Obrada informacija

4. Laboratorijska vježba: Multivarijatni financijski vremenski nizovi

Siječanj 2023.

Upute

U ovoj bilježnici dana je priprema sa svim uputama za 4. laboratorijsku vježbu iz predmeta Obrada informacija - uz bilježnicu su dostupni i podatci u datoteci `ETFprices.csv`.

Vaš zadatak je u bilježnicu na odgovarajuća mjesta dopisati kod Vašeg rješenja, te odgovore na zadana pitanja.

Riješenu bilježnicu potrebno je predati kao izvještaj u .pdf formatu na *Moodle* najkasnije do 24.1.2023. u 23:59h. Datoteka koju predajete se mora zvati *PrezimeIme.pdf*.

Uvod

U laboratorijskoj vježbi razmatra se dinamika cijena vrijednosnica na financijskim tržištima. Dane su povijesne dnevne cijene 8 ETF-ova (eng. *exchange traded fund*) koji prate određene dioničke, obvezničke ili druge indekse.

Ticker	Fond	Klasa imovine
SPY	SPDR S&P 500 ETF Trust	Equity: U.S. - Large Cap
VTI	Vanguard Total Stock Market ETF	Equity: U.S. - Total Market
QQQ	Invesco QQQ Trust	Equity: U.S. - Large Cap
VEA	Vanguard FTSE Developed Markets ETF	Equity: Developed Markets Ex-U.S. - Total Market
AGG	iShares Core U.S. Aggregate Bond ETF	Fixed Income: U.S. - Broad Market, Broad-based Investment Grade
BND	Vanguard Total Bond Market ETF	Fixed Income: U.S. - Broad Market, Broad-based Investment Grade
LQD	iShares iBoxx USD Investment Grade Corporate Bond ETF	Fixed Income: U.S. - Corporate, Broad-based Investment Grade
VCIT	Vanguard Intermediate-Term Corporate Bond ETF	Fixed Income: U.S. - Corporate, Broad-based Investment Grade Intermediate

Pri modeliranju zajedničkog kretanja i rizika vrijednosnica, najčešće se koriste povrati:

$R(t) = \frac{S(t)-S(t-1)}{S(t-1)}$,

gdje je $S(t)$ cijena vrijednosnice u danu t. U sklopu ove laboratorijske vježbe cilj je analizirati kretanje danih ETF-ova i izračunati glavne komponente (PCA) koje utječu na njihovu dinamiku. Laboratorijsku vježbu je potrebno riješiti unutar ove bilježnice i predati riješenu bilježnicu kao izvještaj.

```
from google.colab import drive
import os

drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import pandas as pd
from matplotlib import pyplot as plt

prices = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/ETFprices.csv")
prices.set_index('Time', inplace=True)
prices.index = pd.to_datetime(prices.index)

prices.head()
```

```
/content
```

Time	SPY	VTI	QQQ	VEA	AGG	BND	LQD	VCIT
2022-01-10	458.078430	231.924240	377.329224	49.100105	109.858536	81.452065	125.620148	88.636536
2022-01-11	462.250763	234.236588	382.997437	49.682655	110.063812	81.617920	125.998177	88.753113
2022-01-12	463.500458	234.600677	384.516266	50.216679	110.063812	81.588654	125.910950	88.811386
2022-01-13	457.114075	231.166565	374.897156	49.857430	110.259331	81.725235	126.308357	88.976517
2022-01-14	457.301086	231.255142	377.229950	49.740917	109.643509	81.276474	125.193680	88.481125

Zadatak 1 - Računanje korelacijske matrice i matrice kovarijance povrata

1.1. U prvom zadatku ove laboratorijske vježbe potrebno je prvo iz danih cijena (gore učitanih u Pandas DataFrame) izračunati dnevne povrate za sve pojedine vrijednosnice (prateći formulu danu u uvodu).

Izračunajte srednje povrate i volatilnost (standardnu devijaciju povrata) za svaku pojedinu vrijednosnicu. Pri analizi srednjih povrata i volatilnosti, te se brojke često *anualiziraju* - to znači da se srednji povrati pomnože s 252 (cca. broj trgovinskih dana u godini), a volatilnost s $\sqrt{252}$. Izračunajte anualizirane srednje povrate i volatilnosti te rezultate ispišite u konzolu.

```
#Vaš kod ide ovdje
import numpy as np
from numpy import linalg as LA
from matplotlib import pyplot as plt
```

```
import math
# print(prices.shift(1))

povrati = (prices -prices.shift(1))/prices.shift(1)
# print(povrati)

print("Srednji povrati:", povrati.mean() * 252, '\n\nVolatilnosti:', povrati.std() * math.sqrt(252))
```

```
Srednji povrati: SPY    -0.138092
VTI    -0.148174
QQQ    -0.283459
VEA    -0.092129
AGG    -0.098139
BND    -0.099603
LQD    -0.135087
VCIT   -0.104975
dtype: float64

Volatilnosti: SPY     0.243655
VTI     0.249912
QQQ     0.322061
VEA     0.224686
AGG     0.081305
BND     0.080695
LQD     0.123953
VCIT    0.095454
dtype: float64
```

1.2 Koristeći dnevne povrate, potrebno je izračunati matricu kovarijance Σ i matricu korelacije C svih ETF-ova. Kovarijancu i korelaciju moguće je iz podataka izračunati koristeći Pandas, ali i NumPy ili neke druge biblioteke. Matrice ispišite u konzolu ili vizualizirajte.

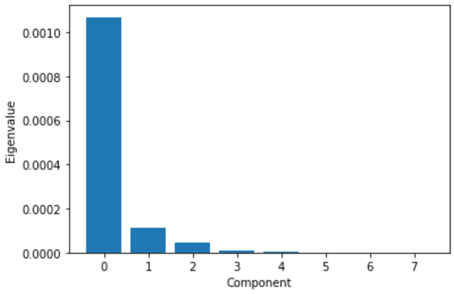
```
#Vaš kod ide ovdje

print("Matrica kovarijacije, Σ:\n")
matrica_kovarijacije = povrati.cov()
print(matrica_kovarijacije)

print("\n\nMatrica korelacije, C:\n")
matrica_korelacije = povrati.corr()
print(matrica_korelacije)

import seaborn as sns
sns.heatmap(matrica_korelacije, annot = True)
```

```
Matrica kovarijacije, Σ:

           SPY       VTI       QQQ       VEA       AGG       BND       LQD  \
SPY   0.000236  0.000241  0.000301  0.000189  0.000026  0.000026  0.000057
VTI   0.000241  0.000248  0.000310  0.000195  0.000027  0.000027  0.000059
```

Zadatak 2 - Analiza glavnih komponenti

2.1. Izračunajte svojstvene vektore i pripadajuće svojstvene vrijednosti matrice kovarijance povrata Σ (svojstvenu dekompoziciju možete pronaći u sklopu biblioteke NumPy https://numpy.org/doc/stable/reference/generated/numpy.linalg.eig.html). Poredajte komponente padajući po svojstvenim vrijednostima i prikažite svojstvene vrijednosti grafički.

```
#Vaš kod ide ovdje

eigenvalues, eigenvectors = np.linalg.eig(matrica_kovarijacije)

# Sort eigenvalues in descending order
idx = eigenvalues.argsort()[::-1]
eigenvalues = eigenvalues[idx]
eigenvectors = eigenvectors[:,idx]

# Plot the eigenvalues
plt.bar(range(len(eigenvalues)), eigenvalues)
plt.xlabel('Component')
plt.ylabel('Eigenvalue')
plt.show()
```



2.2. Izračunajte koliki udio varijance objašnjavaju prve dvije komponente?

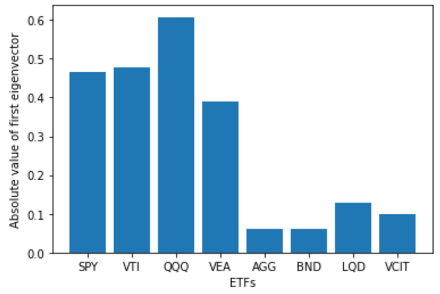
```
#Vaš kod ide ovdje

# Calculate proportion of explained variance by first two components
explained_variance = eigenvalues[:2].sum() / eigenvalues.sum()
print("Proportion of explained variance by first two components: {:.2f}%".format(explained_variance * 100))
```

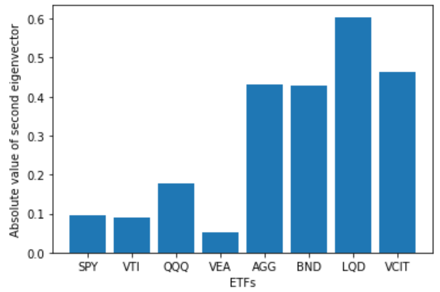
```
Proportion of explained variance by first two components: 95.04%
```

2.3. Komponente PCA će u financijama često opisivati neke zajedničke faktore u podatcima, što je moguće analizirati promatranjem pojedinih elemenata svojstvenih vektora. Ako je neki element određenog svojstvenog vektora velik po magnitudi (pozitivan ili negativan), to znači da ta komponenta opisuje odgovarajuću vrijednosnicu i objašnjava njenu varijancu, za razliku od slučaja kad je element blizu 0, što znači da razmatrana vrijednosnica ne ovisi previše o toj komponenti. Ispišite (ili grafički prikažite) elemente prva dva svojstvena vektora. Pritom pripazite na to što vraća funkcija koju koristite i u kojoj se dimenziji (stupac ili red) nalaze svojstveni vektori.

S obzirom na to koje vrijednosnice opisuju prve dvije komponente, možete li zaključiti koju klasu imovine opisuje prva komponenta (koja odgovara prvom svojstvenom vektoru), a koju klasu druga komponenta? (*odgovor napišite u nastavku*)

```
#Vaš kod ide ovdje

first_eigenvector = np.abs(eigenvectors[:,0])
second_eigenvector = np.abs(eigenvectors[:,1])

# Plot the first eigenvector
plt.bar(povrati.columns, first_eigenvector)
plt.xlabel("ETFs")
plt.ylabel("Absolute value of first eigenvector")
plt.show()

# Plot the second eigenvector
plt.bar(povrati.columns, second_eigenvector)
plt.xlabel("ETFs")
plt.ylabel("Absolute value of second eigenvector")
plt.show()

# Get the absolute values of the first two eigenvectors
first_eigenvector = np.abs(eigenvectors[:,0])
second_eigenvector = np.abs(eigenvectors[:,1])

print(first_eigenvector)
print(second_eigenvector)
```






```
[0.46490446 0.47791577 0.60770765 0.39004316 0.06152524 0.06105256
```

Vjerujem da su najutjecajnije vrijednosnice one s najvecim apsolutnim vrijednostima eigenvectora -> QQQ, VTI za 1. svojstveni vektor -> LQD, VCIT za 2. svojstveni vektor

2.4. Ponovite prethodnu analizu za matricu korelacije povrata C. Koliki udio varijance u tom sučaju objašnjavaju prve dvije komponente? Usporedite elemente prva dva svojstvena vektora u ovom slučaju i u prethodnom slučaju - razlikuje li se interpretacija i kako? (*odgovor napišite u nastavku*)

```
#Vaš kod ide ovdje

# Compute eigenvalues and eigenvectors of the correlation matrix
eigenvalues, eigenvectors = np.linalg.eig(matrica_korelacije)

# Sort the eigenvectors and eigenvalues in descending order
idx = eigenvalues.argsort()[::-1]
eigenvalues = eigenvalues[idx]
eigenvectors = eigenvectors[:,idx]

# Calculate the percentage of variance explained by the first two components
explained_variance = eigenvalues[:2]/sum(eigenvalues)
print("Percentage of variance explained by the first two components:")
print(explained_variance[0] + explained_variance[1])
```

```
Percentage of variance explained by the first two components:
0.9585861905581966
```

Zadatak 3 - Svojstveni portfelji

U primjeni PCA i svojstvenoj dekompoziciji kovarijance u financijama, svojstveni vektori se često zovu i tzv. svojstveni portfelji.

Općenito, portfelj je vektor $w = [w_1, \dots, w_N]$ u kojem svaki element predstavlja težinu ili udio kapitala u određenoj vrijednosnici. Često je dobro pomnožiti njihove težine s predznakom njihove sume - na taj način zapravo samo "okrećemo" predznak svojstvenog vektora tako da mu je suma pozitivna (konačni PCA rastav je i dalje isti ako svojstveni vektor pomnožimo s -1). Također, dobro je i skalirati svojstvene portfelje sa sumom njihovih apsolutnih vrijednosti:

$\tilde{w}_i = \frac{w_i}{\sum_j^N |w_j|}$.

Na taj način se osigurava da visoke magnitude pojedinih elemenata ne uzrokuju velike razlike u volatilnostima svojstvenih portfelja.

Ukoliko znamo povrate $R \in \mathbb{R}^{T \times N}$ (gdje je $R_i \in \mathbb{R}^T$ vektor povrata za vrijednosnicu i) za N vrijednosnica u nekom vremenskom periodu od T dana, povrate portfelja w u tom istom periodu možemo izračunati kao:

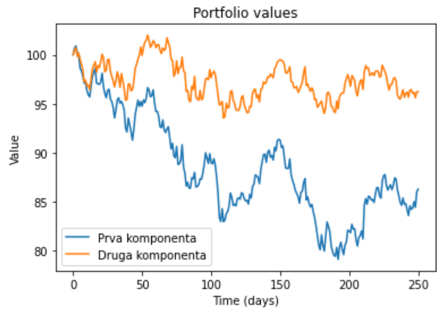
$R_p = \sum R_i w_i = R \cdot w$.

Izračunajte skalirane svojstvene portfelje $\tilde{w}$ koji proizlaze iz prve dvije glavne komponente dobivene iz matrice kovarijance Σ. Za ta dva svojstvena portfelja izračunajte povijesne povrate kroz razmatrani period. Grafički prikažite vremensko kretanje njihovih vrijednosti (njihove povrate "vratite" natrag u cijene, s tim da početna cijena bude jednak za oba portfelja, npr. 100). Također izračunajte anualizirane srednje vrijednosti i volatilnosti svojstvenih portfelja.

```
#Vaš kod ide ovdje

c1 = eigenvectors[:, 0]
c2 = eigenvectors[:, 1]

komponenta1 = np.sign(np.sum(c1)) * c1 / np.sum(np.abs(c1))
komponenta2 = np.sign(np.sum(c2)) * c2 / np.sum(np.abs(c2))

print("Portfolio prve komponente:")
print(komponenta1)
```

```
print("Portfolio druge komponente:")
print(komponenta2)

povrati1 = np.matmul(povrati, komponenta1)
povrati2 = np.matmul(povrati, komponenta2)

vrijednostiPortfolija1 = [0] * len(povrati1)
vrijednostiPortfolija2 = [0] * len(povrati2)

vrijednostiPortfolija1[0] = 100
vrijednostiPortfolija2[0] = 100

for i in range(1, len(povrati1)):
    vrijednostiPortfolija1[i] = vrijednostiPortfolija1[i - 1] * (1 + povrati1[i])
    vrijednostiPortfolija2[i] = vrijednostiPortfolija2[i - 1] * (1 + povrati2[i])

plt.title("Portfolio values")
plt.ylabel("Value")
plt.xlabel("Time (days)")
plt.plot(list(range(len(povrati1))), vrijednostiPortfolija1, label='Prva komponenta')
plt.plot(list(range(len(povrati1))), vrijednostiPortfolija2, label='Druga komponenta')
plt.legend()
plt.show()
```

```
Portfolio prve komponente:
[0.12485224 0.12549028 0.12130324 0.11978049 0.12071949 0.12058677
 0.13323394 0.13403355]
Portfolio druge komponente:
[ 0.13182458  0.13102327  0.13183479  0.11175838 -0.14165837 -0.14159403
 -0.1053407  -0.1049659 ]
```



Ako usporedite dobivene rezultate s kretanjem cijena originalnih vrijednosnica, vidjet ćete sličnosti između vrijednosnica koje pripadaju određenim klasama imovina i pojedinih svojstvenih portfelja. Svojstveni portfelji dakle predstavljaju niže-dimenzionalan prostor tzv. sintetičkih vrijednosnica (u našem slučaju 2 umjesto originalnih 8) koje najbolje opisuju cijeli razmatrani skup podataka. Dobra procjena tih komponenti je ključna u razumijevanju zajedničkog kretanja većih skupova dionica i upravljanju financijskim rizikom.