LUISS

Corso di laurea in Economia e Finanza

Cattedra di Econometria per la Finanza

Prinicpal Portfolios:
Un' analisi sui 25 portafogli di Fama e French

Prof. Federico Carlini

RELATORE

Prof. Nicola Borri

CORRELATORE

Viola Borzillo Matr.755471

CANDIDATO

Anno Accademico 2022/2023

Indice

Capitolo 1

Introduzione

Nel seguente lavoro di tesi verrà effettuata un'analisi approfondita sui diversi fattori che nell'attuale realtà dell'asset pricing hanno un risk premia maggiore. In particolare, i fattori su cui si baserà l'analisi sono i primi 3 dei 25 portafogli dinamici, chiamati "principal portfolios", introdotti dal paper "principal portfolios" scritto da Bryan Kelly, Semyon Malamud e Lasse Heje Pedersen e i primi 3 portafogli "statici", calcolati con il metodo noto della PCA. La motivazione della scelta di questo tema è data dal fatto che, il paper appena menzionato, introduce una modalità alternativa, interessante e semplice di fare portafogli dinamici. Per portafogli dinamici si intende un insieme di attività finanziarie (come azioni, obbligazioni, titoli di Stato, ecc.) che viene regolarmente aggiornato e riequilibrato nel tempo per tenere conto dei cambiamenti nel mercato e delle strategie di investimento dell'investitore. In altre parole, i portafogli dinamici vengono adattati per rispecchiare le aspettative di rendimento e di rischio dell'investitore, nonché le condizioni di mercato prevalenti. L'obiettivo del lavoro di tesi è quello di analizzare entrambe le modalità di asset pricing, la modalità basata su valutazioni dinamica e quella basata su valutazioni statiche, attraverso la principal component analysis.Per valutazione statica si intende un'analisi dell'asset pricing che assume che le condizioni di mercato rimangano costanti nel tempo. In questo contesto, si presume che le variabili rilevanti per la valutazione degli asset finanziari (come i tassi di interesse, i rendimenti attesi, la volatilità, ecc.) rimangano fisse durante l'orizzonte temporale considerato, questo approccio semplifica l'analisi. La metodologia utilizzata per la seguente analisi è avvenuta attraverso il linguaggio di programmazione Python, tutto il codice è inserito nell'appendice 1. In python è stata effettuata l'analisi dinamica dei portafogli, applicando ai 25 portafogli Fama and French, size e book to market, una valutazione dinamica e successivamente una statica, poi il confronto tra essi con la regressione fama Macbeth.

L'elaborato è stato strutturato in tre capitoli, nel primo capitolo vi è la spiegazione delle parti principali del paper "Principal Portfolios", evidenziando i passaggi fondamentali utilizzati nell'analisi. Come secondo capitolo vi è la spiegazione del metodo della PCA, SVD e il confronto fra i due. Infine, come terzo capitolo vi è il confronto fra i due metodi di asset pricing, con particolare riguardo al procedimento seguito, le motivazioni e i risultati raggiunti.

Capitolo 2

Paper Principal Portfolios

ll paper "Principal Portfolios" [1] è incentrato sull'idea di utilizzare la teoria dei portafogli per identificare e isolare i fattori principali che guidano i rendimenti degli asset finanziari. Il paper dimostrerà che i principal portfolios possono essere utilizzati per spiegare la cross section dei returns e i risk premia. È un nuovo asset pricing framework nel quale vengono utilizzati tutti i segnali dei titoli che consentono di prevedere ogni singolo ritorno. Mentre la letteratura si concentra sulla prevedibilità del segnale di ciascun titolo, assumendo una forza uguale tra i titoli, il framework proposto è flessibile e include la prevedibilità incrociata, che porta a tre risultati principali:

- I "principal portfolios" catturano una grande parte della variazione dei rendimenti degli asset finanziari, il che suggerisce che esistono pochi fattori principali che guidano i rendimenti di un ampio universo di asset

- I "principal portfolios" hanno un'alta capacità predittiva sui rendimenti futuri degli asset finanziari, il che suggerisce che la conoscenza dei fattori principali può essere utilizzata per effettuare previsioni sul rendimento degli asset

- I "principal portfolios" sono in grado di spiegare gran parte dei risk premia osservati sul mercato, il che suggerisce che la conoscenza dei fattori principali può essere utilizzata per comprendere le fonti del rischio e del rendimento nel mercato finanziario

In sintesi, il paper "Principal Portfolios" propone un nuovo framework di asset pricing che utilizza la teoria dei portafogli per identificare e isolare i fattori principali che guidano i rendimenti degli asset finanziari, fornendo così una migliore comprensione delle fonti di rischio e di rendimento nel mercato finanziario. In primo luogo, viene derivata la strategia ottimale in forma chiusa. Questa strategia consiste in una matrice di predizione di autovettori, che chiamiamo "portafogli principali". In secondo luogo, scomponiamo il problema in alfa e beta, producendo strategie ottimali con, rispettivamente, zero e l'esposizione del fattore positivo. In terzo luogo, forniamo un nuovo test dei modelli di asset pricing. Empiricamente, i principal portfolios forniscono alfa significativi fuori campione rispetto a fattori standard in diversi dataset. I punti principali del paper includono:

- L'autore propone un metodo per identificare i fattori principali attraverso una tecnica di decomposizione dei rendimenti in una serie di componenti principali. Questi componenti principali rappresentano le fonti primarie di rischio e rendimento dei portafogli

- Utilizzando questa tecnica, l'autore dimostra che i fattori principali dei rendimenti sono altamente persistenti nel tempo, il che suggerisce che gli investitori possono trarre vantaggio dalla loro identificazione e sfruttamento

- L'autore suggerisce anche che i portafogli di fattori principali possono essere utilizzati per costruire portafogli efficienti e diversificati che offrono un'esposizione altamente concentrata ai fattori principali che guidano i rendimenti degli asset finanziari

- Infine, il paper dimostra che i portafogli di fattori principali possono essere utilizzati per creare strategie di investimento a basso costo, che possono essere implementate utilizzando semplici strumenti di investimento come gli ETF

In sintesi, il paper offre un approccio innovativo per identificare e sfruttare i fattori principali dei rendimenti degli asset finanziari, offrendo ai investitori un modo per costruire portafogli efficienti e diversificati. Nel paper "Principal Portfolios" gli autori esaminano l'importanza dei momentum come uno dei fattori principali che guidano i rendimenti degli asset finanziari. Il momentum è un fenomeno osservato nei mercati finanziari in cui gli asset che hanno avuto una performance positiva in passato tendono ad avere una performance positiva anche in futuro, e viceversa per gli asset con una performance negativa. Kelly dimostra che il momentum è un fattore principale attraverso la tecnica di decomposizione dei rendimenti in componenti principali. In particolare, egli identifica uno dei componenti principali come il fattore momentum, che rappresenta la fonte primaria di rischio e rendimento per gli asset finanziari. Inoltre, gli autori suggeriscono che i portafogli di momentum possono essere utilizzati per costruire strategie di investimento altamente redditizie. Infatti, dimostreranno che l'utilizzo di portafogli di momentum ha portato a rendimenti significativamente più elevati rispetto all'utilizzo di un portafoglio diversificato. Tuttavia, il momentum è un fattore che presenta una certa volatilità e che non funziona sempre in modo uniforme su tutti gli asset finanziari. Pertanto, l'utilizzo di strategie basate sul momentum richiedono una rigorosa gestione del rischio e una diversificazione adeguata per evitare di esporre troppo il portafoglio ai movimenti del mercato. Il punto di partenza per gran parte degli asset pricing è un segnale, $S_{i,t}$ che approssima il valore atteso titolo I al tempo t. Nel contesto di un modello di asset pricing, $S_{i,t}$ può rappresentare il beta condizionale rispetto al mercato (o il prezzo di kernel, ossia un fattore di sconto stocastico). Oppure può essere un predittore agnostico delle considerazioni di equilibrio, un segnale o un indicatore che non tiene conto di specifiche considerazioni di equilibrio, come il rapporto di valutazione di ciascun asset o il momento del prezzo recente. Le analisi standard, come la valutazione dei portafogli ordinati per caratteristiche o i test di asset pricing nel senso di Gibbons et al. (1989), si concentrano sui segnali predittivi propri di ciascun asset. In altre parole, l'attenzione è rivolta all'associazione tra il segnale $S_{i,t}$ e il rendimento solo dell'asset "i", rappresentato come $R_{i,t+1}$. Questo significa che invece di considerare fattori o segnali che coinvolgono l'intero universo degli asset o le relazioni tra i prezzi degli asset, l'analisi si concentra solo sulle relazioni tra un particolare segnale $S_{i,t}$ e il rendimento dell'asset specifico "i". Questo approccio è più focalizzato sulle caratteristiche specifiche di ciascun asset e sulla loro relazione con i rendimenti futuri. Ad esempio, invece di considerare l'andamento complessivo del mercato azionario o di un determinato settore, si analizzano le caratteristiche individuali di un singolo titolo, come il suo rapporto prezzo-utili o il suo momento del prezzo, per prevedere il suo rendimento futuro. Questo approccio di analisi delle caratteristiche specifiche di ciascun asset può essere utile per comprendere meglio il ruolo di tali fattori nell'asset pricing e può consentire di identificare relazioni tra segnali specifici e rendimenti che potrebbero non emergere da analisi più ampie sulle dinamiche di mercato o considerazioni di equilibrio. La letteratura attuale e quindi l'analisi standard, come la valutazione dei portafogli ordinati in modo caratteristico o test di determinazione dei prezzi degli asset nello spirito di Gibbons et al. (1989), si concentra sui segnali predittivi degli asset propri; cioè, l'associazione tra $S_{i,t}$ e il ritorno solo sull' asset i, $R_{i,t+1}$. In questo paper viene proposto un nuovo approccio all'analisi dei prezzi degli asset attraverso la "matrice di previsione"o "prediction matrix". Questo approccio mostra come investire in modo ottimale alla luce della cross prevedibilità, per cross-prevedibilità si intende la capacità di utilizzare informazioni da un mercato o asset finanziario per fare previsioni accurate su un altro mercato o asset finanziario correlato, dove ottimale si riferisce alla strategia di massimizzazione del rendimento tra una classe di strategie lineari di trading, come trovare strategie alfa e beta ottimali e un nuovo test di asset pricing model. In questo approccio, la prediction matrix corrisponde a:

$$\Pi = \mathbb{E}(R_{i,t+1}S_t') \in \mathbb{R}^{N\times N}, \quad \text{dove } R_{t+1} = (R_{i,t+1})_{i=1}^N \in \mathbb{R}^N \tag{2.1}$$

e il vettore dei ritorni è $S_t = (S_{i,t})_{i=1}^N \in \mathbb{R}^N$ La parte diagonale della matrice di previsione traccia la previsione degli effetti del proprio segnale, che sono al centro dei prezzi standard degli asset. Ad esempio, se il segnale $S_{i,t}$ rappresenta il momentum di ogni asset, quindi:

$$\Pi_{i,i} = \mathbb{E}(R_{i,t+1}S_{i,t})$$

è il profitto atteso dell'asset di trading basato sul suo segnale che in questo caso, è rappresentato dal momentum. In altre parole, pensiamo al segnale $S_{i,t}$ come al portfolio holding, ossia come alla singola posizione o investimento che compongono un portafoglio di investimenti, per esempio, se consideriamo i 25 portafogli Fama-french il portfolio holding è rappresentato dal segnale relativo al singolo portafoglio, e $R_{i,t+1}S_{i,t}$ come rendimento corrispondente. È importante sottolineare che la matrice di previsione tiene traccia anche dei fenomeni di cross-prevedibilità. Infatti, la parte fuori diagonale della matrice di previsione, $\Pi_{i,j} = \mathbb{E}(R_{i,t+1}S_{j,t})$, mostra come il segnale dell'asset j prevede il rendimento dell'asset I. La cross-prevedibilità esiste molto generalmente nelle attività condizionali modelli di pricing, siano essi di natura equilibrata o puramente statistici. La conoscenza dell'intera matrice di previsione, al contrario della tipica attenzione ai soli elementi diagonali, è fondamentale per ideare portafogli ottimali e comprendere il trade-off rischio-rendimento. Il contributo principale del paper è quello di sviluppare una vasta comprensione teorica della matrice di previsione e delle informazioni sui prezzi degli asset che porta. I principali strumenti di questa analisi sono SVD, analoghe all'utilizzo dell'analisi delle componenti principali (PCA) per studiare le matrici varianza-covarianza. I vettori singolari di Π, noti come "portafogli principali", sono un insieme di portafogli normalizzati ordinati da quelli più prevedibili da S a quelli meno prevedibili. Il top dei portafogli principali sono quelli che offrono i rendimenti più alti e, come tali, offrono rendimenti attesi non condizionali per un investitore che deve affrontare un vincolo di leva finanziaria (uno che non può ricoprire posizioni infinitamente grandi). Una delle intuizioni chiave dell'approccio del paper, è che applicando una SVD direttamente alla matrice di predizione (dei momenti) confonde due fenomeni economici diversi e opposti. Quindi, il paper propone una prima suddivisione Π in una parte simmetrica (che è uguale alla sua trasposizione e indicata con Π_s e una parte antisimmetrica (che è uguale a meno la sua trasposizione e indicata con Π_a e applicando una decomposizione separata alle due matrici Π_s e Π_a, analizzando e scomponendo ogni matrice in modo indipendente, senza considerare la loro relazione o eventuali interazioni. Questo approccio ci consente di evidenziare le caratteristiche individuali, le proprietà o le strutture di Π_s e Π_a separatamente. Può fornire approfondimenti sui modelli specifici, gli autovalori, i valori singolari o altri fattori associati a ciascuna matrice, senza considerare eventuali relazioni o dipendenze tra di loro. La separazione di simmetria di Π: $\Pi = \frac{1}{2}(\Pi + \Pi') + \frac{1}{2}(\Pi - \Pi')$ è un potente dispositivo. Con le decomposizioni degli autovalori di ciascuna parte, è possibile prendere una complicata collezione di associazioni predittive nella matrice Π e decodificarle in un insieme di fatti ben organizzati sui rendimenti attesi. Questi fatti descrivono i) la natura di ogni schema predittivo modello predittivo rappresentato in Π e ii) la forza di questi modelli. La natura di un modello predittivo è descritta dalla sua classificazione come simmetrico o antisimmetrico, che, sorprendentemente, si traducono in beta e alfa. In particolare, il paper dimostra che gli autovettori della matrice simmetrica Π_s sono modi efficaci per ottenere l'esposizione (beta) al fattore basato sul segnale S, mentre gli autovettori della matrice antisimmetrica Π_a sono potenti strategie di neutralizzazione del fattore (alfa) rispetto a questo fattore. Il paper si riferisce alle strategie derivanti dagli autovettori della componente simmetrica come "portafogli di esposizione principale" (PEP) e le strategie derivanti dalla componente antisimmetrica come "portafogli alfa principali" (PAP). Una volta classificati come "esposizione" e "alfa", i modelli di previsione (portafogli principali) vengono ordinati dal più forte al più debole. Questo significa che i portafogli principali (PEPs e PAPs) vengono classificati in base alla loro forza, determinata dall'entità dei loro autovalori. Quelli con autovalori più grandi sono considerati più forti, mentre quelli con autovalori più piccoli sono considerati più deboli. Inoltre, si dimostra che i rendimenti medi incondizionati dei portafogli principali sono strettamente legati ai loro autovalori. Ciò implica che i portafogli principali con autovalori più grandi generano rendimenti medi più elevati, mentre quelli con autovalori più piccoli generano rendimenti medi più bassi. In sostanza, l'ordinamento dei portafogli principali basato sugli autovalori fornisce una misura della loro forza e importanza nella spiegazione dei rendimenti degli asset. I portafogli principali con autovalori maggiori sono considerati più significativi e possono offrire rendimenti medi più alti rispetto a quelli con autovalori minori.

Ciò significa che la scomposizione della matrice Π in componenti simmetriche e antisimmetriche è strettamente collegata all'asset pricing di equilibrio. In un contesto in cui i segnali sono beta rispetto al pricing kernel e non ci sono opportunità di arbitraggio, si osservano le seguenti relazioni:

- Tutti i portafogli di alpha principale (PAPs) devono avere rendimenti attesi in eccesso nulli perché non hanno esposizione al fattore
- Tutti i portafogli di esposizione principale (PEPs) devono fornire rendimenti medi non negativi perché hanno esposizione positiva al pricing kernel

Queste relazioni sono fondamentali per un nuovo test di asset pricing basato sugli autovalori delle componenti simmetriche e antisimmetriche della matrice di previsione. Nel contesto di modelli di asset pricing razionali, non dovrebbe esserci alcuna componente alpha rispetto al pricing kernel. Pertanto, se si scelgono segnali che sono proporzionali alle covarianze con il pricing kernel (come, ad esempio, i beta di mercato), la matrice di previsione corrispondente dovrebbe avere una parte antisimmetrica nulla, ovvero Π dovrebbe essere simmetrica, e non dovrebbero esistere portafogli di alfa. Inoltre, gli autovalori negativi della parte simmetrica di Π corrispondono a strategie con esposizione negativa al fattore ma rendimenti attesi positivi, che rappresentano un'altra forma di alpha. Tuttavia, poiché i modelli di asset pricing razionali escludono sia l'alpha proveniente dai PAPs che quello dalla parte simmetrica di Π, , il test di asset pricing richiede che la matrice Π sia simmetrica e definita positiva. In parole più semplici, quando i segnali catturano l'esposizione al pricing kernel (il fattore che influenza i rendimenti degli asset), tutti i portafogli di esposizione principale (PEPs) dovrebbero generare rendimenti non negativi, mentre tutti i portafogli di alpha principale (PAPs) dovrebbero generare rendimenti nulli. Il paper sviluppa inoltre i fondamenti teorici per l'utilizzo pratico ed empirico della matrice di previsione dal punto di vista della statistica robusta e del machine learning. I principali risultati teorici sono sviluppati nella popolazione, dove Π è noto e, con N asset, ciò richiede la stima di N^2 parametri. Una parametrizzazione così ricca può portare a stime rumorose e sovrapposte che peggiorano la performance fuori campione dei portafogli principali. Nella letteratura e nella pratica finanziaria, i segnali sono spesso analizzati o negoziati nella forma sommatoria, $\widehat{\Pi}_t$,il che limita essenzialmente l'analisi basata sul segnale a testare un singolo parametro pari alla media della propria predicibilità, sommatoria:

$$\mathbb{E}(\mathbf{S}_{i,t}\mathbf{R}_{i,t+1}) \tag{2.2}$$

Sebbene ciò possa beneficiare di una certa robustezza, limitare l'analisi a un problema di un solo parametro è rigido: si perde ogni informazione utile sull'eterogeneità dell'auto prevedibilità. I portafogli principali sono ben adatti a bilanciare l'esigenza di sfruttare informazioni potenzialmente ricche provenienti da tutta la matrice Π controllando al contempo la parametrizzazione per ridurre il rumore e l'overfitting. Le approssimazioni a basso rango di Π (approssimazione della matrice in più fattori rilevanti) e delle sue componenti, basate sulla simmetria, Π_s e Π_a ,offrono un mezzo per bilanciare entrambe le considerazioni in un modo guidato dai dati al fine di ottenere una solida performance di portafoglio fuori dal campione. Il paper implementa la metodologia in modo empirico utilizzando diversi set di dati. Viene condotta un'analisi fuori campione che, in ogni periodo di tempo t, stima la matrice di previsione dai segnali e dai rendimenti passati (ad es.segnali e rendimenti passati (cioè, solo le informazioni disponibili al momento t). La stima della "prediction matrix" non è complessa, infatti, la formula riscontrata nel paper per la stima, è la seguente:

$$\widehat{\Pi}_t = \frac{1}{120} \sum_{\tau=t-120}^{t-1} R_{\tau+1} S'_\tau \tag{2.3}$$

dove, come è possibile osservare anche più avanti nel paper, per molti esempi viene utilizzata una finestra temporale di 120 periodi. Con questa modalità è stimata la matrice di previsione, successivamente le decomposizioni dei valori singolari di Π e delle sue parti simmetriche e antisimmetriche che producono immediatamente PP, PEP e PAP e successivamente viene tracciata la performance fuori dal campione. La diversificazione dei portafogli Fama and French consente di tracciare la performance fuori dal campione. Come nell'analisi empirica che è stata eseguita sulla base di questo paper, anche qui si considera la performance empirica di portafogli principali utilizzando i portafogli standard di Fama-French come asset di base e i momentum di questi portafogli. In seguito, verrà illustrata la rappresentazione empirica nel capitolo 4. In parole più semplici, l'articolo discute dell'uso di segnali per prevedere i rendimenti degli asset nei mercati finanziari. L'obiettivo è classificare questi segnali in due categorie: esposizioni(beta) e alpha. I segnali di esposizione quindi, i beta) aiutano a prevedere i rendimenti basandosi sul pricing kernel, mentre i segnali di alpha no. L'articolo propone un metodo per scomporre le associazioni predittive nella matrice dei segnali in un insieme di fatti ben organizzati sugli expected returns, basati sulla forza e la natura di ciascun modello predittivo. Questo metodo aiuta a bilanciare la necessità di informazioni utili provenienti da tutta la matrice dei segnali, controllando al contempo il rumore e l'eccessiva adattabilità. Il risultato è un insieme di portafogli principali che possono essere utilizzati per ottenere una robusta performance del portafoglio "al di fuori dei dati di allenamento". Quest'ultima puntualizzazione vuol dire che l'insieme di portafogli che si sta considerando mantenga le stesse prestazioni anche su periodo diversi da quelli che si stanno usando per "allenarlo". Infatti "allenamento del portafoglio" vuol dire proprio sviluppo e analisi di quest'ultimo utilizzando dati storici disponibili. I portafogli in questione devono mantenere buone prestazioni anche in situazioni future. Questo sottolinea la capacità dei portafogli di adattarsi alle nuove condizioni di mercato, mantenere una certa solidità e limitare i rischi in periodi che non sono stati considerati precedentemente. Per ottenere una robusta performance del portafoglio è ottimale monitorare continuamente le dinamiche del mercato e modificare i diversi portafogli in base alle nuove informazioni disponibili.

2.1 Matrice di predizione e principal portfolios

Il framework che utilizza il paper per descrivere l'analisi dei portafogli principali (PPA), è il seguente. Per prima cosa introduce il concetto di strategie lineari basate su segnali predittivi, mostra come le strategie lineari siano strettamente connesse alla matrice di previsione, deriva strategie ottimali e introduce la nozione di portafogli principali che implementano strategie ottimali. Iniziamo introducendo l'ambientazione e la notazione che utilizziamo in tutto il testo. L'economia è composta da N titoli negoziati in tempi discreti. Ad ogni istante t, ciascun titolo i genera un rendimento in eccesso rispetto al tasso privo di rischio, $R_{i,t}$. Tutti i rendimenti in eccesso all'istante t vengono raccolti in un vettore, $R_t = (R_{i,t})_{i=1}^N$ e la loro matrice di varianza-covarianza condizionale è $\sum_{R,t} = \mathrm{Var}_t(R_{t+1})$ Ad ogni istante e per ogni titolo, si ha un "segnale" $S_{i,t}$ e tutti i segnali vengono raccolti in un vettore, $S_t = (S_{i,t})_{i=1}^N$. Queste caratteristiche predittive possono essere viste come beta di mercato, rapporti di valutazione, punteggi di momentum o altri segnali osservabili che fungono da proxy per i rendimenti attesi condizionati. Una strategia lineare che si basa su S ha dei pesi di portafoglio: $w_t' = S_t'L$. Ci riferiamo a L $\in \mathbb{R}^{N \times N}$ come matrice delle posizioni perché ogni colonna di L traduce i segnali in una posizione nel portafoglio per ciascun titolo. Ad esempio, la prima colonna $L_1 = (L_{i,1})_{i=1}^N$ di L traduce tutti i segnali in una posizione nel titolo $1, S_t^{\prime L_1}$. Il rendimento di una strategia lineare è naturalmente dato dalla moltiplicazione delle posizioni per i rendimenti, ovvero:

$$R_{t+1}^w = w_t' R_{t+1} = \sum_j (S_t' L_j)(R_{j,t+1}) = S_t' L R_{t+1} \tag{2.4}$$

Dove, ${S_t}' L_j$ corrisponde al valore della posizione in j e invece $R_{j,t+1}$ al return di j. La maggior parte delle strategie di previsione dei rendimenti studiate nella ricerca si focalizza solo sulle previsioni basate sui segnali di ogni titolo, senza considerare la possibilità di utilizzare segnali di altri titoli. Tuttavia, questa analisi ci permette di creare strategie più sofisticate che possono sfruttare sia le previsioni basate sui segnali propri di ogni titolo, sia le previsioni incrociate utilizzando segnali di altri titoli. Un'importante strategia di riferimento è quella che si basa solo sui segnali propri di ogni titolo, senza considerare la prevedibilità incrociata. Questa strategia viene chiamata "fattore semplice" ed è un punto di partenza comune nella letteratura. Quando viene detto che il fattore semplice ha un rendimento medio positivo, significa che i segnali propri dei titoli sono in grado di prevedere positivamente i rendimenti medi dei titoli stessi. Il fattore semplice ha una strategia lineare con una matrice di posizione pari a L= Id e,

$$\tilde{F}_{t+1} = \sum_i (S_t' L_i)(R_{i,t+1}) = S_t' I_d R_{t+1} \tag{2.5}$$

L' approccio affrontato nel paper e poi successivamente dalla mia analisi, consente di esplorare strategie più complesse che tengono conto di segnali incrociati e delle differenze nelle previsioni dei rendimenti tra i vari titoli. Questo permette di ottenere una migliore comprensione dei fattori che influenzano i rendimenti degli investimenti e di creare portafogli ottimali basati su queste informazioni. Per quanto riguarda invece la matrice di predizione, che è la seguente:

$$\Pi = \mathbb{E}(R_{t+1} S_t') \tag{2.6}$$

Se $S_{j,t}$ prevede in media $R_{j,t+1}$ La matrice Π codifica informazioni predittive su come i segnali prevedono tutti i rendimenti, sia basandosi sui segnali propri degli asset che sulla prevedibilità incrociata. Una strategia che sceglie letteralmente una posizione per un asset uguale al suo segnale $S_{i,t}$ guadagna un rendimento di $R_{i,t+1}S_{i,t}$ $\Pi_{i,i}$ rappresenta il valore atteso di questo rendimento. Allo stesso modo, una strategia che assume una posizione nell'asset i basandosi sul segnale di un altro asset j guadagna rendimenti medi pari a $\Pi_{i,i}$. Se $S_(j,t)$ prevede in media $R_{j,t+1}$ u tutti gli asset, questo equivale a dire che la matrice di previsione ha una traccia positiva (tr, la somma dei suoi elementi diagonali):

$$E(\tilde{F}e) = E(\sum_j S_{j,t}R_{j,t+1}) = \mathrm{tr}(\Pi) > 0 \tag{2.7}$$

Questa nozione di prevedibilità propria positiva in media su tutti gli asset è emersa come criterio standard con cui si misurano i segnali predittivi nella letteratura finanziaria empirica e viene valutata tipicamente sulla media campionaria della strategia in (2.5). La prevedibilità propria media non solo astrae dalle informazioni negli elementi fuori diagonale di Π, ma anche dalla eterogeneità degli effetti propri sulla diagonale principale. In breve, le strategie basate sulla prevedibilità propria media sono fortemente vincolate nelle informazioni che considerano riguardo al contenuto predittivo di S. L'intera matrice Π è necessaria (e sufficiente!) per comprendere i rendimenti di strategie lineari più generali. La funzione obiettivo da massimizzare è la seguente: si ha l'obiettivo di massimizzare il rendimento atteso di una strategia lineare soggetta a un vincolo di portafoglio sulla matrice di posizione L:

$$\max_{L:||L||<1} \mathbb{E}(S_t' L R_{t+1}) \tag{2.8}$$

In altre parole, si sta cercando di trovare la combinazione ottimale di posizioni in cui investire, rappresentata dalla matrice L, in modo da massimizzare il rendimento atteso della strategia lineare. Tuttavia, il vincolo di portafoglio impone che la norma della matrice L (misurata tramite la norma Euclidea, indicata come kLk) non superi 1, ovvero la somma dei quadrati dei suoi elementi non deve superare 1. Questo vincolo implica che la strategia deve essere gestita in modo da mantenere un'esposizione complessiva limitata rispetto alle risorse disponibili. L'obiettivo è quindi trovare la combinazione ottimale di posizioni che massimizzi il rendimento atteso della strategia lineare, tenendo conto del vincolo sulle posizioni di portafoglio. La soluzione ottimale in una raccolta di strategie lineari a cui ci riferiamo come portafogli principali (PP) del segnale S. I portafogli principali sono gli elementi costitutivi che si sommano per formare la strategia lineare ottimale nella Proposizione 3:

$$\max_{L:\|L\|<1} \mathbb{E}(S_t' L R_{t+1}) = \sum_{i=1}^{N} \bar{\lambda}_i \tag{2.9}$$

dove $L = M\Pi'$ con $M = (\Pi'\Pi)^{-1/2}$ e $\bar{\lambda}_1 \geq \ldots \geq \bar{\lambda}_N$, che sono i valori singolari di Π e gli autovalori di $(\Pi'\Pi)^{1/2}$. La costruzione dei PP utilizza la SVD di Π. Ciò vuol dire che, sia $\Pi = U\bar{\Lambda}V'$ e $\bar{\Lambda} = \text{diag}((\bar{\lambda}_1), \ldots, (\bar{\lambda}_N))$ è la matrice diagonale dei singolar values e U e V sono le matrici ortogonali con le colonne chiamate u_k e v_k.La matrice L ottimale può essere riscritta in questo modo:

$$(\Pi'\Pi)^{-1/2}\Pi' = V\bar{\Lambda}^{-1}V'V\bar{\Lambda}U' = VU' = \sum_{k=1}^{N} v_k u_k' \tag{2.10}$$

Viene così definito, il k^{th} portafoglio principale come strategia lineare con matrice di posizione $L_k = v_k(u_k)'$ che ha un return di:

$$(PP_{t+1})^k = S_t' u_k v_k' R_{t+1} \tag{2.11}$$

In termini semplici, un portafoglio principale è una strategia di investimento che può essere vista in due modi. Da un lato, è una strategia lineare semplice che assegna una posizione a ciascun asset in base ai segnali provenienti da quegli asset. Dall'altro lato, è una strategia che negozia un portafoglio basato sui segnali provenienti da un altro portafoglio. La costruzione di un portafoglio principale è molto semplice: basta utilizzare un programma informatico per calcolare la "decomposizione ai valori singolari" di una matrice chiamata Π, ottenendo due insiemi di vettori. Questi vettori rappresentano i portafogli principali. La decomposizione dei portafogli principali è simile a una decomposizione che si fa per la varianza, ma invece di separare la varianza, separiamo il rendimento atteso. Ogni portafoglio principale ha un rendimento atteso che è rappresentato da un valore chiamato "valore singolare". Il rendimento atteso di ciascun portafoglio principale è il suo valore singolare:

$$E(PP_{t+1}^k) = \text{tr}(\Pi v_k u_k') = \text{tr}(U\bar{\Lambda}V' v_k u_k') = \text{tr}(U\bar{\Lambda}e_k u_k') = \text{tr}((\bar{\lambda}_k)u_k u_k') = \bar{\lambda}_k \tag{2.12}$$

Quindi, Il rendimento atteso di ciascun portafoglio principale è dato dal suo corrispondente valore singolare (proposizione 4), $E(PP^i_{i,t+1}) = \bar{\lambda}_i$ e la somma dei portafogli principali è la strategia lineare ottimale:

$$\max \mathbb{E}(S_t' L R_{t+1})_{\|L\| \leq 1} = \mathbb{E}\left(\sum_{i=1}^{N} PP^i_{t+1}\right) = \sum_{i=1}^{N} \bar{\lambda}_i \tag{2.13}$$

Si deriva quindi il rendimento delle strategie ottimali alfa e beta e si mostra come queste possano essere scomposte in portafogli principali, proprio come nella soluzione generale di cui alle Proposizioni 3-4. Decomposizione in portafogli principali, proprio come nella soluzione generale di cui alle Proposizioni 3-4.

2.2 Decomposizione della simmetria alfa-beta

La decomposizione della simmetria in alfa e beta fa riferimento alla separazione dei componenti di un portafoglio principale in due parti: una parte chiamata "beta" e una parte chiamata "alfa". Il componente beta di un portafoglio principale rappresenta l'esposizione a fattori comuni o di mercato che influenzano il rendimento di un'ampia gamma di asset. È essenzialmente la parte del rendimento del portafoglio principale che può essere spiegata da fattori sistemici o condivisi. Il componente beta è spesso associato all'andamento generale del mercato o a fattori di rischio ampiamente riconosciuti. D'altra parte, il componente alfa di un portafoglio principale rappresenta il rendimento che non può essere spiegato dai fattori comuni o di mercato. Rappresenta l'eccesso di rendimento del portafoglio principale rispetto a quanto ci si aspetterebbe in base all'esposizione ai fattori sistemici. Il componente alfa può essere interpretato come il rendimento generato dalle capacità di selezione degli asset manager o dalle opportunità di arbitraggio specifiche. La decomposizione della simmetria in alfa e beta ci consente di comprendere meglio la natura dei rendimenti di un portafoglio principale. Identificando i componenti beta e alfa, possiamo distinguere tra il rendimento del portafoglio principale che è attribuibile ai fattori di mercato e quello che è attribuibile a fattori specifici o idiosincratici.

Questa decomposizione ci aiuta a valutare il contributo relativo di questi due componenti al rendimento complessivo del portafoglio principale.

2.3 Risultati empirici paper

La prima applicazione del paper utilizza i portafogli principali sui 25 portafogli Fama-French, perché si tratta di uno degli insiemi di dati più semplici e più studiati della finanza, è un contesto empirico ideale per dimostrare le proprietà del metodo in modo trasparente. In particolare, i portafogli Fama e French includono tre fattori principali: il fattore di mercato (Market Factor), il fattore di dimensione (Size Factor) e il fattore di valore (Value Factor) poiché sono costruiti selezionando i titoli statunitensi in base alla loro dimensione (misurata dalla capitalizzazione di mercato) e al rapporto di valutazione (book-to-market) e utilizzano dati giornalieri dal luglio 1963 alla fine del 2019. Come segnale predittivo variabile nel tempo per ogni portafoglio, viene utilizzato il momentum. Il rendimento mensile ritardato di un portafoglio è un forte predittore positivo dei rendimenti mensili successivi in un'ampia gamma di portafogli azionari in tutto il mondo (Gupta e Kelly,2019), nonché in altre classi di attività (Moskowitz et al., 2012). Per ogni asset di ciascun campione, viene calcolato il suo rendimento cumulativo negli ultimi 20 giorni di negoziazione (circa un mese), quindi si standardizza il segnale in ogni periodo convertendolo in un rank cross-sectional e dividendolo per il numero di asset e poi sottraendo la media (mappando il segnale nel campo di variazione [-0.5,0.5]). Lo utilizziamo per prevedere i successivi rendimenti mensili (cumulativi di 20 giorni) di ogni portafoglio. Stimiamo la matrice di previsione come la controparte campionaria della definizione

$$\Pi = \mathbb{E}(R_{i,t+1}S_t^{'}) \tag{2.14}$$

utilizzando una "finestra di addestramento" mobile. La finestra di addestramento è costituita dagli ultimi 120 periodi di tempo non sovrapposti, periodi di tempo di 20 giorni. La matrice di previsione stimata al periodo t è la seguente,

$$\widehat{\Pi}_t = \frac{1}{120} \sum_{\tau=t-120}^{t-1} R_{\tau+1} S_{\tau}^{'} \tag{2.15}$$

Sulla base di questa matrice di previsione empirica, calcoliamo i suoi vettori singolari per formare i portafogli principali e calcoliamo gli autovettori delle sue parti simmetriche e antisimmetriche. Dando origine ai portafogli principali empirici di esposizione e ai portafogli principali di alfa. Confrontiamo questi ultimi con il semplice fattore F_t definito nella (2.5). Per limitare gli effetti indebiti dell'illiquidità sulle nostre conclusioni, aggiungiamo sempre un cuscinetto di 1 giorno in più tra l'ultimo giorno del campione di formazione e il primo giorno del campione di previsione. Verifichiamo se i portafogli principali empirici si comportano in conformità con le nostre previsioni teoriche. I valori singolari della matrice di previsione corrispondono ai rendimenti attesi dei corrispondenti PP. Viene riscontrato, che i rendimenti realizzati corrispondono all'incirca alla forma dei valori singolari ex ante, con PP con un numero basso di autovalori e rendimenti realizzati elevati. Tuttavia, mentre questa relazione sarebbe perfetta su base campionaria, è possibile osservare naturalmente una certa degradazione dei rendimenti realizzati rispetto agli autovalori quando si guarda al di fuori del campione. Per quanto riguarda gli autovalori delle parti simmetriche e antisimmetriche della matrice di previsione, e di conseguenza, rendimenti realizzati fuori campione dei corrispondenti PEP e PAP, rispettivamente. Anche in questo caso una stretta relazione tra i rendimenti previsti ex ante e quelli realizzati fuori campione. In questo campione, solo i primi due PP e i primi due PEP sembrano avere un rendimento significativo fuori dal campione, e solo il rendimento del primo PAP è significativo. La Figura 2. 1 riassume la performance fuori campione corretta per il rischio dei PP, dei PEP e dei PAP. Per semplicità, viene riportato solo il rendimento della somma dei tre principali portafogli (per ogni versione: PP, PEP e PAP) e della combinazione dei primi 3 PEP e dei primi 3 PAP. In ogni caso, viene confrontata la loro performance con quella del fattore semplice, che è solo la somma del pro-

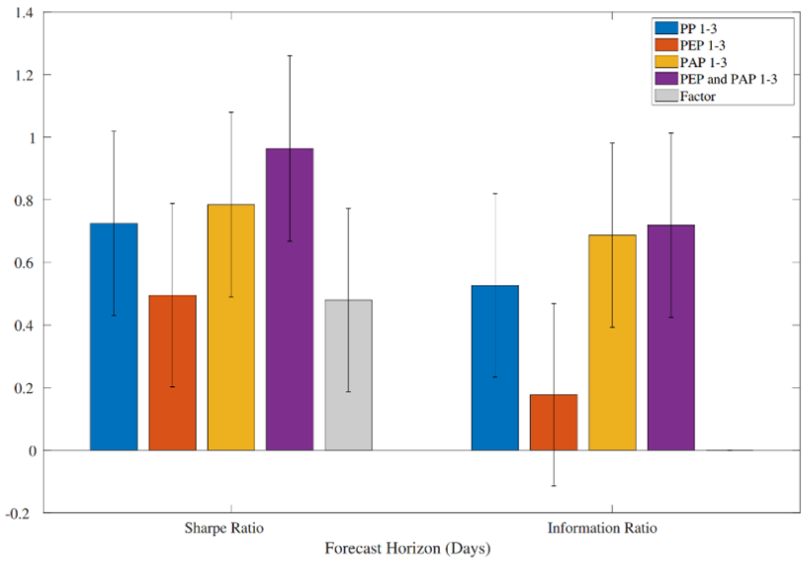

Figura 2.1: Indici di performance dei Principal Portfolios

dotto dei segnali e dei rendimenti. Nell'analizzare la performance del fattore, utilizziamo lo stesso segnale per il fattore e per i PP e si valutano entrambi sugli stessi orizzonti di previsione, per cui ogni gruppo di barre rappresenta un confronto a parità di condizioni fuori dal campione. Il PEP ha uno Sharpe ra-

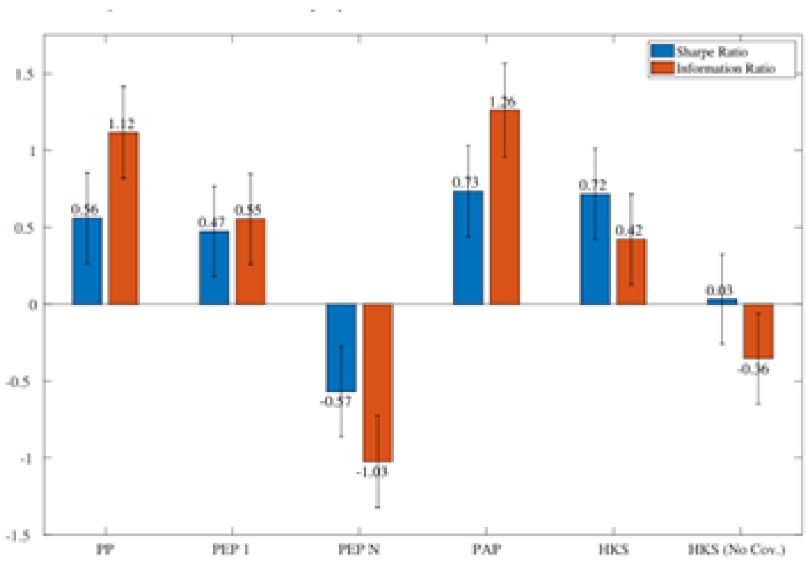

Figura 2.2: Performance delle strategie di temporizzazione dei fattori combinate tra i segnali

tio (SR) simile a quello del fattore semplice, dove SR è il rendimento medio in eccesso diviso per la volatilità. Il PAP ha un SR più elevato e la combinazione di PEP e PAP è ancora più elevato, con un SR più che doppio rispetto a quello del fattore semplice. La strategia PP ha una performance simile a quella del PAP e batte nettamente il fattore semplice. La migliore performance complessiva migliore è quella ottenuta dalla combinazione di PEP e PAP. In tutti i casi, quando riportiamo gli indici di Sharpe e gli indici informativi, riportiamo anche le barre di errore standard di ± 2 intorno a ciascuna stima in base alla formula dell'errore standard approssimativo di Lo (2002). Il grafico 2 riporta anche l'information ratio (IR) fuori campione e il suo intervallo di confidenza come misura del rendimento corretto per il rischio dei portafogli principali. In particolare, l'IR è è calcolato regredendo il rendimento del PP (o del PEP, del PAP o della loro combinazione) sul fattore semplice (Fe) e i cinque fattori di Fama-French (il mercato MKT, il fattore dimensionale SMB, il fattore valore HML, il fattore redditività RMW e il fattore investimento CMA):

$$PP_t = \alpha + \beta^0 \cdot F_t + \beta^1 \cdot MKT_t + \beta^2 \cdot SMB_t + \beta^3 \cdot HML_t + \beta^4 \cdot RMW_t + \beta^5 \cdot CMA_t + \varepsilon_t \tag{2.16}$$

L'IR è l'alfa diviso per la volatilità residua, IR $= \frac{\alpha}{\sigma(\varepsilon_t)}$, che può essere interpretato come il rapporto Sharpe quando tutti i fattori sul lato destro sono coperti (cioè l'alfa espresso come rapporto Sharpe). Il PEP non ha un alfa significativo (o, equivalentemente, un IR significativo), ma il PAP è un alfa significativo (statistica t di 4,42) e lo è anche la strategia PP e la combinazione di PEP e PAP. È interessante notare che la PEP ha un carico altamente significativo sul fattore semplice con un R2 elevato, mentre, al contrario, il PAP ha un carico altamente significativo sul fattore semplice con un R2 elevato. Mentre, al contrario, la strategia PAP ha un carico piccolo e non significativo sul fattore e un R2 basso. Questi risultati sono coerenti con l'idea che il PEP fornisca un'esposizione ai fattori mentre il PAP fornisca un alfa non correlato.

Capitolo 3

Principal component analysis e Singular value decomposition

In questo capitolo vengono spiegati approfonditamente e confrontati due metodi di analisi e scomposizione che sono stati utilizzati nell'analisi empirica svolta nel terzo capitolo. La PCA (Principal component analysis) e l'SVD (singular value decomposition), sono entrambe due tecniche utilizzate per analizzare i dati e per l'apprendimento automatico per la riduzione della dimensionalità di un campionee l'estrazione delle caratteristiche principali. Sono entrambe strettamente correlate, infatti la PCA può essere interpretata come l'applicazione dell'SVD per la covarianza dei dati. Nella PCA, la matrice di covarianza dei dati viene decomposta tramite SVD per ottenere le componenti principali. La PCA, successivamente, utilizza gli autovalori e gli autovettori della matrice di covarianza per calcolare le componenti principali; invece, l'SVD utilizza i valori singolari della matrice per decomporla. Sia la PCA che l'SVD sono utilizzate per ridurre la dimensionalità dei dati, ed estrarre quindi le caratteristiche più significative. Successivamente verrà analizzata l'applicazione pratica dei due metodi, in questo capitolo ci si limita all'analisi teorica.

3.1 PCA

Come già accennato, la Principal Component Analysis (PCA), nota anche come analisi delle componenti principali, è una tecnica statistica utilizzata per semplificare un insieme di dati. È stata sviluppata da Pearson (1901) e Hotelling (1933), mentre il riferimento moderno più valido è Jolliffe (2002). Lo scopo di questo metodo è ridurre la dimensionalità dei dati multivariati preservando il maggior numero possibile di informazioni rilevanti. È una forma di apprendimento non supervisionato in quanto si basa esclusivamente sui dati di input senza fare riferimento ai dati di destinazione corrispondenti (infatti il criterio da massimizzare è la varianza). La PCA, come ampiamente descritto nel libro "An introduction to Statistical Learning" [2] e nel libro "Financial Econometric Modeling" [3] è una trasformazione lineare che trasforma i dati in un nuovo sistema di coordinate tale che il nuovo insieme di variabili, chiamate componenti principali, siano funzioni lineari delle variabili originali, siano non correlate e la maggiore varianza ottenuta dalla proiezione dei dati si trovi sulla prima coordinata, la seconda maggiore varianza sulla seconda coordinata e così via (successivamente verrà fornito un esempio pratico). In pratica, ciò viene ottenuto calcolando la matrice di covarianza dell'intero set di dati. Successivamente, vengono calcolati gli autovettori e gli autovalori della matrice di covarianza e vengono ordinati in base all'autovalore decrescente. È importante notare che il bias della PCA non è sempre appropriato; le caratteristiche con bassa varianza potrebbero avere effettivamente un'elevata rilevanza predittiva, dipende dall'applicazione. Quando ci si trova con un ampio insieme di variabili correlate, le componenti principali permettono di riassumere questo insieme con un numero più piccolo di variabili rappresentative che spiegano collettivamente la maggior parte della variabilità dell'insieme originale. Le direzioni delle componenti principali sono viste come direzioni nello spazio delle caratteristiche lungo le quali i dati originali sono altamente variabili. Queste direzioni definiscono anche linee e sottospazi che si avvicinano il più possibile alla nuvola di dati. Per eseguire regressione a componenti principali, si usano semplicemente le componenti principali come variabili indipendenti o esplicative in un modello di regressione al posto dell'insieme originale di variabili. Quindi, l'analisi delle componenti principali (PCA) si riferisce al processo mediante il quale le componenti principali vengono utilizzate come variabili di input nel modello di regressione e successivamente, l'utilizzo di tali componenti per la comprensione dei dati. Come detto precedentemente, la PCA è un approccio non supervisionato. Non supervisionato infatti vuol dire che l'algoritmo di apprendimento è applicato su dati per i quali non viene fornita nessuna supervisione o guida sulla variabile di output desiderata. Non supervisionato nel senso che non viene fornita nessuna supervisione sulla corretta risposta o sulla variabile target desiderata. Questo per identificare modelli o relazioni significative, gli approcci non supervisionati sono utilizzati per l'esplorazione di dati, la scoperta di cluster o grupppi, riduzione di dimensionalità oppure rilevazione di anomalie. Invece un approccio supervisionato può essere, ad esempio, la previsione del prezzo delle case basata sulle loro caratteristiche. In questo caso, l'obiettivo è addestrare un modello che possa stimare il prezzo di una casa utilizzando un set di dati etichettati che includono informazioni come dimensioni della casa, numero di camere da letto, ubicazione geografica, ecc. Spesso, l'approccio non supervisionato dal momento che coinvolge solo un insieme di caratteristiche X1,X2,...,Xp, e nessuna risposta associata Y. Oltre a produrre variabili derivate da utilizzare in problemi di apprendimento supervisionato, la PCA serve anche come strumento per la visualizzazione dei dati (visualizzazione delle osservazioni o visualizzazione dei dati). Può anche essere utilizzata come strumento per l'imputazione dei dati, cioè per riempire i valori mancanti in una matrice di dati. Cosa sono le componenti principali? Si suppone di voler visualizzare n osservazioni con misure su una serie di p caratteristiche, X1,X2,...,Xp come parte di un'analisi esplorativa dei dati. A tale scopo si possono esaminare i diagrammi di dispersione bidimensionali dei dati, ciascuno dei quali contiene le misurazioni delle n osservazioni su due delle caratteristiche. Tuttavia, ci sono $\binom{p}{2} = \frac{p(p-1)}{2}$ diagrammi di dispersione; per esempio, con p = 10 ci sono 45 diagrammi! Se p è grande, non sarà sicuramente possibile guardarli tutti; inoltre, molto probabilmente nessuno di essi sarà informativo, poiché ognuno di essi contiene solo una piccola frazione delle informazioni totali presenti nell'insieme di dati. È chiaro che è necessario un metodo migliore per visualizzare le n osservazioni quando p è grande. In particolare, si cerca una rappresentazione a bassa dimensionalità dei dati che catturi la maggior parte delle informazioni possibili. Ad esempio, se si può ottenere una rappresentazione bidimensionale dei dati che catturi la maggior parte delle informazioni, allora è possibile tracciare le osservazioni in questo spazio a bassa dimensione. La PCA fornisce uno strumento per fare proprio questo. Trova una rappresentazione a bassa dimensione di un insieme di dati che contiene la maggior parte possibile della variazione. L'idea è che ciascuna delle n osservazioni vive in uno spazio p-dimensionale, ma non tutte le dimensioni sono ugualmente interessanti. La PCA cerca un numero ridotto di dimensioni che siano il più possibile interessanti, dove il concetto di interesse è misurato dalla quantità di variazione (varianza) delle osservazioni lungo ciascuna dimensione. Ciascuna delle dimensioni individuate dalla PCA è una combinazione lineare delle caratteristiche p. Si illustra il modo in cui vengono trovate queste dimensioni, o componenti principali. La prima componente principale di un insieme di features X1,X2,. . . ,Xp è la combinazione lineare normalizzata dei features:

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p \tag{3.1}$$

che ha la varianza maggiore. Per normalizzazione si intende che la sommatoria:

$$\sum_{j=1}^{P} \phi_{J1}^2 = 1 \tag{3.2}$$

Gli elementi $\phi_{11}, \cdots, \phi_{P1}$ sono visti come i loadings della prima componente principale; insieme, i loadings costituiscono il vettore di loadings della componente principale, $\phi_1 = \begin{pmatrix} \phi_{11} \\ \phi_{21} \\ \vdots \\ \phi_{P1} \end{pmatrix}$. Vengono vincolati i caricamenti in modo che la loro somma dei quadrati sia uguale a uno, poiché altrimenti l'impostazione di questi elementi in valore assoluto potrebbe risultare in una varianza arbitrariamente grande. Dato un insieme di dati n $\times$ p , X, come si calcola la prima componente principale? Poiché l'interesse è sempre solo alla varianza, si assume che ogni variabile di X sia stata centrata in modo da avere media zero (cioè, le medie delle colonne di X sono zero). Cerchiamo quindi la combinazione lineare dei valori delle caratteristiche del campione nella forma:

$$z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} + \cdots + \phi_{p1}x_{ip} \tag{3.3}$$

che ha la maggiore varianza campionaria, soggetta al vincolo: $\sum_{j=1}^{P} \phi_{J1}^2 = 1$.In altre parole, il primo vettore di carico della componente principale risolve il problema di ottimizzazione:

$$\max_{\phi_{11},\ldots,\phi_{P1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \phi_{j1}x_{ij} \right)^2 \right\} \tag{3.4}$$

Dalla formula (3.3) si può scrivere l'obiettivo nella formula (3.4) come $\frac{1}{n}\sum_{i=1}^{n} \phi_{j1}^2 = 1$. Poichè $\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0$, la media dei valori $z_{11}, z_{21}, \ldots, z_{n1}$ sarà anche zero. Pertanto, l'obiettivo che si sta massimizzando nella formula (3.4) è semplicemente la varianza campionaria dei n valori di z_{i1}. Ci si riferisce a $z_{11}, z_{21}, \ldots, z_{n1}$ come i punteggi del primo componente principale. Il problema (3.4) può essere risolto tramite una decomposizione degli autovalori, una tecnica standard dell'algebra lineare. C'è un' interpretazione geometrica per il primo componente principale. Il vettore di caricamento ϕ_1 con elementi $\phi_{11}, \phi_{21}, \ldots, \phi_{P1}$ definisce una direzione nello spazio delle caratteristiche lungo la quale i dati variano di più. Se si proiettano i punti dati $x_1, x_2, \ldots, x_n$ su questa direzione, i valori proiettati saranno i punteggi del componente principale $z_{1,1}, \ldots, z_{n,1}$ stessi. Dopo che il primo componente principale Z_1 delle caratteristiche è stato determinato, possiamo trovare il secondo componente principale Z_2. Il secondo componente principale è la combinazione lineare di $X_1, \ldots, X_n$ che ha la massima varianza tra tutte le combinazioni lineari non correlate con Z_1. I punteggi del secondo componente principale $z_{1,2}, \ldots, z_{n,2}$ hanno la forma:

$$z_{i2} = \phi_{12}x_{i1} + \phi_{22}x_{i2} + \ldots + \phi_{p2}x_{ip} \tag{3.5}$$

dove ϕ_2 è il vettore di caricamento del secondo componente principale, con elementi $\phi_{12}, \phi_{22}, \ldots, \phi_{P2}$. Si scopre che vincolare Z_2 ad essere non correlato con Z_1 è equivalente a vincolare la direzione ϕ_2 ad essere ortogonale (perpendicolare) alla direzione ϕ_1. Nell'esempio della Figura 3.1, le osservazioni si trovano in uno spazio bidimensionale (poiché p = 2), quindi una volta trovato ϕ_1, c'è solo una possibilità per ϕ_2, che è mostrata come una linea tratteggiata blu. Ma in un set di dati più ampio con p > 2 variabili, ci sono più componenti principali distinte e vengono definite in modo simile. Per trovare ϕ_2, risolviamo un problema simile a (3.4) con ϕ_2 al posto di ϕ_1 e con il vincolo aggiuntivo che ϕ_2 sia ortogonale a ϕ_1. Una volta calcolate le componenti principali, è possibile rappresentarle una contro l'altra per ottenere visualizzazioni a bassa dimensionalità dei dati. Ad esempio, si può rappresentare il vettore di punteggi Z_1 contro Z_2, Z_1 contro Z_3, Z_2 contro Z_3 e così via. Geometricamente, ciò equivale a proiettare i dati originali sullo spazio sotteso da ϕ_1, ϕ_2 e ϕ_3 e rappresentare i punti proiettati. Sulla base di questo, viene rappresentato un esempio l'uso della PCA sul set di dati "USArrests. Per ciascuno dei 50 stati degli Stati Uniti, il set di dati contiene il numero di arresti ogni 100.000 abitanti per tre crimini: aggressione, omicidio e stupro. Viene registrato anche "UrbanPop" (la percentuale della popolazione di ciascuno stato che vive in aree urbane). I vettori di punteggi dei componenti principali hanno lunghezza n = 50, mentre i vettori di caricamento dei componenti principali hanno lunghezza p = 4. La PCA è stata eseguita dopo aver standardizzato ogni variabile in modo da avere media zero e deviazione standard uno. La Figura 3.1 rappresenta i primi due componenti principali di questi dati. La figura rappresenta sia i punteggi dei componenti principali che i vettori di caricamento in un'unica visualizzazione chiamata biplot. Nella Figura 3.1, si può osservare che il primo vettore di caricamento assegna un peso approssimativamente uguale ad aggressione, omicidio e stupro, ma con un peso molto inferiore su UrbanPop. Pertanto, questa componente corrisponde approssimativamente a una misura delle tariffe complessive dei crimini gravi. Il secondo vettore di caricamento assegna la maggior parte del suo peso a UrbanPop e molto meno peso alle altre tre caratteristiche. Pertanto, questo componente corrisponde approssimativamente al livello di urbanizzazione dello stato. In generale, è possibile osservare che le variabili legate al crimine (omicidio, aggressione e stupro) sono collocate vicine l'una all'altra, mentre la variabile UrbanPop è distante dalLle altre tre. Ciò indica che le variabili correlate al crimine sono correlate tra loro: gli Stati con alti tassi di omicidio tendono ad avere alti tassi di aggressione e stupro, mentre la variabile UrbanPop è meno correlata con le altre tre. Possiamo esaminare le differenze tra gli Stati attraverso i due vettori di punteggio dei componenti principali mostrati nella Figura 3.1. La discussione sui vettori di caricamento suggerisce che gli Stati con punteggi positivi elevati sul primo componente, come California, Nevada e Florida, hanno alti tassi di criminalità, mentre gli Stati con punteggi bassi sul primo componente, come il North Dakota, hanno tassi di criminalità bassi. La California ha anche un punteggio elevato sul secondo componente, indicando un alto livello di urbanizzazione, mentre lo stesso non si può dire per gli Stati come il Mississippi. Gli Stati che si avvicinano a zero su entrambi i componenti, come l'Indiana, hanno livelli medi sia di criminalità che di urbanizzazione.

Questa rappresentazione fornisce una visione compatta dei dati in uno spazio a due dimensioni, evidenziando le relazioni tra le variabili e le caratteristiche dei singoli stati. La PCA può essere un potente strumento per l'analisi dei dati, in quanto consente di ridurre la dimensionalità dei dati complessi e identificare le principali tendenze e modelli nascosti. Tuttavia, è importante considerare che la PCA è un metodo statistico e che le interpretazioni dei risultati devono essere fatte in base al contesto specifico del problema di studio.

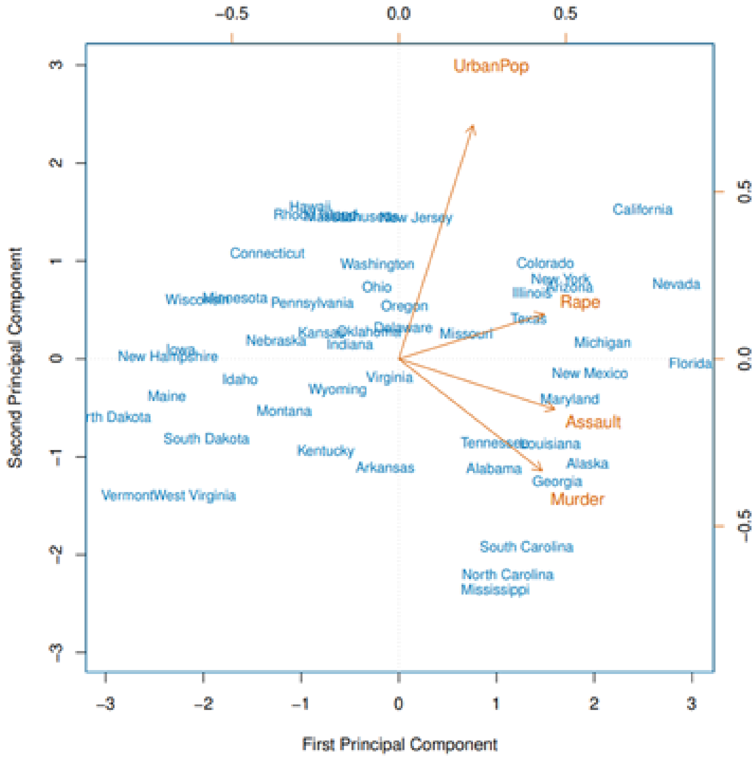

Figura 3.1: Esempio PCA

Le prime due componenti principali per i dati USArrests. I nomi degli stati blu rappresentano i punteggi per le prime due componenti principali. Le frecce arancioni indicano i primi due vettori di caricamento dei componenti principali (con gli assi in alto e a destra). Ad esempio, il caricamento per Stupro sul primo componente è 0,54 e il suo caricamento sul secondo componente principale 0,17 (Lo stupro è centrato nel punto (0,54, 0,17)). Questa figura è nota come biplot, perché visualizza sia i punteggi che i loadings delle componenti principali.

Quindi in sintesi, i passi principali pe il calcolo della PCA sono i seguenti:

- Standardizzazione dei dati: Se necessario, i dati vengono standardizzati per garantire che tutte le variabili abbiano la stessa scala.
- Calcolo delle matrici di covarianza o correlazione: Viene calcolata la matrice di covarianza se i dati non sono stati precedentemente standardizzati, o la matrice di correlazione se i dati sono stati standardizzati. Queste matrici mostrano le relazioni tra le variabili.

- Decomposizione della matrice: Successivamente, viene effettuata la decomposizione della matrice di covarianza o correlazione per ottenere gli autovettori e gli autovalori associati. Gli autovettori rappresentano le direzioni principali dei dati, mentre gli autovalori indicano l'importanza di queste direzioni.

- Selezione delle componenti principali: Le componenti principali vengono selezionate in base all'importanza degli autovalori. Solitamente, le prime componenti principali che spiegano la maggior parte della varianza totale dei dati considerati per l'analisi successiva.

- Proiezione dei dati: Infine, i dati vengono proiettati nel nuovo sistema di coordinate definito dagli autovettori delle componenti principali selezionate. Questa proiezione riduce la dimensionalità dei dati originali e crea nuove variabili non correlate chiamate componenti principali.

La PCA viene utilizzata in vari campi, tra cui l'analisi dei dati, il riconoscimento dei modelli, la riduzione del rumore nei dati e la visualizzazione dei dati ad alta dimensione in uno spazio a dimensione ridotta.

3.2 SVD

Un ruolo fondamentale in algebra lineare (teorica o numerica) è svolto dalle decomposizioni, o fattorizzazioni, delle matrici, che rappresentano uno strumento potente per l'analisi dei problemi e per la progettazione di algoritmi risolutivi. La decomposizione ai valori singolari (SVD) di una matrice è uno strumento molto usato in algebra lineare numerica e generalizza il Teorema Spettrale dalle matrici simmetriche n $\times$ n alle matrici qualsiasi m $\times$ n, questa è una modalità potente per la risoluzione dei problemi. La nascita della SVD viene fatta risalire al 1873, ad opera di E. Beltrami, per matrici quadrate non singolari (con determinante diverso da 0); I l'estensione a matrici complesse (matrici in cui ogni elemento è un numero complesso) è dovuta a L. Autonne nel 1913. Invece per le matrici rettangolari e la generalizzazione delle principali proprietà della decomposizione risalgono al 1936-39, ad opera di C. Eckart e G. Young. Qui verranno prese d'esempio soltanto matrici reali. Una delle applicazioni più note della SVD è proprio quella della Principal Component Analysis o Analisi delle Componenti Principali, nota come PCA equivale ad un'approssimazione a basso rango della matrice ottenuta dalla SVD. Perché quest'ultima si basa sulla selezione delle componenti principali più significative e sulla riduzione del set di dati originale. La SVD ha molteplici applicazioni, viene utilizzata in particolare, per il riconoscimento facciale presente in diversi dispositivi elettronici. I Riconoscimento di volti mediante "eigenfaces" (ref. [4]) che si basano sulla PCA per rappresentare i volti come combinazioni lineari di "eigenfaces". Gli eigenfaces sono i vettori propri della matrice di varianza e covarianza delle immagini dei volti. Questi vettori rappresentano le caratteristiche distintive dei volti presenti nel set di addestramento. Ciascuna eigenface rappresenta un diverso "pattern" di variazione dei pixel facciali. Infatti, nel processo di riconoscimento dei volti con eigenfaces, prima si acquisiscono un insieme di immagini facciali di addestramento. Le immagini poi vengono preelaborate per la rimozione del rumore e allineate per una corretta posizione. Poi si calcola la matrice di covarianza delle immagini facciali e si determinano le eigenfaces con la SVD della matrice. Una volta ottenute le eigenfaces, si possono rappresentare le nuove immagini facciali come combinazioni lineari delle eigenfaces, proiettando le immagini sullo spazio delle eigenfaces. Per identificare o riconoscere un volto sconosciuto, si confronta la sua proiezione nello spazio delle eigenfaces con le proiezioni delle immagini facciali di addestramento. Si può utilizzare una misura di similarità, ad esempio la distanza euclidea, per determinare la corrispondenza più vicina e quindi identificare il volto. Data una matrice $A \in \mathbb{R}^{m \times n}$ e p=min(m,n), una decomposizione ai valori singolari (SVD) di A è una fattorizzazione della forma:

$$A = U\mathcal{E}^T V \tag{3.6}$$

dove U = $(u_1, \ldots, u_m) \in \mathbb{R}^{m \times m}$ e V=$(v_1, \ldots, v_n) \in \mathbb{R}^{n \times n}$ sono ortogonali e $\mathcal{E} \in \mathbb{R}^{m \times n}$ è pseudodiagonale con elementi diagonali $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$.

- se $m \neq n$, sia $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$, la matrice $\mathcal{E}$ ha forma a blocchi $\begin{bmatrix} 0 & \Sigma \end{bmatrix}$
- se $m \geq n$; e, il caso opposto, $m \leq n$,può essere trattato come il primo operando con la matrice trasposta. Teorema per l'esistenza della SVD: sia $A \in \mathbb{R}^{m \times n}$, con $m \geq n$. Allora esistono due matrici ortogonali U = $(u_1, \ldots, u_m) \in \mathbb{R}^{m \times m}$ e V=$(v_1, \ldots, v_n) \in \mathbb{R}^{n \times n}$ tali che $U^T AV = \begin{pmatrix} 0 \\ \Sigma \end{pmatrix}$ con $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$ e $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$ L'esistenza della SVD è dimostrata da Gasparo [5].

Viene definita l'unicità della $\mathcal{E}$ che è unicamente determinata, cioè i valori singolari di una matrice sono unicamente determinati, esiste una sola soluzione possibile per il sistema di equazioni lineari che la coinvolge. Mentre le matrici U e V possono essere scelte in diversi modi e quindi la SVD di una matrice non è unica. A causa dell'unicità dei valori singolari, e a dispetto della non unicità dei vettori singolari si è soliti parlare della SVD di una matrice, e non di una SVD. D'altra parte, la non unicità dei vettori singolari si riferisce al fatto che i vettori singolari associati a valori singolari multipli possono variare o essere multipli. Questo significa che le matrici U e V nella decomposizione SVD possono avere diverse possibili scelte di vettori singolari per valori singolari multipli. Se $A \in \mathbb{R}^{m \times n}$, con $n \geq n$ è una matrice quadrata simmetrica allora i valori singolari coincidono con il valore assoluto (modulo nel caso complesso) degli autovalori. Infatti, essendo A reale e simmetrica, i suoi autovalori λ_i per i=1,..,n , è possibile scrivere: $Av_i = \lambda_i v_i \iff (A^T A)\lambda_i = \lambda_i (A^T A) v_i \iff (A^T A) v_i = \lambda_i^2 v_i$. I valori singolari di A sono proprio $\sqrt{\lambda_i^2} = |\lambda_i|$ per i= 1,...,n. Inoltre A è diagonalizzabile quindi ammette la scomposizione $A = QD^t Q$ con $D = (\lambda_1, \ldots, \lambda_n)$. In altri termini, i valori singolari sono i valori assoluti degli autovalori e i vettori singolari sono autovettori. Se $A \in \mathbb{R}^{m \times n}$, con $m \geq n$, non è simmetrica e i valori singolari e i vettori singolari non hanno legami diretti con i suoi autovalori e autovettori, ma sono invece strettamente collegati a quelli delle due matrici simmetriche $A^t A \in \mathbb{R}^{n \times n}$ e $AA^t \in \mathbb{R}^{m \times m}$. Infatti dalla

(3.6) si deduce che

$$A^tA = V\Sigma^2 V^t e A^t A = U \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} U^t \tag{3.7}$$

e quindi entrambe le matrici ammettono $\sigma_1^2, \ldots, \sigma_n^2$ fra i loro autovalori. Inoltre A^tA ha altri m-n autovalori nulli. Dalla (3.7) si deduce anche che le colonne di V sono autovettori di A^tA e quelle di U lo sono per AA^t; in particolare, i vettori singolari "non essenziali" $u_{n+1}, \ldots, u_m$ sono autovettori corrispondenti all'autovalore nullo. Per quanto riguarda l'interpretazione geometrica dell'SVD, essa è molto semplice e interessante. Considero e_i^n e e_i^m l'i-esimo vettore della base canonica di R^n e R^m rispettivamente, e osserviamo che

$$Av_i = U\mathcal{E}^T V v_i = U\mathcal{E} e_i^n = \sigma_i u_i, \forall i = 1, \ldots, n. \tag{3.8}$$

Si può osservare che le colonne di U e quelle di V costituiscono una base ortonormale di R^m e di R^n rispettivamente. La relazione (1) dice allora che per ogni trasformazione lineare da R^n in R^m è possibile individuare delle basi ortonormali dei due spazi tali che l'i-esimo vettore della base di R^n viene trasformato in un multiplo dell'i-esimo vettore della base di R^m. Rispetto a queste basi, la trasformazione è identificata dalla matrice $\mathcal{E}$.

Quindi in sintesi, le diverse fasi per il calcolo della SVD sono le seguenti:

- Si calcola per prima cosa $C = A^tA$
- Si diagonalizza C, ossia si calcola la fattorizzazione $C = V\Lambda V^T$ con V ortogonale e $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ con $\lambda_1 \geq \cdots \geq \lambda_n$.
- Poi si calcola $\mathcal{E} = \sqrt{\Lambda}$
- Poi si calcolano le colonne di U dalla relazione $AV = U\mathcal{E}$, ovvero $u_i = \frac{Av_i}{\sigma_i}$ per $i = 1, \ldots, r$. Successivamente, si completano le colonne $u_{r+1}, \ldots, u_m$ per ottenere una base ortogonale di $\mathbb{R}^m$.

Questo tipo di operazione è chiamato da Stewart [6] “algoritmo cartesiano”, è infatti, un’operazione tra due vettori tridimensionali che restituisce un terzo vettore ortogonale ai due vettori originali.

3.3 Confronto

La PCA (Principal Component Analysis) e SVD (Singular Value Decomposition) sono due tecniche correlate utilizzate nell’analisi dei dati e nell’estrazione delle componenti principali. PCA è un metodo statistico utilizzato per ridurre la dimensionalità dei dati, identificando le direzioni principali (componenti principali) lungo le quali i dati variano di più. L’obiettivo di PCA è proiettare i dati originali su un nuovo sistema di coordinate costituito dalle componenti principali, in modo da ridurre la complessità del problema senza perdere informazioni significative. SVD è una tecnica matematica più generale utilizzata per decomporre una matrice in tre componenti: una matrice di sinistri singolari, una matrice diagonale di valori singolari e una matrice di destra singolari. Questa decomposizione consente di rappresentare una matrice in modo più compatto e di identificare le direzioni principali lungo le quali i dati variano di più. Il collegamento tra PCA e SVD è che PCA può essere visto come un’applicazione specifica di SVD. Quando si esegue PCA sui dati, si calcola la matrice di covarianza dei dati e quindi si esegue la decomposizione SVD di questa matrice per ottenere le componenti principali. Entrambi PCA e SVD sono utilizzati per l’analisi dei dati e l’estrazione delle componenti principali, ma SVD ha un’applicazione più ampia in altri campi come la compressione delle immagini, il filtraggio dei segnali e la ricostruzione dei dati mancanti. PCA, d’altra parte, è spesso utilizzato come una tecnica di preelaborazione dei dati prima di applicare altri algoritmi di apprendimento automatico o analisi dei dati. La scelta tra PCA e SVD dipende dall’obiettivo specifico dell’analisi dei dati e dalle circostanze specifiche. Non è corretto dire che uno sia "meglio" dell’altro in senso assoluto, poiché entrambi hanno le loro applicazioni e vantaggi. PCA è comunemente utilizzato per la riduzione della dimensionalità e la visualizzazione dei dati, in quanto consente di identificare le direzioni principali lungo le quali i dati variano di più. Questo può semplificare l'interpretazione dei dati e consentire una rappresentazione più compatta dei dati. Inoltre, PCA può essere utilizzato come tecnica di preelaborazione dei dati prima di applicare altri algoritmi di apprendimento automatico. D'altra parte, SVD ha una maggiore flessibilità e un'applicazione più ampia. Oltre all'analisi dei dati, SVD è utilizzato in molti altri campi come la compressione delle immagini, il filtraggio dei segnali, la ricostruzione dei dati mancanti e la risoluzione di problemi di equazioni lineari sovradeterminate. SVD fornisce una scomposizione matematica dei dati che può essere utilizzata in modo più generale rispetto a PCA. In sintesi, se l'obiettivo principale è ridurre la dimensionalità e ottenere una rappresentazione compatta dei dati, PCA può essere una scelta appropriata. Se si desidera una scomposizione matematica più generale dei dati o se si ha l'intenzione di utilizzare SVD per scopi diversi dall'analisi dei dati, allora SVD potrebbe essere più adatto. In ogni caso, entrambe le tecniche sono utili strumenti nell'analisi dei dati e possono essere utilizzate in combinazione a seconda delle esigenze specifiche.

Capitolo 4

Confronto PCA con Principal Portfolios

In quest'ultimo capitolo verranno confrontati in maniera empirica i due metodi esposti precedentemente, in particolare verranno "valutati" i portafogli principali in maniera diversa, in maniera dinamica, come descritto dal capitolo 1, attraverso la metodologia spiegata dal paper di Kelly. In seguito, i portafogli ottimali verranno calcolati applicando la PCA. Successivamente, le due metodologie verranno confrontate applicando la regressione Macbeth, a quel punto verrà calcolato l'R quadro di ciascuna regressione e verrà fatto il test di significatività al fine di verificare quale dei due modelli si adatta meglio ai dati.

4.1 Dati

I dati utilizzati per effettuare le diverse analisi sono stati presi dal sito Fama and French in particolare, sono stati utilizzati i 25 portafogli giornalieri di Fama e French "size e book to market". Il database va dal 1926 al 2022, nella nostra analisi sono stati estratti i portafogli giornalieri del primo giorno disponibile per ogni mese, e quindi il fattore temporale sarà composto da 1153 mesi. La scelta dell'utilizzo di questo dataset è che, storicamente, le strategie di investimento basate sui fattori fama e french hanno dimostrato di generare rendimenti superiori rispetto agli investimenti passivi nell'indice di mercato generale, questi infatti sfruttano le anomalie nel prezzo di mercato. Inoltre, utilizzare tutti i 25 portafogli per la costruzione delle strategie di investimento comporta maggior diversificazione, rispetto all'utilizzo di un solo titolo. E poi, i 25 portafogli dimensione e book-to-market sono basati su caratteristiche fondamentali della società, come la capitalizzazione di mercato e il rapporto prezzo/utili. Gli svantaggi possono essere alti costi di transazione, infatti utilizzare strategie basate su portafogli giornalieri richiede operazioni di compravendita frequenti che richiedono alti costi che potrebbero ridurre i rendimenti generati.

4.2 Procedimento

L'obiettivo dell'analisi è il calcolo dei principal portfolios e della pca sulla base del dataset precedentemente illustrato per capire quali dei due performa meglio. A tal proposito verrà prima illustrato il procedimento per il calcolo dei principal portrfolios, sulla base di quanto dimostrato nel paper di Kelly e successivamente verrà illustrato il procedimento utilizzato per il calcolo della PCA.

4.2.1 Principal portfolios

Come primo passo, per il calcolo dei principal portfolios, occorre calcolare la matrice dei momentum, la matrice degli R al tempo t+1 e successivamente la matrice dei segnali. Per il calcolo della matrice dei momentum è stata applicata la seguente formula: $M_i(t_1) = \sum_{S=-20}^{t} \frac{R_{i,S-1}}{N} \sum_{i=1}^{N} \sum_{S=-20}^{t} R_{i,S}$ arrivando fino a t_n. Il risultato è una matrice 1153x25. In seguito, sono stati ricavati gli R_{t+1}. I dati presenti in entrambe le matrici sono stati standardizzate, in questo punto si è seguito un metodo più semplificato di quello proposto nel paper (rif.[1]). Infatti, per quanto riguarda la matrice R_{t+1}, essendo espressa in forma percentuale è stata semplicemente divisa per 100, per normalizzare invece la matrice M_t si è effettuato questo procedimento: $\frac{M_t - \mu_{M_t}}{3\sigma_{M_t}}$ (dove $\mu_(M_t)$

corrisponde al vettore media dei momentum per ciascuno dei 25 portafogli e σ_{M_t} corrisponde alla deviazione standard). Questo è il procedimento che si avvicinava maggiormente all'obiettivo proposto nel paper, ossia quello di far rientrare i valori della matrice dei momentum, e conseguentemente quello dei segnali, in un range [-0.5,0.5]. A questo punto la matrice dei segnali $S_t = (M_t'\ddot{O}R_{t+1})/1153$. Dopodiché si calcola la singolar value decomposition alla matrice dei segnali, il risultato sono tre diversi dataframe, rispettivamente U, Σ,V. Dove, nella matrice U saranno presenti i left singular values, sulla diagonale della matrice diagonale Σ, i valori singolari e nella matrice V i right singolar value, come descritto nel capitolo 2. A questo punto, si hanno tutti i dati necessari per calcolare i tre principal portfolio nella metodologia precedentemente descritta. Ovvero: $PP_1 = (R_{t+1} * U_1) * (M_t * V_1)$ dove U_1 e V_1, corrispondono rispettivamente alla prima colonna della matrice U e alla prima colonna della matrice V. Questo tipo di operazione è chiamata "Hadamart product", noto anche come prodotto elemento per elemento, è un'operazione matematica che consiste nella moltiplicazione degli elementi corrispondenti di due matrici o vettori.

4.2.2 PCA

Successivamente è stata calcolata la PCA sulla matrice degli R_{t+1}. Come abbiamo visto, la PCA è utile per ridurre la dimensionalità del nostro dataset, consentendo di rappresentare i dati con un numero inferiore di variabili e mantenendo gran parte dell'informazione. Nel modello fattoriale di Fama French, la PCA può essere utilizzata per identificare le combinazioni lineari delle variabili che spiegano la maggior parte della varianza dei rendimenti del mercato azionario. Ossia riscrivere i fattori di Mkt-RF, SMB e HML come combinazione lineare per ottenere un fattore unico, ad esempio PCA1. Queste combinazioni lineari possono poi essere utilizzate come regressori nel modello fattoriale di Fama French. È di particolare interesse della letteratura in tema di asset pricing dimostrare che la PCA Analysis è un ottimo strumento di approssimazione rispetto al generico modello fattoriale di Fama-French per Data

Set con pochi portafogli. Infatti, come vedremo nell'analisi della varianza, la PCA costruita sui 25 portafoglio di Fama-French riesce a spiegare in maniera efficace la maggior parte della variabilità dei dati. Contrariamente, quando si utilizzano dataset più grandi sono necessari strumenti diversi rispetto alla Principal Component Analysis. Prima di procedere al calcolo dei PCA factors è necessario calcolare gli autovettori e gli autovalori della matrice di varianza-covarianza dei rendimenti dei 25 portafoglio di Fama-French oridinati secondo il fattore SMB e HML. Quindi, tramite la libreria numpy di python si trova la matrice varianza-covarianza degli R_{t+1}. I PCA factors vengono ottenuti come combinazione lineare dei 25 portafogli. I coefficienti della combinazione lineare sono gli autovettori della matrice di varianza-covarianza del dataset di partenza.

4.3 Regressione fama Macbeth e risultati

Come ultimo step si ha l'applicazione della regressione fama macbeth (rif. [7],[8]) ai risultati ottenuti, quindi, si avrà una regressione Fama-macbeth con variabile indipendente i 3 PP trovati e l'altra regressione con variabile indipendente le prime 3 PCA. La variabile dipendente invece, rimane uguale per entrambe, ossia i rendimenti dei 25 portafogli di Fama-French formati sulla dimensione e sul Book-to-Market ratio. Per quanto riguarda i PP come prima regressione si avrà la seguente: $R_{t+1} = \alpha + \beta_1 PP_1 + \beta_2 PP_2 + \beta_3 PP_3 + \epsilon_t$ Per quanto riguarda le PCA invece: $R_{t+1} = \alpha + \beta_1 PCA_1 + \beta_2 PCA_2 + \beta_3 PCA_3 + \epsilon_t$

Accanto a questo approccio lineare, verrà presentato l'approccio cross-sectional di Fama-MacBeth. Quest'ultimi presentano uno step ulteriore alla regressione lineare. Infatti, i beta stimati vengono utilizzati come regressori contro la media dei rendimenti dei singoli portafogli. E quindi come seconda regressione si avrà: $\mu_{R_{t+1}} = \gamma + \gamma_1 \hat{\beta}_1 + \gamma_2 \hat{\beta}_2 + \gamma_3 \hat{\beta}_3 + \varepsilon_t$

L'output del secondo step dell'approccio di Fama-MacBeth è la stima del risk premium per ogni risk factor inserito nel modello di asset pricing. Una volta stimati i coefficienti della regressione lineare e salvati nel DataFrame, è possibile calcolare i risk premium del fattore di mercato seguendo l'approccio di Fama-MacBeth. In questo caso la regressione stima i gamma che rappresentano proprio il risk premium. I beta stimati con le PCA sono diversi da quelli stimati con i PP anche se il procedimento è lo stesso.

Per capire quale delle due regressioni performa meglio, è stato calcolato l'R quadro di entrambe e il test di significativà. Producendo risultati migliori per quanto riguarda l'utilizzo della PCA. In particolare, L'R quadro della PCA è maggiore di quello delle PP, inoltre il P-value delle tre PCA è sempre uguale a 0, ciò vuol dire che le tre PCA sono statisticamente significative, i 3 PP hanno un P-value abbastanza elevato, quindi non lo sono.

Questo implica che la probabilità che l'effetto osservato sia dovuto al caso è elevata, e quindi non si può rigettare l'ipotesi nulla che non ci sia alcuna relazione tra la variabile indipendente e quella dipendente. In altre parole, un alto p-value suggerisce che i risultati della regressione non sono statisticamente significativi e che il modello potrebbe non essere utile per fare previsioni o per comprendere la relazione tra le variabili. Il valore soglia per considerare un p-value alto dipende dal contesto specifico dell'analisi e può variare a seconda delle convenzioni e delle prassi accettate in quel campo. In generale, un p-value superiore a 0,05 (o 0,1) viene considerato alto e suggerisce di non rigettare l'ipotesi nulla.

Capitolo 5

Conclusioni

Questo studio ha cercato di rispondere alla domanda: "Hanno risk premia migliore i portafogli valutati con il metodo della PCA o quelli valutati con il metodo Principal Portfolios?" A tal fine, è stata condotta un'analisi tramite Python delle due modalità, calcolati 3 principal portfolios dei 25 portafogli Fama -French size e book to market (con modalità ampiamente descritta), successivamente è stata applicata la PCA al data-set e sono state prese in considerazione le prime tre Principal components. In seguito, il confronto tra i due metodi è avvenuto confrontando l'R quadro delle regressioni Fama Macbeth calcolate utilizzando i tre portafogli prima di un metodo e poi dell'altro come variabili indipendenti. Il risultato ottenuto è che l'R quadro della regressione che ha come variabili indipendenti le tre PCA è maggiore di quello della regressione che ha come variabili indipendenti i tre Principal Portfolios. Inoltre, il test di significatività applicato alla regressione che ha come variabili indipendenti i tre principal portfolios ha un p-value che supera nettamente lo 0.05, ciò vuol dire che i dati non forniscono prove convincenti a sostegno dell'ipotesi che la variabile indipendente abbia un impatto significativo sulla variabile dipendente (rappresentata dai dati 25 portafogli Fama and French). Ciò può essere interpretato come una mancanza di evidenza statistica per affermare che la variabile indipendente abbia un effetto significativo sul risultato in questione nel contesto specifico dell'analisi. Invece il p-value della regressione delle tre PCA è sempre molto vicino allo zero ciò vuol dire che non c'è alcuna possibilità statistica che l'ipotesi nulla sia vera. Pertanto, si può concludere che vi è una relazione statistica significativa tra la variabile indipendente considerata e la variabile dipendente nel modello di regressione.

Bibliografia

[1] Bryan Kelly, Semyon Malamud, and Lasse Heje Pedersen. Principal portfolios. *The Journal of Finance*, 2023.

[2] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R.* Springer, second edition, 2021.

[3] Stan Hurn, Vance L. Martin, Peter C. B. Phillips, and Jun Yu. *Financial Econometric Modeling.* Oxford University Press, 2021.

[4] L. Montefusco. Esercitazione face recognition.

[5] M.G. Gasparo. Metodi numerici per il calcolo di autovalori e autovettori, valori singolari e vettori singolari di matrici reali, 2021.

[6] G.W. Stewart. *Matrix Algorithms Vol. II: Eigensystems.* SIAM, 1999.

[7] Linear model setup for second-pass regression.

[8] Kevin Sheppard. Example: Fama-macbeth.

Capitolo 6

Appendice

Codice Python

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plot


X = pd.read_excel(r"C:\Users\lucio\Downloads\MOMENTUM_first_day.
    xlsx", index_col = "Date")
X

R = pd.read_excel(r"C:\Users\lucio\Downloads\Rt+1.xlsx", index_col
     = "Date")
R

INDICE = []
INDICE.append(R.index)
INDICE

M_s= (X - X.mean()) / (3*X.std())
M_s

R_t= R/100
R_t

```

```
S0 = np.matmul(np.transpose(M_s),R_t)
S = S0/1153
S

# calcola la Singular Value Decomposition della matrice X
U, s, V = np.linalg.svd(S)

# crea i dataframe a partire dalle matrici U, s e V
df_U = pd.DataFrame(U)
df_s = pd.DataFrame(np.diag(s))
df_V = pd.DataFrame(V)

# stampa i dataframe df_U, df_s e df_V
#print("Matrice U:")
#print(df_U)
#
#print("Matrice s:")
#print(df_s)
#
#print("Matrice V:")
#print(df_V)

print(df_U)

print(df_s)

print(df_V)

# Trovo il primo portafoglio dinamico con le prime colonne di df_U
    e df_V
U_1=df_U.T[0]
V_1=df_V.T[0]
A=np.dot(R_t,U_1)
B=np.dot(M_s,V_1)
P1=np.multiply(A,B)
P1

# Trovo il secondo portafoglio dinamico con le seconde colonne di
```

```
    df_U e df_V
U_2=df_U.T[1]
V_2=df_V.T[1]
C=np.dot(R_t,U_2)
D=np.dot(M_s,V_2)
P2=np.multiply(C,D)
P2

# Trovo il terzo portafoglo dinamico con le terze colonne di df_U
    e df_V
U_3=df_U.T[2]
V_3=df_V.T[2]
E=np.dot(R_t,U_3)
F=np.dot(M_s,V_3)
P3=np.multiply(E,F)
P3

#CREO UN DATAFRAME CON I TRE PP
PP=pd.DataFrame({'PP1': P1, 'PP2': P2, 'PP3': P3})
PP.set_index(INDICE, inplace=True)
PP

#Prima di tovare i pca factors  necessario calcolare gli
    autovettori e gli autovalori della matrice di varianza-
    covarianza dei rendimenti dei 25 portafoglio di Fama-French
Var_Cov = R_t.cov()
Var_Cov

w, v = np.linalg.eig(Var_Cov);
V = pd.DataFrame(v)
V
PCA_Factors = np.matmul(R_t,V)
PCA_Factors= PCA_Factors.rename(columns = {col: "PCA " + str(i+1)
    for i,col in enumerate(PCA_Factors.columns)})
PCA_Factors

#ESTRAZIONE PCA FACTORS
PCA_3Factors = PCA_Factors.iloc[:,0:3]
```

```
PCA_3Factors.columns=["PCA1","PCA2","PCA3"]
PCA_3Factors


#INIZIO A FARE LE DUE REGRESSIONI IMPORTANDO PRIMA LE LOBRERIE E
    POI I 25 PORTAFOGLI
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from scipy.stats import t as t_student
import statsmodels.api as sm
from scipy import stats
from sklearn.metrics import r2_score

#REGRESSIONE FAMA MC BETH CON TRE FATTORI PCA come variabili
    indipiendenti
Portfolios = pd.read_excel(r"C:\Users\lucio\Downloads\25
    _Portfolios_months.xlsx", index_col = "Date")
Portfolios

def fama_french_regression(df, n, independent_df, independent_vars
    ):
     results = pd.DataFrame()
     for i in range(n):
          y = df.iloc[:,i]
          X = independent_df[independent_vars]
          X = sm.add_constant(X)
          model = sm.OLS(y, X).fit()
          coef = model.params
          coef.name = df.columns[i]
          results = pd.concat([results, coef], axis=1)
     return results

#REGRESSIONE FAMA MC BETH CON TRE FATTORI PCA come variabili
    indipiendenti
coefficient_df = fama_french_regression(Portfolios,25,PCA_3Factors
      , ["PCA1","PCA2","PCA3"])
X2=coefficient_df.transpose()
X2
```

```
INDICE1 = []
INDICE1.append(X2.index)
INDICE1

Y = pd.read_excel(r"C:\Users\lucio\Downloads\Media_pond_25_port.xlsx")
Y.set_index(INDICE1, inplace=True)
Y

coeffs = np.linalg.inv(X2.transpose() @ X2) @ X2.transpose() @ Y
coeffs.index = X2.columns
coeffs

# Aggiungere una costante
#X2 = sm.add_constant(X2)
# Creare il modello di regressione
model = sm.OLS(Y, X2).fit()

# Stampare il summary del modello
print(model.summary())

#regressione con i tre PP
coefficient_df = fama_french_regression(Portfolios,25, PP, ["PP1",
    "PP2","PP3"])
X3=coefficient_df.transpose()
X3

coeffsPP = np.linalg.inv(X3.transpose() @ X3) @ X3.transpose() @ Y
coeffsPP.index = X3.columns
coeffsPP


# Creare il modello di regressione
model = sm.OLS(Y, X3).fit()

# Stampare il summary del modello
print(model.summary())
```

Singular Value Decomposition.py

Capitolo 7

Riassunto

Nel seguente lavoro di tesi è stata effettuata un'analisi approfondita sui diversi fattori che nell'attuale realtà dell'asset pricing hanno un risk premia maggiore. In particolare, i fattori su cui si baserà l'analisi sono i primi 3 dei 25 portafogli dinamici, chiamati "principal portfolios", introdotti dal paper "principal portfolios" scritto da Bryan Kelly, Semyon Malamud e Lasse Heje Pedersen e i primi 3 portafogli "statici", calcolati con il metodo noto della PCA. Per portafogli dinamici si intende un insieme di attività finanziarie (come azioni, obbligazioni, titoli di Stato, ecc.) che viene regolarmente aggiornato e riequilibrato nel tempo per tenere conto dei cambiamenti nel mercato e delle strategie di investimento dell'investitore. In altre parole, i portafogli dinamici vengono adattati per rispecchiare le aspettative di rendimento e di rischio dell'investitore, nonché le condizioni di mercato prevalenti. L'obiettivo del lavoro di tesi è quello di analizzare entrambe le modalità di asset pricing, la modalità basata su valutazioni dinamica e quella basata su valutazioni statiche, attraverso la principal component analysis. La struttura è la seguente, nel primo capitolo sono stati riassunti i passaggi principali del paper e quelli più utili per l'analisi. In particolar modo, nel primo paragrafo, è stato illustrato il metodo con cui gli autori del paper hanno ottenuto la matrice di predizione e i principal portfolios. Nel paragrafo seguente, è stata accennato il passaggio successivo dell'analisi del paper, ossia la separazione di simmetria della matrice di predizione. Vi è stato inserito solo un accenno perché nell'analisi svolta dalla sottoscritta questo passaggio non è stato effettuato. Successivamente sono stati riportati i risultati empirici del paper con anche gli indici di misurazione della performance. Nel secondo capitolo è stata riportata la modalità alternativa di valutazione dei portafogli dinamici. Quindi la PCA prende in input un insieme di dati multivariati e ne calcola le componenti principali. Le componenti principali sono combinazioni lineari delle variabili originali che catturano la massima varianza dei dati. Queste componenti principali sono ordinate in base alla quantità di varianza che spiegano, con la prima componente principale che spiega la maggiore varianza. Nel secondo paragrafo di questo capitolo vi è l'approfondimento della metodologia SVD, che è il metodo attraverso il quale è stata scomposta la matrice di predizione nell'analisi, come è possibile vedere nel codice presente in appendice 1. Successivamente sono state messe a confronto le due metodologie, sono due tecniche utilizzate nell'analisi dei dati e nella riduzione della dimensionalità, ma differiscono nel modo in cui operano e negli obiettivi che perseguono. In termini di obiettivi, la PCA cerca di spiegare la massima varianza nei dati e individuare le componenti principali più significative. La SVD, d'altra parte, decompone una matrice in componenti singolari, senza focalizzarsi specificamente sulla varianza. In termini di applicabilità a PCA è comunemente utilizzata per analizzare dati multivariati e ridurre la dimensionalità dei dati. La SVD può essere applicata a una vasta gamma di matrici, inclusi dati multivariati, immagini, matrici sparse, ecc. Per quanto riguarda il formalismo matematico: La PCA si basa sulla matrice di covarianza dei dati, mentre la SVD utilizza una decomposizione matriciale basata sugli autovalori e gli autovettori. Per quanto riguarda l'interprerabilità le componenti principali ottenute dalla PCA sono combinazioni lineari delle variabili originali, rendendo possibile interpretare il significato delle componenti. Le componenti singolari ottenute dalla SVD possono essere meno facilmente interpretabili dal punto di vista dei dati originali. In conclusione, la PCA e la SVD sono entrambe tecniche utili per l'analisi dei dati e la riduzione della dimensionalità, ma si differenziano per l'obiettivo specifico e l'approccio matematico utilizzato. La scelta tra PCA e SVD dipenderà dalla natura dei dati e dagli obiettivi dell'analisi.

Successivamente nel terzo capitolo è descritta approfonditamente la metodologia svolta per l'analisi, il procedimento e risultati. Tutto il codice è inserito nell'appendice. In python è stata effettuata l'analisi dei portafogli, applicando ai 25 portafogli Fama and French, size e book to market una valutazione dinamica con la metodologia dei principal portfolios, descritta nel primo capitolo, e successivamente una statica applicando la PCA alla matrice degli R(t+1) , poi il confronto tra essi con la regressione fama Macbeth. Il risultato ottenuto è che l'R quadro della regressione che ha come variabili indipendenti le tre PCA è maggiore di quello della regressione che ha come variabili indipendenti i tre Principal Portfolios. Inoltre, il test di significatività applicato alla regressione che ha come variabili indipendenti i tre principal portfolios ha un p-value che supera nettamente lo 0.05, ciò vuol dire che i dati non forniscono prove convincenti a sostegno dell'ipotesi che la variabile indipendente abbia un impatto significativo sulla variabile dipendente (rappresentata dai dati 25 portafogli Fama and French). Ciò vuol dire che una valutazione statica dei 25 portafogli Fama e French, size e book to market a preferibile a una dinamica con il metodo dei Principal portfolios.