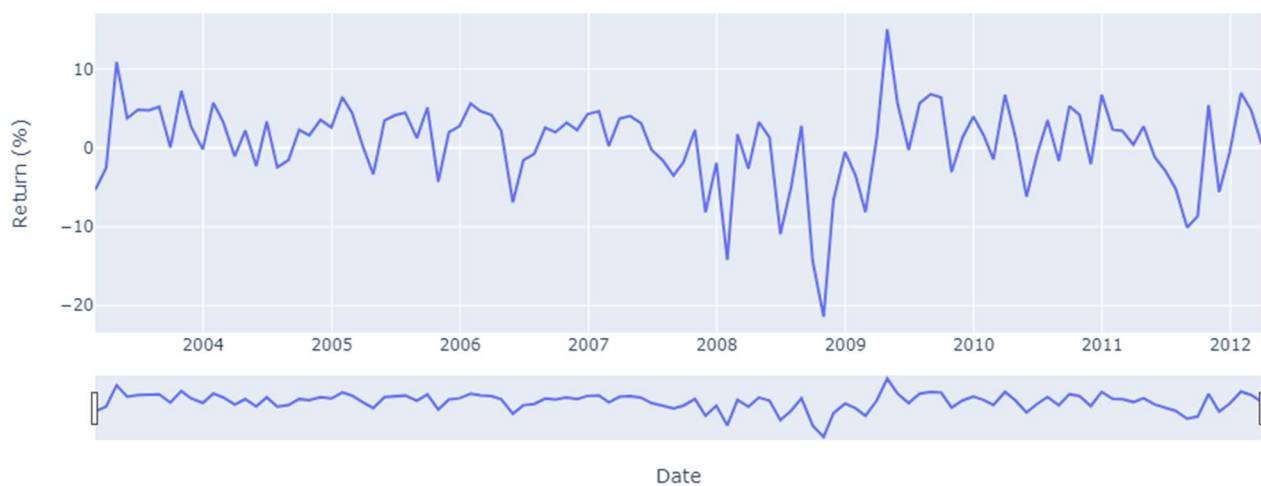


Figura 1: Rendimento logaritmico del benchmark (per ogni mese)

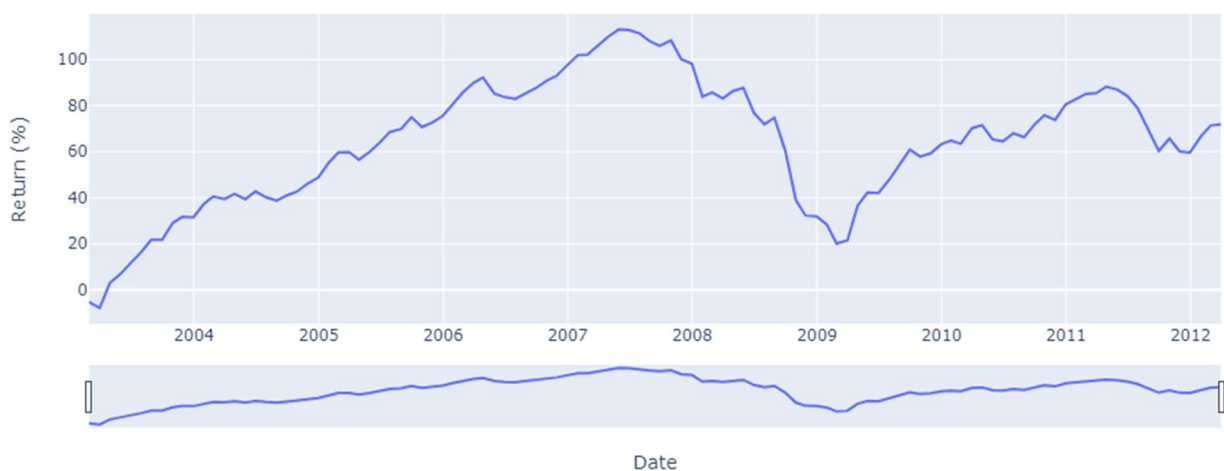


Figura 2: Rendimento composto del Benchmark (sull'intero periodo)

Statistics	Value
Return AVG (%)	7.8443
Return STD (%)	18.1806
Downside return STD (%)	14.5467
Return Total (%)	71.9063
Return Risk Adj (%)	43.1466
Sharpe Ratio	0.0617
Sortino Ratio	0.2671

Figura 3: Statistiche del benchmark

1.2 Analisi complessiva dei risultati per tutte le strategie univariate

Possiamo vedere i risultati delle strategie univariate esaminando il rapporto di informazione.

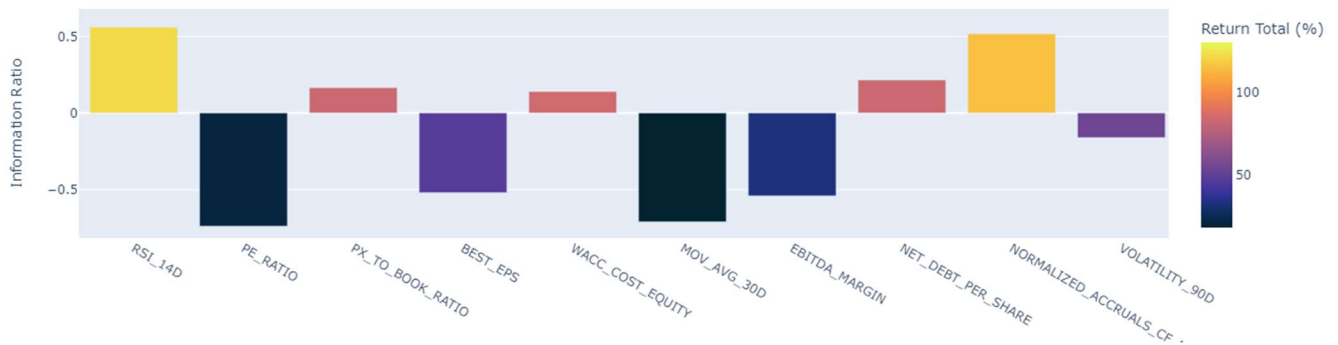


Figura 4: Indice di informazione per strategie univariate (con commissioni)

Possiamo notare che le migliori strategie sono RSI_14D, NORMALIZED_ACCRUALS_CF, NET_DEBT_PER_SHARE, PX_TO_BOOK_RATIO e WACC_COST_EQUITY (per le altre cinque è negativo). Queste strategie hanno un rapporto informativo positivo, da 0,56 a 0,14, il che significa che hanno un rendimento superiore al benchmark a parità di rischio assunto. Ma dobbiamo considerare che non è abbastanza alto per essere considerato una strategia solida (come ci aspettavamo), poiché questi risultati non sono statisticamente significativi e potrebbero essere dovuti alla fortuna.

La performance di questi portafogli può essere osservata anche analizzando gli alfa.



Figura 5: Strategie univariate vincenti, Alpha ad ogni mese.

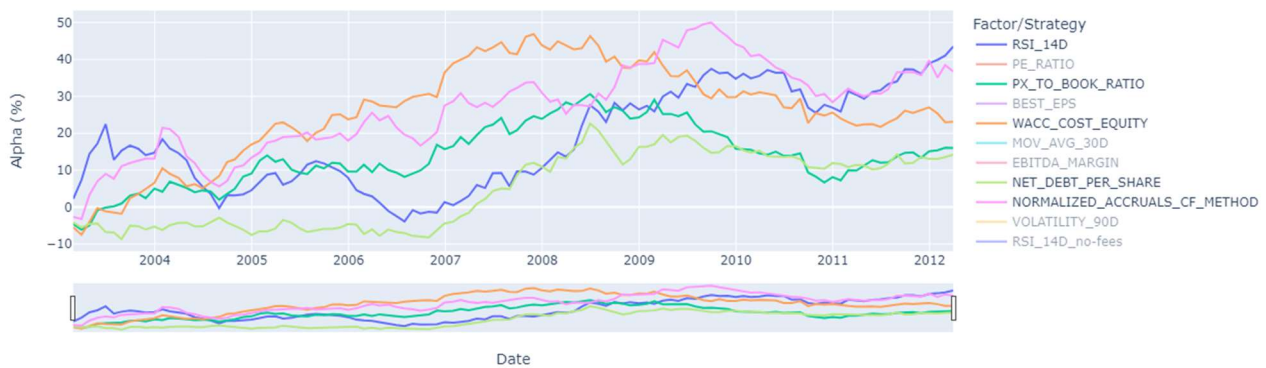


Figura 6: Strategie vincenti univariate, Alpha composto.

Strategies/Factors	Infor... Ratio ↓	Return AVG (%)	Return STD (%)	Downside Return STD (%)	Return Total (%)	Alpha AVG (%)	Alpha Total (%)	Return Risk Adj (%)	Sharpe Ratio	Beta	Treynor Ratio	Sortino Ratio
RSI_14D	0.5604	13.3361	23.5328	17.9864	122.2479	4.7441	43.4878	56.6705	0.115	1.1925	0.0786	0.5213
NORMALIZED_ACCRUALS_	0.5169	12.576	23.1969	17.3362	115.2802	4.0043	36.7062	54.2142	0.1072	1.1872	0.0726	0.497
NET_DEBT_PER_SHARE	0.2153	9.0891	17.6789	14.4545	83.3169	1.5469	14.1799	51.4122	0.0838	0.9222	0.0556	0.3549
PX_TO_BOOK_RATIO	0.1657	9.0008	16.7167	13.0199	82.5076	1.743	15.9775	53.8432	0.0871	0.849	0.0594	0.3872
WACC_COST_EQUITY	0.1397	9.167	14.7014	11.3305	84.0313	2.5216	23.1151	62.3548	0.1023	0.6914	0.0753	0.4596
VOLATILITY_90D	-0.1609	5.9682	11.5833	9.4188	54.7083	0.0766	0.7026	51.5238	0.0501	0.4973	0.0404	0.2133
BEST_EPS	-0.5199	5.1905	17.8932	14.0844	47.5797	-2.4398	-22.3644	29.0083	0.0199	0.9449	0.013	0.0874
EBITDA_MARGIN	-0.5412	3.6144	15.0978	12.0803	33.1317	-3.2681	-29.9576	23.9397	-0.0066	0.7524	-0.0046	-0.0286
MOV_AVG_30D	-0.7103	2.0341	15.1495	12.6661	18.6456	-4.8234	-44.2144	13.4266	-0.0367	0.746	-0.0258	-0.152
PE_RATIO	-0.7389	2.2814	21.9859	17.0704	20.9131	-6.128	-56.1731	10.3768	-0.022	1.1455	-0.0146	-0.0983

Figura 7: Strategie univariate, statistiche ordinate per Information Ratio.

Dai risultati è evidente che le strategie vincenti riescono a sconfiggere il benchmark, fornendo un buon alfa sia per il fattore tecnico che per quello fondamentale. Tuttavia, come si può notare dalle statistiche, le strategie vincenti hanno una deviazione standard più elevata e un drawdown più alto, il che significa che sono più rischiose del benchmark; ciò è confermato dallo Sharpe ratio che è piuttosto basso. Possiamo anche osservare che (nonostante un IR più basso) la strategia WACC_COST_EQUITY ha un alfa più alto della terza migliore strategia, ma con una volatilità più bassa; quindi, fornisce un miglior rendimento aggiustato per il rischio.

1.3 Analisi delle 3 migliori strategie

Possiamo approfondire le strategie migliori esaminando lo stato del portafoglio durante ogni periodo; in dettaglio, per questa parte prenderò in considerazione le due strategie migliori: RSI_14D (RSI), NORMALIZED_ACCRUALS_CF (NAC) e WACC_COST_EQUITY (WCE), per le ragioni illustrate in precedenza.

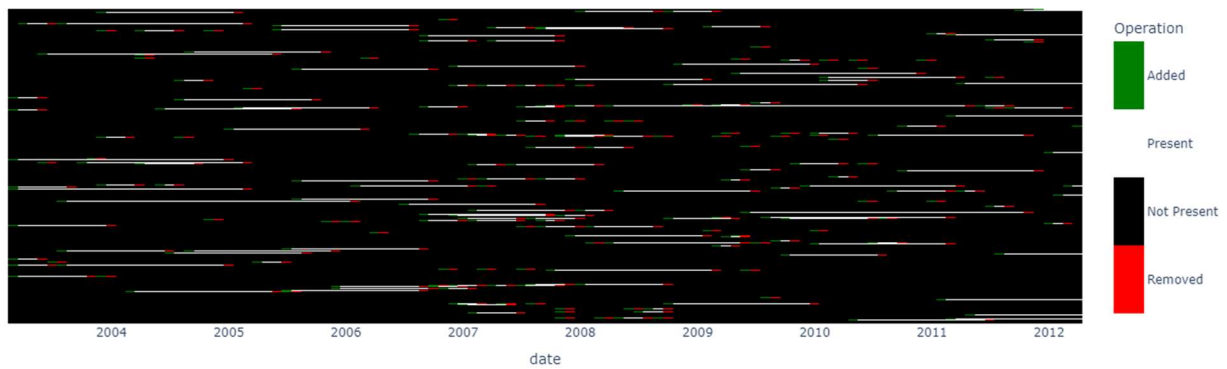


Figura 8: Stato del portafoglio per RSI_14D

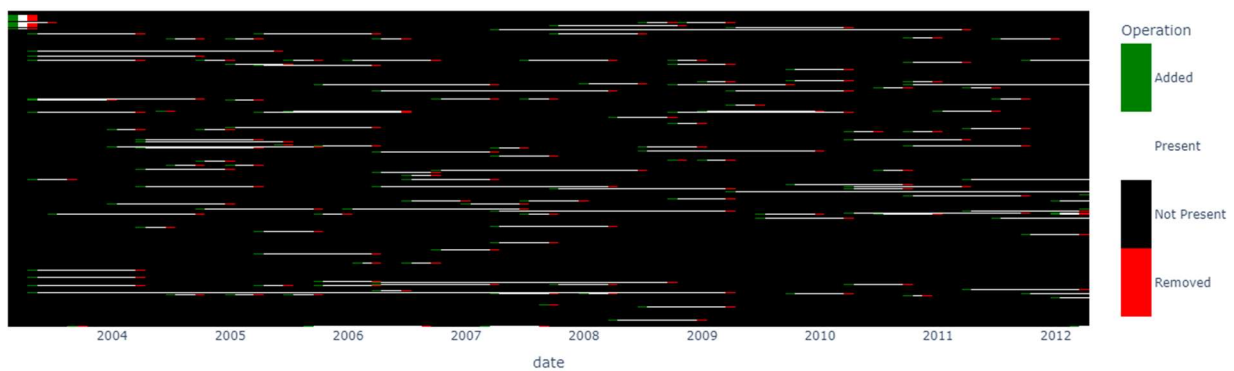


Figura 9: Stato del portafoglio per NORMALIZED_ACCRUALS_CF

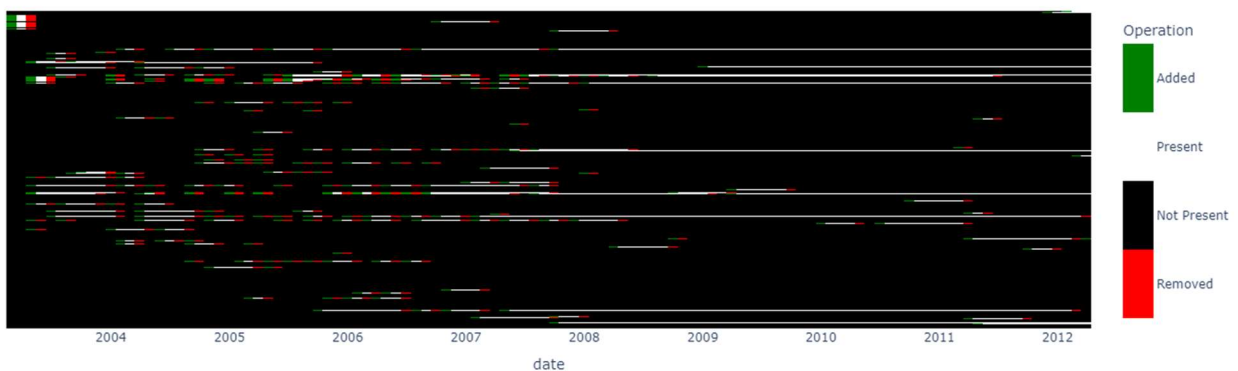


Figura 10: Stato del portafoglio per WACC_COST_EQUITY

Osservando lo stato del portafoglio, possiamo notare che RSI ha un turnover elevato (tipico delle strategie tecniche), mentre NAC e WCE hanno un turnover inferiore. Questo è un buon segno, poiché un turnover elevato significa commissioni più elevate e un rendimento inferiore. Possiamo anche notare che RSI e NAC hanno un turnover stabile, mentre WCE ha un turnover elevato all'inizio e poi quasi nullo; ciò è probabilmente dovuto al fatto che la maggior parte dei dati mancanti si trova all'inizio del periodo, quindi la strategia non è in grado di prendere una posizione forte.

Capitolo 2: Strategie multivariate

In questo capitolo presenterò diversi tipi di strategie multivariate che utilizzano più fattori. Inoltre, proporrò la mia soluzione per risolvere il problema della correlazione tra le strategie univariate e analizzerò i risultati.

Nota: le tasse sono considerate nelle strategie multivariate.

2.1 Screening sequenziale

La prima strategia multivariata che presenterò è il Sequential Screening. Questa strategia si basa sull'idea di selezionare i titoli migliori utilizzando una sequenza di fattori. I fattori vengono utilizzati per classificare i titoli e i titoli migliori vengono selezionati per il portafoglio. I fattori vengono utilizzati in un ordine specifico e i titoli vengono selezionati in base alla classifica del fattore precedente.

I fattori utilizzati nello screening sequenziale sono i migliori 4 fattori delle strategie univariate, utilizzati in ordine di IR (dal peggiore al migliore); in dettaglio:

- All'inizio utilizziamo NET_DEBT_PER_SHARE per selezionare i 400 titoli migliori.
- Quindi utilizziamo il PX_TO_BOOK_RATIO per selezionare i 200 titoli migliori.
- Quindi utilizziamo NORMALIZED_ACCRUALS_CF per selezionare i 100 titoli migliori.
- Alla fine utilizziamo RSI_14D per selezionare i 30 titoli migliori.

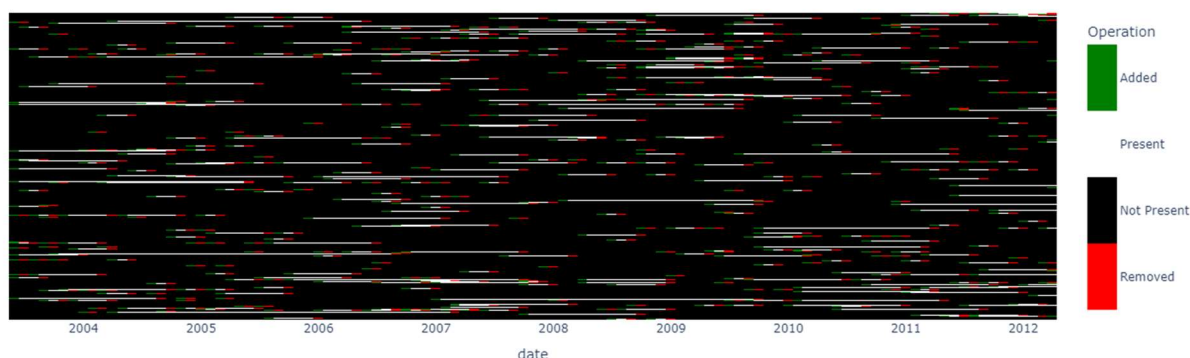


Figura 11: Stato del portafoglio per lo screening sequenziale

Il risultato dello screening sequenziale è un portafoglio con un elevato turnover, dovuto all'elevato turnover dell'RSI, che fornisce un IR leggermente superiore alla migliore strategia univariata, con metriche comparabili e alfa e beta leggermente superiori.

2.2 Screening simultaneo con ZScore

La seconda strategia multivariata che presenterò è lo Screening simultaneo con zscore, che si basa sull'idea di aggregare diversi fattori in uno solo, per poi selezionare i titoli migliori in base al ranking del fattore aggregato. I fattori utilizzati nello screening simultaneo sono i migliori 4 fattori delle strategie univariate.

2.2.1 Semplice

La prima versione del Simultaneous Screening with zscore è la versione semplice, in cui calcolo la media, in ogni periodo, dello zscore dei fattori vincenti; quindi, classifico ogni azione in base alla media dello zscore e seleziono i 30 titoli migliori.

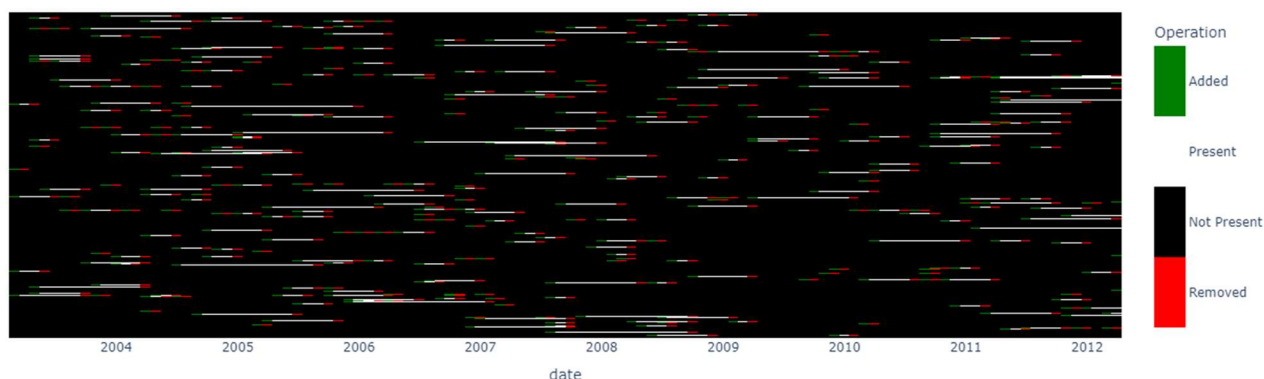


Figura 12: Stato del portafoglio per lo screening semplice con zscore

Il risultato dell'impiego del metodo di screening z-score diretto è un portafoglio con un turnover ancora più elevato rispetto allo screening sequenziale, ma migliore sotto ogni altro aspetto, con un IR più elevato, metriche migliori e un alfa più elevato e un beta più basso.

2.2.2 Ponderato

In questa versione ho cercato di risolvere il problema della correlazione dei fattori utilizzando una media ponderata dello zscore, dove i pesi sono l'inverso della correlazione tra i fattori. In dettaglio, i pesi sono calcolati calcolando la matrice di correlazione dei fattori (zscore) su una finestra mobile di 12 mesi, poi calcolo la media per ogni fattore e alla fine utilizzo l'inverso della media assoluta come pesi.

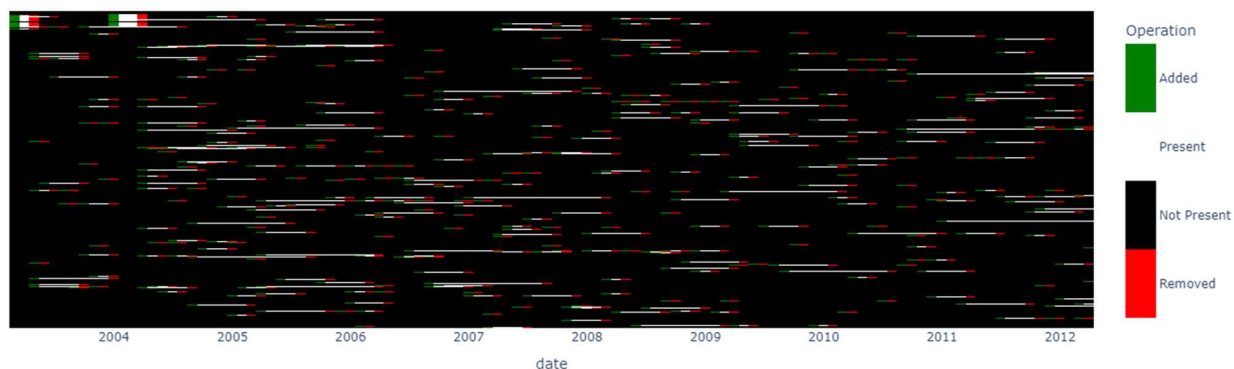


Figura 13: Stato del portafoglio dello screening ponderato per lo zscore

Come si può vedere dai risultati, ha un turnover leggermente inferiore rispetto alla versione semplice e, data l'instabilità dei pesi all'inizio, assume posizioni strane all'inizio; ma, dopo qualche mese, si stabilizza e fornisce un alfa leggermente superiore con un IR leggermente inferiore.

2.3 Analisi complessiva dei risultati per tutte le strategie multivariate

Possiamo vedere i risultati delle strategie multivariate esaminando l'information ratio, l'alfa raggiunto e le statistiche.

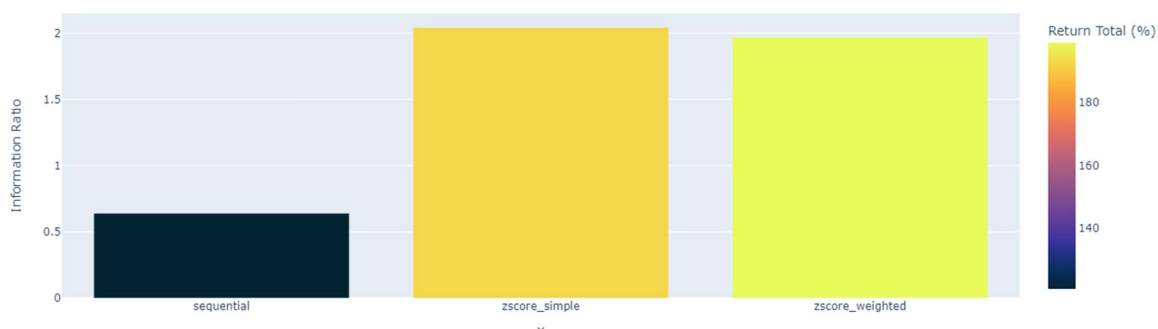


Figura 14: Rapporto informativo delle strategie multivariate



Figura 15: Alfa composto delle strategie multivariate

Strategies/Factors	Infor... Ratio ↓	Return AVG (%)	Return STD (%)	Downside Return STD (%)	Return Total (%)	Alpha AVG (%)	Alpha Total (%)	Return Risk Adj (%)	Sharpe Ratio	Beta	Treynor Ratio	Sortino Ratio
zscore_simple	2.043	20.9928	18.4625	14.1511	192.4337	13.3312	122.2024	113.705	0.2663	0.953	0.1787	1.2037
zscore_weighted	1.9701	21.7109	20.3453	15.4492	199.0168	13.6677	125.2869	106.7122	0.2519	1.0512	0.1689	1.149
sequential	0.6408	13.5252	23.6731	17.8589	120.5993	4.9991	44.5749	57.1329	0.1167	1.2121	0.0789	0.5358

Figura 16: Statistiche delle strategie multivariate

Osservando le performance, è evidente che le strategie z-score superano lo screening sequenziale, che ha un turnover inferiore, ma il suo alfa e il suo IR sono molto più bassi rispetto alle strategie z-score. In confronto, la versione semplice della strategia z-score ha un IR leggermente superiore alla versione ponderata e un alfa leggermente inferiore, ma sono abbastanza simili in tutti gli altri aspetti (tranne che per il beta che è più alto per la ponderata).

Possiamo anche osservare le metriche di rischio (come il rendimento aggiustato per il rischio, il rapporto Sharpe, Sortino e Treynor) che confermano che le strategie z-score sono migliori dello screening sequenziale.

Conclusione

In questo progetto, ho sviluppato un modello di fattori basato sul dataset fornito "Euro.xls". Ho eseguito strategie univariate su 10 fattori e ho presentato i risultati. Successivamente, ho presentato diversi tipi di strategie multivariate che utilizzano più fattori e ho proposto la mia soluzione per risolvere il problema della correlazione tra le strategie univariate. Ho anche analizzato i risultati e confrontato le prestazioni delle strategie per valutare i risultati.

Si può concludere che le migliori strategie univariate sono RSI_14D, NORMALIZED_ACCRUALS_CF, NET_DEBT_PER_SHARE, PX_TO_BOOK_RATIO e WACC_COST_EQUITY. Queste strategie presentano un rapporto informativo positivo, da 0,56 a 0,14, il che significa che hanno un rendimento superiore al benchmark a parità di rischio assunto. Tuttavia, i risultati non sono sufficientemente elevati per essere considerati una strategia solida.

Ho utilizzato questi fattori per creare strategie multivariate e ho scoperto che le migliori strategie multivariate sono le versioni semplice e ponderata dello Screening simultaneo con zscore, che superano lo screening sequenziale. Lo zscore semplice e quello ponderato sono abbastanza simili in termini di rischio e di performance, ma se eliminiamo i primi mesi, la versione ponderata è migliore. Lo screening sequenziale ha un turnover inferiore, ma la sua performance è inferiore a quella delle strategie z-score.

In futuro potremmo provare a utilizzare una combinazione di fattori diversa per le strategie multivariate, invece di assumere che il fattore migliore per le strategie univariate sia anche il migliore per quelle multivariate, e potremmo anche provare ad affrontare il problema della correlazione in modi diversi (ad esempio utilizzando la PCA). Un altro possibile sviluppo, con dati più frequenti, potrebbe essere l'utilizzo di un modello di apprendimento automatico per imparare la migliore combinazione di fattori per le strategie multivariate.

Appendice: Dettagli tecnici

Il codice utilizzato per questa relazione è scritto in Python ed è suddiviso in tre parti principali: `dataset_extraction.ipynb` responsabile dell'elaborazione iniziale dei dati; `analysis.ipynb` per l'analisi delle strategie; `app.py` per la visualizzazione dei risultati.

Nel file `dataset_extraction.ipynb` ho estratto i dati dal dataset fornito "Euro.xls" (rinominato "data_initial.xlsx") e li ho trasformati in un DataFrame di pandas (multi-indicizzato per data e azioni, e fattori come colonne) senza perdere o modificare alcuna informazione. Questa parte è importante, perché nel file di analisi cambio i fattori da utilizzare senza alcuna difficoltà.

Nel file `analysis.ipynb` ho eseguito le strategie univariate e multivariate e poi ho valutato i risultati. Per essere efficiente dal punto di vista computazionale ho utilizzato molte operazioni vettoriali (da pandas e numpy) e la libreria numba; tuttavia, questo rende il codice meno comprensibile a prima vista, quindi ho separato il codice in diverse funzioni e ho aggiunto molti commenti per renderlo più leggibile. Ogni risultato viene salvato nella cartella "data/output" su più file csv, da utilizzare nell'applicazione.

Nel file `app.py`, ho utilizzato le librerie Dash e Plotly per creare un'applicazione web interattiva per visualizzare i risultati (non era necessario, ma facilita l'analisi dei risultati); per eseguire l'applicazione, è necessario installare le librerie richieste (`requirements.txt`) ed eseguire il file.

Esiste anche un quarto file, `offline_plot.ipynb`, che viene utilizzato per generare tracciati simili a quelli dell'applicazione, ma in un ambiente notebook, per eliminare la necessità di eseguire l'applicazione per visualizzare i risultati (anche se l'applicazione è consigliata).