



# UNIVERSITÀ DI SIENA

1240

Dipartimento di  
Dipartimento di Economia Politica e Statistica

Corso di Laurea in  
Scienze Economiche e Bancarie

## Sintesi di un indicatore di climate risk tramite riduzione dimensionale dei Carbon Allowance Certificate

Relatori:

Prof. Federico Crudu

Prof. Piero Mazzarisi

Candidato:

Giacomo Scali

Sessione autunnale

Anno Accademico

2023-2024

# Indice

<b>1</b>	<b>Introduzione</b>	<b>2</b>
<b>2</b>	<b>Metodologia</b>	<b>5</b>
2.1	Raccolta, pulizia e analisi preliminare dei dati . . . . .	5
2.1.1	Pulizia dei dati . . . . .	5
2.1.2	Analisi preliminare . . . . .	6
2.2	Riduzione dimensionale . . . . .	7
2.2.1	Analisi per componenti principali . . . . .	7
2.2.2	Autoencoder . . . . .	9
2.3	Hedging . . . . .	11
<b>3</b>	<b>Costruzione e analisi della performance dell'indice</b>	<b>13</b>
3.1	Costruzione indice con PCA . . . . .	13
3.2	Costruzione indice con autoencoder . . . . .	16
3.3	Performance degli indici . . . . .	19
<b>4</b>	<b>Discussione dei risultati</b>	<b>21</b>

# Capitolo 1

## Introduzione

Negli ultimi anni, la crescente consapevolezza dei rischi legati al cambiamento climatico ha portato il settore finanziario a integrare fattori ambientali nei modelli di valutazione del rischio e nelle strategie di investimento. La finanza climatica è un campo emergente all'intersezione tra finanza tradizionale e scienza del clima che si concentra sull'analisi e sulla gestione dei rischi e delle opportunità finanziarie derivanti dalle dinamiche climatiche.

Nel marzo 2000 la Commissione Europea ha presentato un documento contenente alcune prime idee sulla progettazione dell'EU ETS (EU Emissions Trading System) che è servita come base per numerose discussioni con le parti interessate che hanno ulteriormente contribuito a modellare la legislazione europea sul clima. La direttiva EU ETS (Direttiva 2003/87/CE del Parlamento europeo), che istituisce un sistema per lo scambio di quote di emissioni di gas a effetto serra nella comunità, è stata adottata nel 2003 e il sistema è stato lanciato nel 2005. Il volume complessivo di gas serra che può essere emesso dalle centrali elettriche, dalle industrie e dal settore dell'aviazione è coperto dal sistema di scambio delle emissioni dell'UE e limitato da un tetto massimo sul numero di quote di emissione. All'interno di questo limite, le aziende ricevono o acquistano quote di emissione (certificati), che possono scambiare secondo le necessità [Klepper and Peterson, 2004]. Il numero di quote diminuisce ogni anno, garantendo che le emissioni totali si riducano.

Ogni quota dà al detentore il diritto di emettere una tonnellata di anidride carbonica ( $CO_2$ ), oppure una quantità equivalente di altri gas serra, come l'ossido di diazoto ( $N_2O$ ) e i perfluorocarburi (PFC). Esistono quote di emissione con diverse scadenze (periodo di tempo in cui si può esercitare il diritto all'emissione), questi strumenti derivati sono interessanti poiché, a differenza di altri fattori ambientali, come gli indicatori ESG (*Environmental, Social, Governance*), che sono disponibili solo con frequenza annuale o trimestrale, i prezzi di questi derivati vengono aggiornati giornalmente. Questo permette di esplorare nuove frontiere nella modellizzazione dei rischi climatici ad alta frequenza.

L'obiettivo di questa tesi è sviluppare un indicatore climatico attraverso la riduzione dimensionale dei prezzi dei certificati con diverse scadenze, utilizzando tecniche di *machine learning* come la *principal component analysis* (PCA) e gli *autoencoder*, da utilizzare come strumento di *hedging*, ovvero per proteggere gli investimenti finanziari dalle fluttuazioni legate ai cambiamenti climatici.

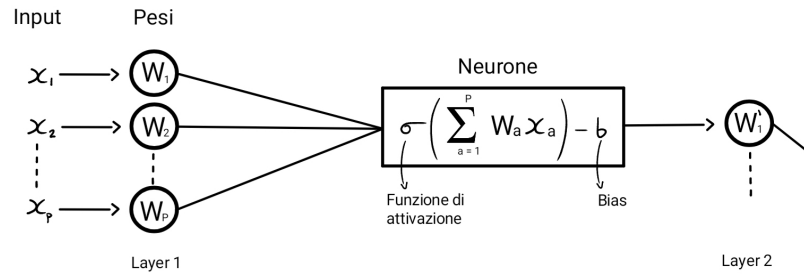
Il *machine learning* è lo studio e l'utilizzo di algoritmi che siano in grado di migliorare in modo autonomo le proprie prestazioni attraverso l'analisi dei dati.<sup>1</sup> Gli algoritmi di *machine learning* possono essere di diversi tipi a seconda del compito per cui sono stati pensati. Se a questi algoritmi viene fornito un numero sufficiente di dati, attraverso processi di ottimizzazione, sono in grado di eseguire il compito assegnato in modo sempre più efficiente senza il bisogno di un intervento esterno.

Nel campo del *machine learning*, una rete neurale artificiale è un modello costituito da un gruppo di interconnessioni neuroni artificiali (o nodi), ogni nodo si connette ad altri e ha il suo peso e la sua soglia (*bias*) associati tramite i quali passa l'informazione ricevuta dal layer precedente a quello successivo (processo schematizzato nella figura 1.1).<sup>2</sup> Come ogni algoritmo di *machine learning*, una rete neurale riceve in input un set di dati, denominato *training set*, da cui impara iterativamente modificando i propri parametri tramite l'ottimizzazione di una certa funzione detta funzione di costo (o *loss function*)  $L(x; \theta)$ , che si calcola a partire dai dati e che dipende dai parametri della rete  $\theta$ . Tale funzione di costo ha un'espressione diversa a seconda degli obiettivi dell'algoritmo e la sua minimizzazione si può raggiungere attraverso diversi metodi. Un problema di questo tipo di algoritmi sta nel fatto che è necessaria una

---

<sup>1</sup>I concetti presentati in questo paragrafo riprendono la trattazione di Goodfellow et al. (2016).

<sup>2</sup>Il bias in questo contesto determina se il neurone debba attivarsi e passare l'informazione al layer successivo.



**Figura 1.1:** Schema del passaggio dell'informazione all'interno del neurone, il significato di  $\sigma$  viene descritto nella sezione 2.2.2.

grande mole di dati per un apprendimento efficace, questa comporta però costi computazionali elevati. Una possibile attenuante a questo problema consiste nella elaborazione *batch* (per lotti), una tecnica in cui i dati vengono suddivisi in lotti più piccoli e immessi nella rete in blocchi separati, anziché tutti in una volta. Ciò consente al calcolatore di utilizzare in modo più efficiente le sue risorse e ridurre significativamente il tempo necessario per l'addestramento del modello.

## Capitolo 2

# Metodologia

### 2.1 Raccolta, pulizia e analisi preliminare dei dati

Per studiare il problema proposto è necessario raccogliere dati relativi ai prezzi a termine giornalieri dei certificati di emissione a diverse scadenze.<sup>1</sup> Questi dati sono stati acquistati da un'azienda specializzata nella raccolta, l'analisi e la vendita di dati, kibot.com. La raccolta copre un arco temporale che va dal 01-01-2008 al 30-12-2019, un totale di 3132 osservazioni per ogni certificato.

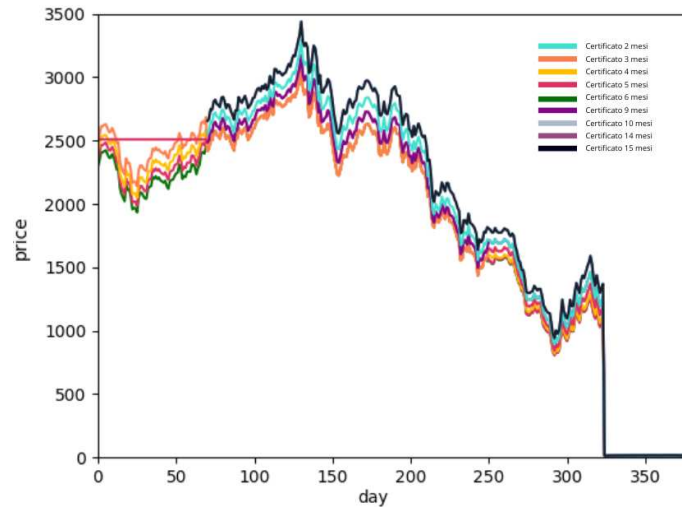
#### 2.1.1 Pulizia dei dati

Per garantire l'integrità e l'affidabilità dei dati, è stata effettuata una verifica preliminare della qualità, correggendo ed eliminando eventuali valori anomali o mancanti. Ad una prima analisi della figura 2.1 si nota subito un comportamento anomalo riguardante i prezzi dei primi 325 giorni per tutte le scadenze. Per alcune scadenze infatti i prezzi dei primi 70 giorni sono mancanti mentre per tutte i prezzi in questo intervallo sono 2 ordini di grandezza superiori al resto del dataset. Osservando però che questa difformità nei prezzi termina esattamente dopo 325 valori per tutte le scadenze e che i valori sono per il resto del dataset registrati fino alla seconda cifra decimale, concludo quindi che deve trattarsi di un errore di inserimento delle cifre decimali.

Per quanto riguarda i valori mancanti, 70 su 3132 circa il 2% delle osservazioni, procedo

---

<sup>1</sup>Il prezzo a termine è il prezzo che l'acquirente si impegna a versare per un bene (o un'attività finanziaria) la cui consegna avviene in un momento diverso rispetto a quello in cui viene stipulato il contratto di compravendita.

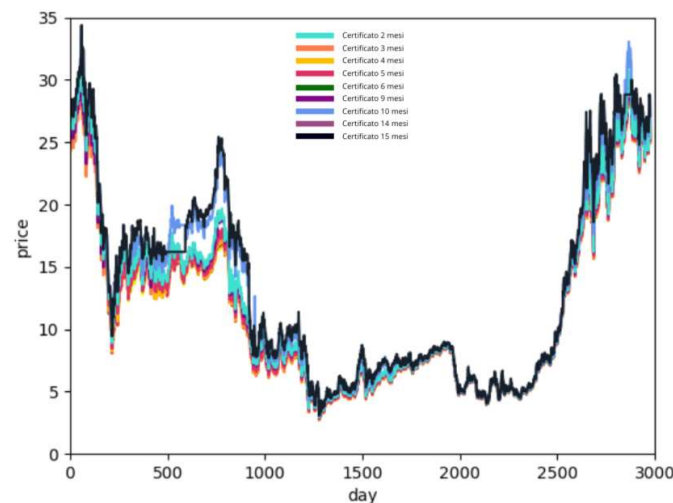


**Figura 2.1:** Prezzi primi 400 giorni di tutte le scadenze, si osserva che i primi 325 valori oscillano tra 1000 e 3500 mentre i restanti fra 4 e 35.

rimuovendo tali date per tutte le scadenze. Il dataset ripulito copre quindi i prezzi di tutte le scadenze dal 08-04-2008 al 30-12-2019 per un totale di 3062 osservazioni.

### 2.1.2 Analisi preliminare

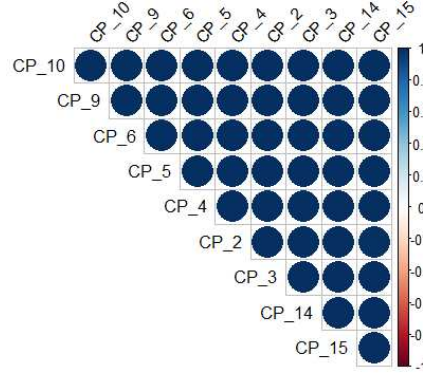
Procedo allora con una analisi preliminare dei dati a disposizione. Sono disponibili i prezzi per diverse scadenze dei certificati: 2, 3, 4, 5, 6, 9, 10, 14, 15 mesi, osservando i grafici dell'andamento dei prezzi nel tempo (figura 2.2), i loro andamenti sembrano molto correlati.



**Figura 2.2:** Prezzi di tutti i certificati per tutte le date disponibili.

Questi derivati infatti sono identici in tutto tranne che per la scadenza, quindi possiamo aspettarci correlazione molto alta fra scadenze diverse. Analizzando la figura 2.3 che riporta

con una codifica cromatica la matrice di correlazione, risulta che le variabili sono fra loro tutte molto correlate che conferma l'ipotesi fatte osservando il grafico dei prezzi. Valori così elevati all'interno della matrice di correlazione supportano la riduzione ad 1 unica dimensione per la creazione di un indice.



**Figura 2.3:** Matrice di correlazione fra i prezzi di tutte le scadenze dei certificati disponibili, codificate come  $CP\_2$  che ad esempio si riferisce al certificato con scadenza a due mesi.

## 2.2 Riduzione dimensionale

### 2.2.1 Analisi per componenti principali

L'analisi per componenti principali è una metodologia statistica di riduzione dimensionale lineare. L'obiettivo dell'analisi per componenti principali (PCA) consiste nella sostituzione delle variabili di partenza con un numero inferiore di nuove variabili non correlate tra loro, tali che conservano la maggior parte dell'informazione presente nei dati originali. Data quindi una matrice di dati di dimensione  $n \times p$  con variabili tutte quantitative si effettua una trasformazione ortogonale tale che le nuove variabili:

1. sono combinazioni lineari delle variabili di partenza,
2. sono tra loro incorrelate,
3. sono in ordine decrescente rispetto alla loro varianza.

Data  $\tilde{X}$  la matrice dei dati centrati sulla media (definita nell'equazione 2.1 dove  $1_n$  è un vettore  $n$  dimensionale di 1 e  $I$  la matrice identità), la prima componente principale  $y_1$  è calcolata come combinazione lineare delle variabili  $\tilde{X}a_1$  (dove  $a_1$  rappresenta il vettore dei coefficienti



della trasformazione lineare), avente massima varianza.

$$\tilde{X} = \left[ \left( I - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^t \right) X \right]. \quad (2.1)$$

Si può dimostrare allora [Mardia et al., 1979] che la massima varianza dei punteggi della prima componente principale si ottiene con  $a_1 = \gamma_1$ , dove  $\gamma_1$  è un autovettore di norma unitaria relativo al primo autovalore  $\lambda_1$  della matrice varianza covarianza  $S$  di  $\tilde{X}$  calcolata come:

$$S = \frac{\tilde{X}^t \tilde{X}}{n}, \quad (2.2)$$

e che in generale per la  $k$ -esima componente vale  $y_k = \tilde{X} \gamma_k$  con  $k \leq \text{rank}(\tilde{X})$  e con  $\gamma_k$  autovettore associato al  $k$ -esimo autovalore, in ordine decrescente di  $S$ .

Valgono inoltre le seguenti proprietà: la varianza della  $k$ -esima componente principale è uguale all'autovalore associato  $\text{var}(y_k) = \lambda_k$ , data la simmetria di  $S$  vale che  $\text{tr}(S) = \sum_{k=1}^p \lambda_k$ , la covarianza tra la  $k$ -esima componente principale e la  $j$ -esima variabile ( $X_{.j}$  indica al colonna  $j$ -esima della matrice dei dati) vale  $\text{Cov}(y_k, X_{.j}) = \lambda_k \gamma_{jk}$ , infine vale che all'interno della classe dei metodi di riduzione dimensionale lineari la PCA è la trasformazione che preserva meglio l'informazione [Fukunaga and Koontz, 1970]. Possiamo allora, date le proprietà sopra elencate, calcolare la porzione di varianza spiegata dalla  $k$ -esima componente principale come  $\frac{\lambda_k}{\sum_{k=1}^p \lambda_k}$  e di conseguenza scegliere il numero delle componenti principali da utilizzare, con l'obiettivo di tener conto di una percentuale sufficientemente elevata della varianza totale. In questo caso voglio verificare che la riduzione ad una sola componente sia in grado di spiegare una quota elevata della varianza totale.

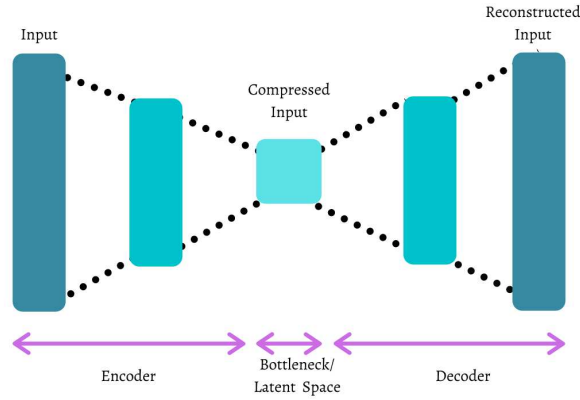
Quando le variabili sono espresse in diverse unità di misura e/o presentano ordini di grandezza molto differenti, l'analisi per componenti principali partendo dalla matrice di covarianza è inappropriata. Si fa riferimento quindi alle variabili espresse in termini di scostamenti standardizzati, che equivale ad assumere come punto di partenza la matrice di correlazione  $R$  invece di quella di varianza covarianza  $S$ .

L'analisi per componenti principali consiste quindi in una trasformazione lineare che mappa i dati in uno spazio di dimensioni ridotte, dove ogni componente principale è una combinazione lineare delle variabili originali.

### 2.2.2 Autoencoder

Un *autoencoder* è una rete neurale artificiale utilizzata per l'apprendimento non supervisionato di codifiche efficienti. Può essere considerato una generalizzazione della PCA, che è in grado di catturare relazioni non lineari nei dati (non linear PCA). In generale la sua struttura è composta da tre parti principali [Kramer, 1991]:

1. *encoder* ( $\Phi$ ): insieme di livelli nell'architettura *autoencoder* che prende in ingresso i dati originali ad alta dimensione e li mappa su di una rappresentazione a bassa dimensione, chiamata codice o *bottleneck*;
2. spazio latente (o codice): livello che contiene la rappresentazione più compressa dell'input con un numero di nodi pari al numero desiderato di variabili che vogliamo estrarre dal modello;
3. *decoder* ( $\Psi$ ): insieme di livelli di un *autoencoder* che decodificano la rappresentazione codificata dei dati, ricostruendo i dati nella loro forma originale precedente alla codifica.



**Figura 2.4:** Struttura generica autoencoder, reperita da: towards data science.

L'obiettivo consiste quindi nel trovare due mappe non lineari  $\Phi$  e  $\Psi$  tali che dato  $X \subset R^p$  e  $F \subset R^m$  ( $X$  mappato su  $R^m$ ) con  $m < p$  minimizzano l'errore di ricostruzione (definito come il quadrato della differenza tra l'output ed i dati in input) [Kramer, 1991]:

$$\begin{aligned}
 \Phi : R^p &\longrightarrow R^m, \\
 \Psi : R^m &\longrightarrow R^p, \\
 \hat{\Phi}, \hat{\Psi} &= \arg \min ||X - (\Psi \circ \Phi)(X)||^2.
 \end{aligned}
 \tag{2.3}$$

Un autoencoder può essere strutturato in vari modi, quello che verrà utilizzato in questo contesto è il più semplice possibile, ovvero quello con un solo layer nascosto (compreso tra input e output layer). L'encoder prende in input  $x \in X$  e lo mappa su  $h \in F$  con

$$h = \sigma(Wx + b). \quad (2.4)$$

Questa immagine  $h$  viene indicata come codice (o rappresentazione latente),  $\sigma$  è una funzione d'attivazione sigmoidea  $\sigma(i) = \frac{1}{1+e^i}$ , necessaria per consentire alla rete di sviluppare rappresentazioni complesse che non sarebbero possibili con un semplice modello lineare.<sup>2</sup>  $W$  e  $b$  sono rispettivamente una matrice dei pesi e un vettore di bias inizializzati in modo casuale e poi aggiornati iterativamente tramite algoritmi di backpropagation; per questa specifica rete ho utilizzato l'algoritmo Adam [Kingma and Ba, 2017]. Il decoder invece si occupa della fase di decodifica, nel quale si passa dalla mappa  $h$  alla ricostruzione  $\tilde{x}$  dei dati

$$\tilde{x} = \sigma(W'x + b'). \quad (2.5)$$

Gli autoencoder sono addestrati per ridurre il più possibile l'errore di ricostruzione, questo avviene minimizzando una funzione di costo. Per questa rete la funzione di costo utilizzata è lo scarto quadratico medio quindi:

$$C(x, \tilde{x}) = \|x - \tilde{x}\|^2 = \|x - \sigma(W'(\sigma(Wx + b) + b))\|^2. \quad (2.6)$$

Nei casi in cui lo spazio latente  $F$  abbia dimensionalità inferiore allo spazio di input  $X$ , il vettore  $\Phi(x)$  può essere considerato come una rappresentazione compressa dell'input  $x$ . Quindi visto che l'obiettivo è quello della sintesi di un indice la dimensione di tale spazio deve essere 1 (quel layer deve essere costituito da un unico nodo).

---

<sup>2</sup>Se l'attivazione fosse lineare l'autoencoder degenererebbe alla PCA [Wang et al., 2016]

## 2.3 Hedging

In finanza, un *hedge* (detto anche copertura) è una strategia di investimento utilizzata per ridurre il rischio legato un'altra attività o posizione [Van Gestel, 2008].

In questo contesto, con rischio ci si riferisce al grado di incertezza e/o potenziale perdita finanziaria derivante da un investimento. Esistono diverse possibili misure di rischio per un asset, fra le più utilizzate tuttavia vi sono la varianza e la deviazione standard dei rendimenti giornalieri, calcolati come  $R_i = \frac{P_i - P_{i-1}}{P_{i-1}}$ .

In questo contesto sono interessato ad esplorare se e quanto una esposizione all'indice sintetizzato in un portafoglio costituito da titoli, soggetti alle normative europee sul clima e monitoraggio delle emissioni, riduca il rischio ad esso associato. Una volta definito l'indice, è necessario calcolare la quota ottimale da allocarci (o *hedge ratio*) per ridurre al massimo il rischio, definita come il rapporto della copertura di una posizione aperta rispetto alla posizione complessiva. A tal proposito definisco il vettore dei rendimenti giornalieri del titolo da coprire  $R_A$ , il vettore dei rendimenti giornalieri dell'indice  $R_I$ ,  $\rho$  come il coefficiente di correlazione tra i due e  $\sigma_A, \sigma_I$  come le rispettive deviazioni standard.<sup>3</sup> Allora la varianza del portafoglio costituito dal titolo  $A$  e la vendita di una quota  $h$  dell'indice è [Saltari et al., 2019]

$$v = \sigma_A^2 + h^2 \sigma_I^2 - 2\rho h \sigma_A \sigma_I. \quad (2.7)$$

Per trovare allora il valore di  $h$  che minimizza la varianza procedo con i calcoli in 2.8.<sup>4</sup>

$$\frac{\partial v}{\partial h} = 2h \sigma_I - 2\rho \sigma_A \sigma_I = 0,$$

controllo le condizioni del secondo ordine

$$\frac{\partial^2 v}{\partial h^2} = 2\sigma_I^2 > 0, \quad (2.8)$$

trovo quindi il minimo risolvendo poi per  $h$

$$h = \rho \frac{\sigma_A}{\sigma_I}.$$

---

<sup>3</sup>È necessario ragionare in termini di rendimenti perché rappresentano una misura adimensionale e quindi indipendente dalla scala dei dati di partenza, cosa non vera in generale per la varianza dei prezzi.

<sup>4</sup>Assumendo che sia possibile effettuare una vendita allo scoperto senza alcun tipo di commissione.

Per quantificare in fine quanto la copertura con una quota  $h$  in portafoglio sia efficace, dati  $Var(U)$  varianza dei rendimenti del portafoglio non coperto e  $Var(H)$  varianza dei rendimenti del portafoglio coperto definisco  $e$  (equazione 2.9) [Ederington, 1979]

$$e = \frac{Var(U) - Var(H)}{Var(U)} = 1 - \frac{Var(H)}{Var(U)}. \quad (2.9)$$

Il valore così ottenuto rappresenta la massima riduzione della varianza del portafoglio ottenibile tramite la vendita o l'acquisto di una quota dell'indice, ad esempio un valore di 0.8 indica una riduzione dell'80% della varianza del portafoglio.

## Capitolo 3

# Costruzione e analisi della performance dell'indice

### 3.1 Costruzione indice con PCA

Per la costruzione dell'indice tramite analisi per componenti principali ho utilizzato una libreria Python chiamata `sklearn` [Pedregosa et al., 2011] che contiene funzioni utili per standardizzare i dati e per calcolare le componenti principali.

```
1  # Standardizzazione delle variabili
2  x = StandardScaler().fit_transform(x)
3  # Calcolo prima componente principale
4  pca = PCA(n_components=1)
5  components = pca.fit_transform(x)
6  principalDf = pd.DataFrame(data = components , columns =
7  ['principal component 1'])
```

**Listing 3.1:** Codice per calcolo prima componente principale.

A riga 2 del listing 3.1 i dati vengono standardizzati con la funzione `StandardScaler()` che, data  $D$  una matrice diagonale con l'inverso delle deviazioni standard dei vettori colonna di  $X$  sulla diagonale come in equazione 3.1,

$$D = \begin{pmatrix} \frac{1}{\sqrt{s_1}} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \frac{1}{\sqrt{s_p}} \end{pmatrix}, \quad (3.1)$$

restituisce  $Z$  la matrice dei dati standardizzati

$$Z = \tilde{X}D, \quad (3.2)$$

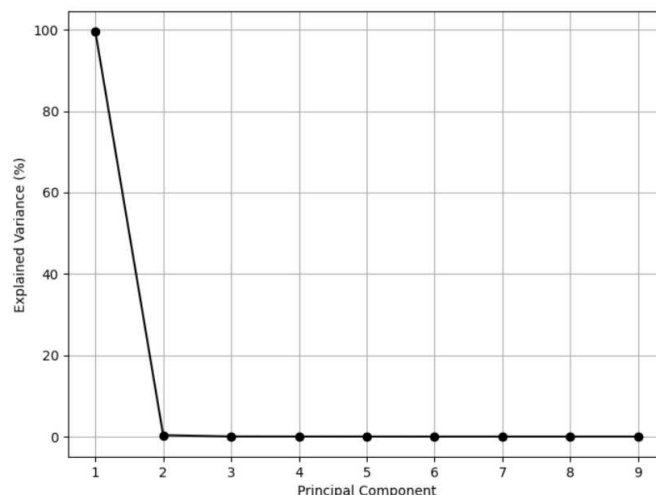
che non è altro che l'estensione matriciale della standardizzazione a media 0 e varianza 1 unidimensionale. A riga 4 del listing 3.1 creo un oggetto di tipo PCA (importato dalla libreria) con solo una componente principale, a riga 5 si calcola infine il vettore dei punteggi della prima componente principale.

Rimane adesso da verificare se una sola componente principale tiene conto di una percentuale sufficientemente elevata della varianza totale (calcolata nel listing 3.2).

```
1  # Calcolo della % di varianza spiegata
2  total_var = pca.explained_variance_ratio_.sum() * 100
```

**Listing 3.2:** Codice per il calcolo della quota di varianza spiegata dalla prima componente principale.

La quota di varianza spiegata è del 99.59%. Anche osservando la figura 3.1, che riporta la percentuale di varianza spiegata da ogni componente principale, la riduzione ad una unica componente preserva quasi tutta l'informazione.



**Figura 3.1:** Grafico della quota di varianza spiegata da ogni componente principale (scree-plot).

Riporto infine nella tabella 3.1 il vettore  $\gamma_1$  (definito nella sezione 2.2.1), ovvero i coefficienti che moltiplicano le variabili di partenza (o *loadings*), questi valori descrivono il peso di una variabile (quindi in questo caso quale scadenza dei certificati) nella definizione della componente principale.<sup>1</sup> Nei fatti si tratta di moltiplicare il vettore  $\gamma_1$  per la radice quadrata del corrispondente autovalore:  $Loadings = \gamma_1 \sqrt{\lambda_1}$ .

$CA_2$	$CA_3$	$CA_4$	$CA_5$	$CA_6$	$CA_9$	$CA_{10}$	$CA_{14}$	$CA_{15}$
0.3194	0.3202	0.3210	0.3255	0.3289	0.3348	0.3325	0.3625	0.3644

**Tabella 3.1:** Loadings prima componente principale,  $CA_2$  ad esempio, si riferisce al certificato con scadenza a due mesi.

Essendo i vettori dei prezzi molto simili fra loro i pesi attribuiti alle diverse scadenze sono molto simili.

---

<sup>1</sup>Possono essere anche utilizzati come pesi per replicare l'indice.



## 3.2 Costruzione indice con autoencoder

Per la costruzione dell'indice con l'*autoencoder* ho utilizzato *keras* [Chollet et al., 2015] una libreria open source per l'apprendimento automatico scritta in Python.

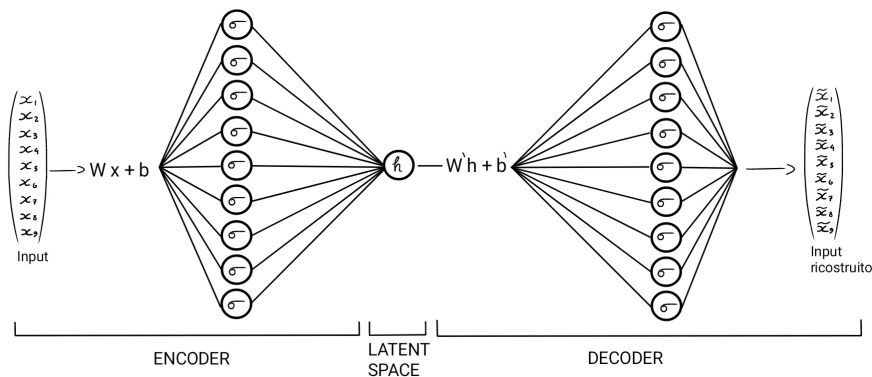
Riporto nel listing 3.3 la parte di codice dove definisco l'architettura della rete (schematizzata nella figura 3.2), ovvero 3 layer così composti: il primo layer di input con 9 nodi con funzione di attivazione sigmoide, il secondo con un unico nodo e un output layer con 9 nodi e funzione di attivazione sigmoide.

```

1 encoding_dim = 1
2 input = keras.Input(shape=(9,))
3 encoded = layers.Dense(encoding_dim, activation='sigmoid')(input)
4 decoded = layers.Dense(9, activation='sigmoid')(encoded)
5 autoencoder = keras.Model(input, decoded)
6 encoder = keras.Model(input, encoded)
7 autoencoder.compile(optimizer='adam', loss='mse')

```

**Listing 3.3:** Codice per la definizione della struttura dell'autoencoder.



**Figura 3.2:** Rappresentazione autoencoder.

Prima di allenare la rete è necessario normalizzare i dati; a questo proposito ho utilizzato il min max scaler:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (3.3)$$

Procedo allora ad allenare il modello dividendo il dataset iniziale in training set e validation set per moderare l'overfitting<sup>2</sup>.

```
1 autoencoder.fit(training_data, training_data,
2                 epochs=60,
3                 batch_size=10,
4                 shuffle=True,
5                 validation_data=(validation_data, validation_data)
6                 )
7
8 encoded_data = encoder.predict(data_std)
```

**Listing 3.4:** Codice per l'addestramento dell'autoencoder.

La parte del codice che regola l'addestramento della rete è riportata nel listing 3.4 dove vengono anche definiti i seguenti iperparametri:<sup>3</sup>

- **epochs=60**: un iperparametro che specifica il numero di volte in cui l'intero training set passerà attraverso il modello durante l'addestramento;
- **batch size=10**: un iperparametro che definisce il numero di campioni da elaborare prima di aggiornare i parametri del modello interno;
- **shuffle=True**: riordina in modo casuale i dati di training prima di ogni epoch per evitare che il modello apprenda modelli specifici dell'ordine dei dati, migliorandone così la generalizzazione [Goodfellow et al., 2016].

Per quanto riguarda la scelta dei valori per questi iperparametri, l'unico modo è determinarli attraverso una serie di tentativi. Questo processo si chiama regolazione degli iperparametri ed in sostanza, si tratta di addestrare il modello in sequenza con diversi set di iperparametri scegliendo poi alla fine quello con cui si sono raggiunti i risultati migliori.

---

<sup>2</sup>Ovvero la tendenza del modello ad adattarsi a caratteristiche che sono specifiche solo del training set, ma che non hanno riscontro nella distribuzione tipica del resto dei casi. Il modello impara le peculiarità del training set e non riesce ad adattarsi a dati nuovi.

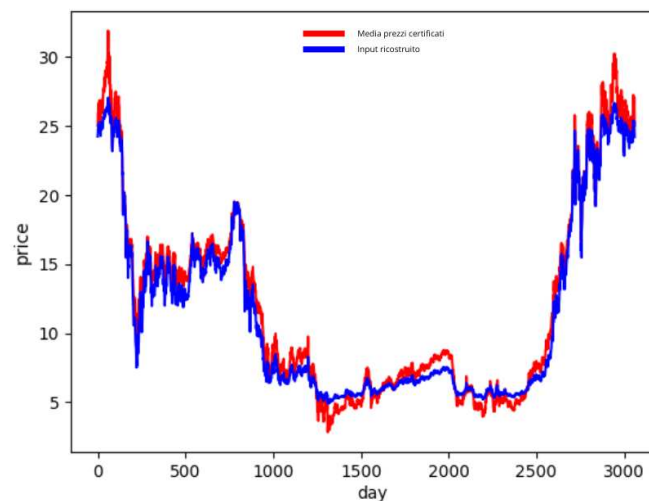
<sup>3</sup>All'interno degli algoritmi di apprendimento automatico ci si riferisce con il termine iperparametro ad un parametro che specifica il processo di apprendimento, questo tipo di parametri a differenza di quelli interni al modello non può essere stimato in maniera automatica.

Dopo aver allenato la rete, posso recuperare la rappresentazione del dataset nello spazio latente (riga 6 del listing 3.4) e anche confrontare input e input ricostruito (calcolato a riga 1 del listing 3.5).

```
1 encoded_data = encoder.predict(data_std)
```

**Listing 3.5:** Codice per il calcolo dei dati codificati.

Indicando con il colore blu i dati ricostruiti e col rosso un portafoglio *equally weighted* composto da tutte i certificati, riporto in figura 3.3 i due grafici.<sup>4</sup>



**Figura 3.3:** Confronto fra input e input ricostruito dall'*autoencoder*.

---

<sup>4</sup>Un portafoglio *equally weighted* attribuisce ad ogni titolo la stessa ponderazione.

### 3.3 Performance degli indici

I titoli che studierò sono SHELL, ENI e RYANAIR. La scelta di tali titoli non è casuale, infatti questi titoli operando principalmente in Europa sono vincolati dalle norme vigenti; inoltre il settore delle centrali elettriche, delle industrie e dell'aviazione sono coperti dal sistema di scambio delle emissioni (EU ETS) e sono soggetti a vincoli più stringenti in termini di emissioni. Risulta interessante valutare anche indici al posto di singoli titoli dato che questi potrebbero essere soggetti a dinamiche soggettive più impattanti rispetto alle politiche riguardanti le emissioni. A tal proposito analizzerò i seguenti indici: l'indice MSCI Europe Energy 35/20, l'indice MSCI Europe Materials 35/20 e l'indice MSCI Europe ESG Broad CTB. Il primo è un indice ponderato per la capitalizzazione di mercato e corretto per il flottante libero<sup>5</sup> che offre un parametro di riferimento per le società europee di dimensioni medio-grandi, classificate nel settore dell'energia in base allo standard globale di classificazione settoriale (GICS). Il secondo è un indice a capitalizzazione di mercato ponderato sulla base del flottante, ideato per misurare il rendimento azionario degli elementi costitutivi rientranti nella categoria Materiali (sulla base degli Standard di Classificazione Industriale Mondiale) nell'indice MSCI Europe. L'ultimo indice invece esclude le società i cui prodotti hanno un impatto sociale o ambientale negativo, sovrappesando al contempo le società con un forte punteggio ESG e con una *carbon footprint*<sup>6</sup> di almeno il 20 % inferiore rispetto l'Indice principale (MSCI Europe).

---

<sup>5</sup>La percentuale delle azioni in circolazione di una società che può essere acquisita dagli investitori retail.

<sup>6</sup>Una misura sia quantitativa sia qualitativa delle emissioni di gas serra generate.

I dati nella tabella 3.2 sono relativi ai rendimenti giornalieri dei titoli e degli indici analizzati/sintetizzati. Si può notare come i rendimenti medi siano vicini ad 0, questo deriva dal fatto che gli stessi sono calcolati su base giornaliera, quindi valori in questo ordine di grandezza non sono anomali. Inoltre la variabilità degli indici sintetizzati appare sensibilmente più elevata rispetto ai titoli analizzati, questo rispecchia l'elevata volatilità osservabile nei prezzi dei certificati nell'intervallo temporale considerato (figura 2.2).

Titolo	Osservazioni	Media	Deviazione standard
PCA	3061	$5.29 \cdot 10^{-4}$	0.0656
Autoencoder	3061	$4.13 \cdot 10^{-4}$	0.0342
ENI	2966	$-1.81 \cdot 10^{-5}$	0.0177
Shell	2962	$2.16 \cdot 10^{-4}$	0.0149
Ryanair	2942	$5.23 \cdot 10^{-4}$	0.0180
MSCI Energy	2969	$2.47 \cdot 10^{-4}$	0.0149
MSCI Materials	2969	$1.64 \cdot 10^{-4}$	0.0165
MSCI ESG	430	$3.53 \cdot 10^{-4}$	0.0071

**Tabella 3.2:** Statistiche preliminari sui rendimenti giornalieri degli indici sintetizzati e dei titoli analizzati.

Nella tabella 3.3 riporto invece i dati ottenuti calcolando hedge ratio ( $h$ ) e quota di varianza ridotta ( $e$ ) come definiti nella sezione 2.3 per tutti i titoli persi in esame, sia per l'indice sintetizzato con la PCA che per l'indice sintetizzato con l'*autoencoder*.

Titolo	Hedge ratio PCA	$e$ PCA	Hedge ratio auto.	$e$ autoencoder
ENI	0.025	1.92%	-0.075	2.27%
Shell	0.002	1.45%	0.003	2.96%
Ryanair	-0.004	0.02%	-0.006	0.01%
MSCI Energy	0.025	2.27%	-0.082	3.51%
MSCI Materials	0.039	3.41%	0.076	4.48%
MSCI ESG	0.015	2.19%	-0.018	2.82%

**Tabella 3.3:** Qualità della copertura ottenuta con gli indici sintetizzati.

Valori negativi per l'edge ratio indicano che l'indice va acquistato e non venduto dato che per i calcoli di  $h$  nella sezione 2.3 ho considerato un portafoglio costituito dal titolo  $A$  e dalla vendita di una quota  $h$  dell'indice sintetizzato.<sup>7</sup> Valori negativi richiedono quindi l'acquisto di una quota  $h$  dell'indice, mentre valori positivi la vendita di tale quota.

---

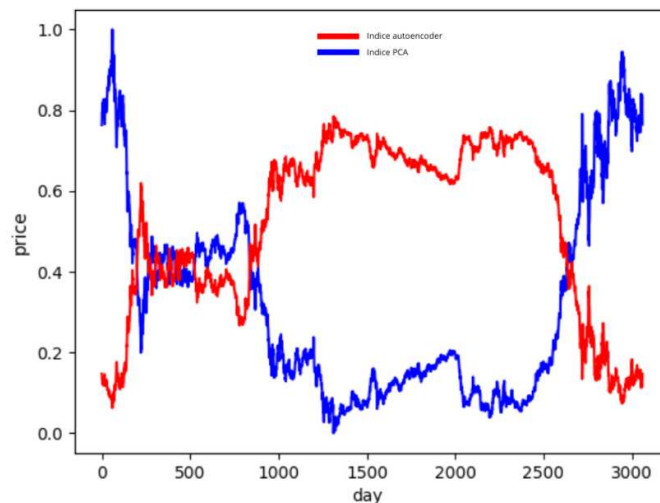
<sup>7</sup>Rispetto al totale investito nel titolo da coprire, ad esempio se il controvalore del titolo da coprire fosse di 100€ e avessi  $h = 0.1$  dovrei costruire un portafoglio con 100€ nel titolo e vendere 10€ dell'indice.

## Capitolo 4

# Discussione dei risultati

Osservando la percentuale di varianza ridotta ( $e$ ) nella tabella 3.3, l'indice sintetizzato con l'autoencoder risulta migliore in termini di varianza ridotta rispetto a quello sintetizzato con la PCA su quasi tutti i titoli considerati. La quota di varianza ridotta raggiunge i livelli più elevati sugli indici MSCI Energy e MSCI Materials, questi indici contengono infatti le aziende che contribuiscono maggiormente alle emissioni di gas serra. Questo risultato supporta allora l'ipotesi fatta riguardante i settori maggiormente influenzati dalle politiche europee sul clima.

Vale la pena osservare gli hedge ratio nella tabella 3.3. Questi differiscono molto fra i due indici, contrariamente a quanto ci si potrebbe aspettare. Questo è da imputare alla differente rappresentazione dei dati adottata dai due algoritmi. Infatti osservando i grafici dei prezzi degli indici nella figura 4.1, si può notare come l'andamento dei due sia simile a meno di



**Figura 4.1:** Grafico dei prezzi degli indici sintetizzati.

una rotazione attorno alla retta  $price = 0.4$ . Questa differenza però non è rilevante dato che la PCA è invariante per rotazione e quindi anche modificando il vettore dei fattori (*loadings* 3.1) tramite una trasformazione ortogonale, per ottenere un valore di  $h$  più vicino a quello dell'*autoencoder*, l'input ricostruito dal modello rimarrebbe invariato.<sup>1</sup>Tale caratteristica dipende dal fatto che l'analisi per componenti principali ammette l'esistenza di più set di parametri che sono *observationally equivalent* [Fisher, 1976].

---

<sup>1</sup>Anche per la PCA è possibile ottenere una ricostruzione dell'input come  $\tilde{x} = y\gamma^t + \mu$ , con  $\mu$  vettore delle medie delle variabili di partenza.

# Bibliografia

- [Chollet et al., 2015] Chollet, F. et al. (2015). Keras. <https://keras.io>.
- [Ederington, 1979] Ederington, L. H. (1979). The hedging performance of the new futures markets. *The journal of finance*, 34(1):157–170.
- [Fisher, 1976] Fisher, F. (1976). *The Identification Problem in Econometrics*. Krieger.
- [Fukunaga and Koontz, 1970] Fukunaga, K. and Koontz, W. L. (1970). Application of the karhunen-loeve expansion to feature selection and ordering. *IEEE Transactions on computers*, 100(4):311–318.
- [Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- [Kingma and Ba, 2017] Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.
- [Klepper and Peterson, 2004] Klepper, G. and Peterson, S. (2004). The eu emissions trading scheme allowance prices, trade flows and competitiveness effects. *European Environment*, 14(4):201–218.
- [Kramer, 1991] Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal*, 37(2):233–243.
- [Mardia et al., 1979] Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. Probability and Mathematical Statistics : a series of monographs and textbooks. Academic Press.
- [Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.



- [Saltari et al., 2019] Saltari, E., Esculapio, S., and Di Pietro, M. (2019). *Introduzione all'economia finanziaria*. Esculapio.
- [Van Gestel, 2008] Van Gestel, T. (2008). *Credit Risk Management: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. OUP Oxford.
- [Wang et al., 2016] Wang, Y., Yao, H., and Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, 184:232–242. RoLoD: Robust Local Descriptors for Computer Vision 2014.