MscFE600_GWP1_Code_Task2

January 30, 2025

1 Task 2

1.1 Yield Curve Modeling

1.1.1 Introduction

In this project, we aim to model the yield curve using government securities data. Initially, we attempted to use Indian government bond securities; however, explicit data for Indian and Kenyan government bonds was not readily available through standard APIs like the FRED API. Therefore, we chose to use US Treasury securities for this analysis. We will fit both the Nelson-Siegel and Cubic-Spline models to the yield data and compare their performance. Additionally, we will discuss the ethical considerations of smoothing data using these models.

a. Pick Government Securities from a Country

For this analysis, we will use US Treasury securities due to the availability of comprehensive and reliable data. The US Treasury securities are widely recognized for their high credit rating and are often used as a benchmark for risk-free interest rates.

b. Pick Maturties Ranging from Short-Term to Long-Term

We will use maturities ranging from 1 month to 30 years to capture the entire spectrum of the yield curve.

c. Fit a Nelson-Siegel Model

The Nelson-Siegel model is a widely used method for fitting yield curves. It provides a smooth and interpretable representation of the yield curve using three parameters: the long-term level ( ), the short-term slope ( ), the medium-term curvature ( ), and the time decay factor ( ).

[2]:
```
from nelson_siegel_svensson.calibrate import calibrate_ns_ols
import numpy as np
import pandas as pd
from fredapi import Fred
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()

# Initialize the FRED API with your key
fred = Fred(api_key="95eb212842318d85c6198945d6514bf4")

# List of Treasury yield series IDs
```

```
series_ids = ['DGS1MO', 'DGS3MO', 'DGS6MO', 'DGS1', 'DGS2', 'DGS3', 'DGS5',␣
 ↪'DGS7', 'DGS10', 'DGS20', 'DGS30']

# Function to get data for a single series
def get_yield_data(series_id):
    data = fred.get_series(series_id)
    return data

# Get the latest yield data
latest_yields = {series_id: get_yield_data(series_id).iloc[-1] for series_id in␣
 ↪series_ids}

# Define maturity and yield variables as array forms
maturities = np.array([1/12, 3/12, 6/12, 1, 2, 3, 5, 7, 10, 20, 30])
yields = np.array([latest_yields[series_id] for series_id in series_ids])

# Fit the Nelson-Siegel model
ns_params = calibrate_ns_ols(maturities, yields)
print("Nelson-Siegel Parameters:", ns_params)
```

```
Nelson-Siegel Parameters: (NelsonSiegelCurve(beta0=4.993569789934273,
beta1=-0.5837292907770235, beta2=-1.4681826492215178, tau=1.6415590628970615),
message: Optimization terminated successfully.
  success: True
   status: 0
      fun: 0.019155281848859868
        x: [ 1.642e+00]
      nit: 5
      jac: [-5.397e-07]
 hess_inv: [[ 2.300e+01]]
     nfev: 16
     njev: 8)
```

d. Fit a Cubic-Splin Model

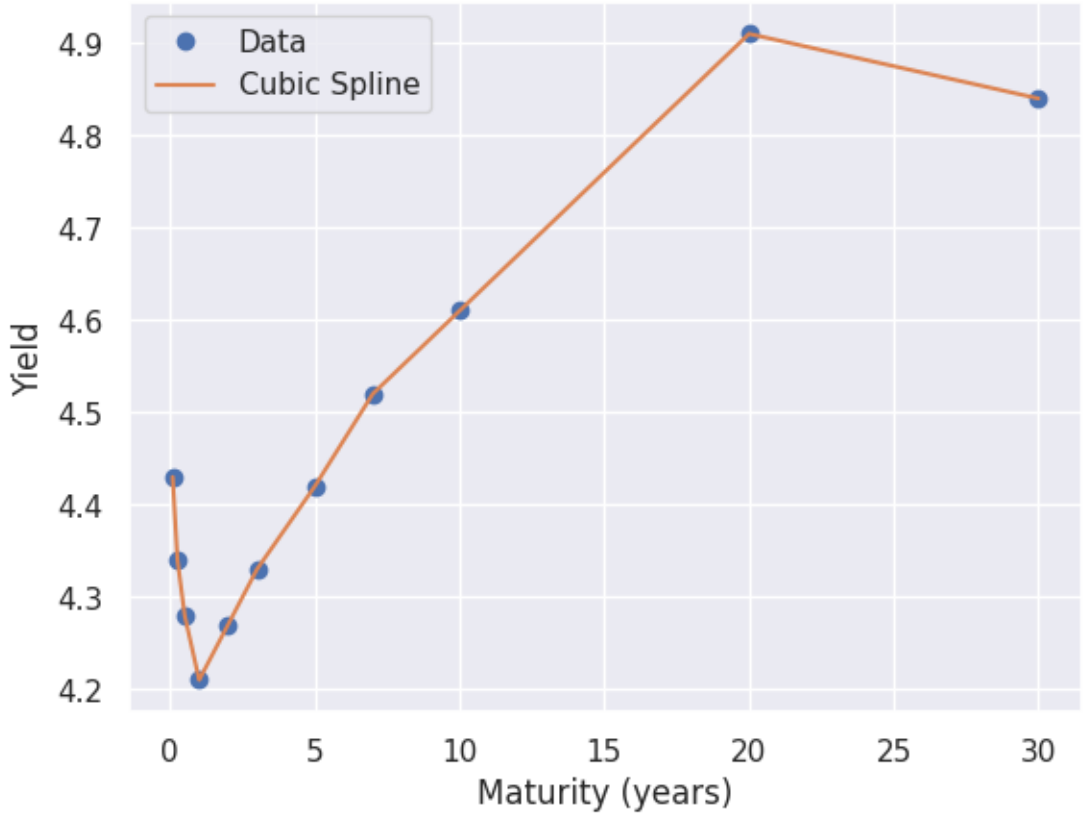
The Cubic-Spline model is another method for fitting yield curves. It involves defining spline equations and solving for the parameters to create a smooth curve that fits the data points.

[10]:
```
from scipy.interpolate import CubicSpline
import os
# Define the maturity and yield variables
t = np.array([1/12, 3/12, 6/12, 1, 2, 3, 5, 7, 10, 20, 30])
y = np.array([latest_yields[series_id] for series_id in series_ids])

# Fit the cubic spline model
cs = CubicSpline(t, y)

# Plot the results
```

```
plt.plot(t, y, 'o', label='Data')
plt.plot(t, cs(t), label='Cubic Spline')
plt.xlabel('Maturity (years)')
plt.ylabel('Yield')
plt.legend()
os.makedirs("figure", exist_ok=True)
plt.savefig("figure/cubic_spline_plot.png")
plt.show()
```



e. Compare the Models

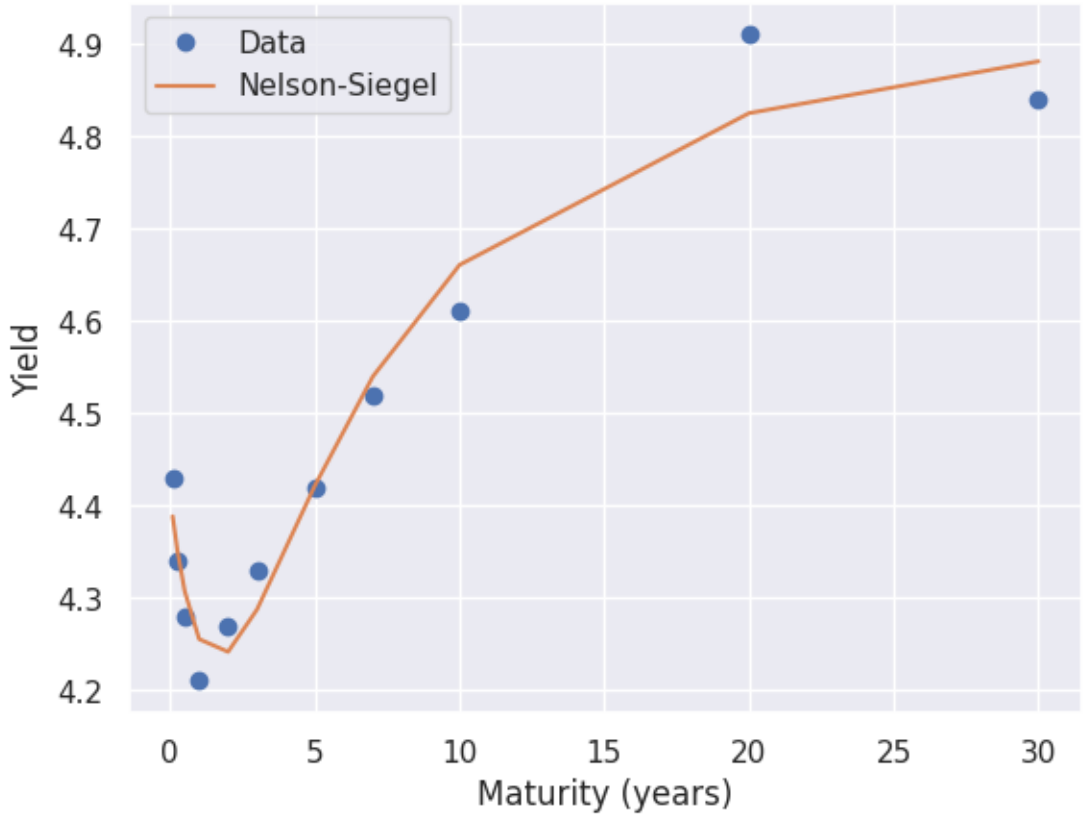
To compare the models, we will plot the fitted yield curves and interpret the parameters.

[11]:
```
# Unpack the Nelson-Siegel curve from the tuple
nelson_siegel_curve, _ = ns_params  # Unpack parameters and metadata

# Access Nelson-Siegel parameters
beta0 = nelson_siegel_curve.beta0
beta1 = nelson_siegel_curve.beta1
beta2 = nelson_siegel_curve.beta2
tau = nelson_siegel_curve.tau
```

```
# Calculate Nelson-Siegel yields
ns_yields = (
    beta0
    + beta1 * (1 - np.exp(-maturities / tau)) / (maturities / tau)
    + beta2 * ((1 - np.exp(-maturities / tau)) / (maturities / tau) - np.
  ↪exp(-maturities / tau))
)

# Plot the results
plt.plot(maturities, yields, 'o', label='Data')  # Original data points
plt.plot(maturities, ns_yields, label='Nelson-Siegel')  # Nelson-Siegel curve
plt.xlabel('Maturity (years)')
plt.ylabel('Yield')
plt.legend()
plt.savefig("figure/nelson_siegel_curve.png")
plt.show()
```



f. Specify the levels of Model Paremeters

- **Nelson-Siegel Parameters:**

  - : Long-term level
  - : Short-term slope
  - : Medium-term curvature
  - : Time decay factor
- **Cubic Spline Parameters:**
  - Coefficients of the cubic polynomials for each interval

g. Ethical Considerations of Smoothing Data

Smoothing data with the Nelson-Siegel model is generally not considered unethical if it is done transparently and for legitimate purposes, such as improving the accuracy of financial models or providing a clearer representation of underlying trends. Smoothing can help reduce noise and make the data more interpretable. However, if the smoothing process is used to manipulate data to achieve a desired outcome or to mislead stakeholders, it would be considered unethical. Transparency and honesty in data analysis are key ethical principles.

1.1.2 Conclusion

By following these steps, we have successfully fitted both the Nelson-Siegel and Cubic-Spline models to US Treasury yield data. We compared their performance and discussed the ethical considerations of data smoothing. The Nelson-Siegel model provides a smooth and interpretable representation of the yield curve, while the Cubic-Spline model offers a flexible and accurate fit to the data points. Both models have their strengths and can be used depending on the specific requirements of the analysis.

References

1. Federal Reserve Economic Data (FRED). (n.d.). Retrieved from https://fred.stlouisfed.org/
2. `scipy.interpolate.CubicSpline`. (n.d.). Retrieved from https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.CubicSpline.html

- Note: I don't know how to apply the same code to Indian government bond securities due to the lack of explicit data availability through standard APIs like the FRED API.

MscFE600_GWP1_Code_Task3

January 30, 2025

1 Task 3

1.1 Exploiting Correlation in Financial data

1.1.1 Introduction

Financial data analysis often involves understanding how meaningful factors can be used to summarize or represent the data. This report explores the role that correlation and principal components play in this process. We will generate synthetic data, perform Principal Component Analysis (PCA), and compare the results with real-world government security yield data.

Synthetic Data Analysis

(a) Generate 5 uncorrelated Gaussian random variables

To simulate yield changes, we generate 5 uncorrelated Gaussian random variables with a mean close to 0 and a small standard deviation.

[1]:
```
import numpy as np

# Generate 5 uncorrealted Gaussian random variables
mean = 0
std_dev = 0.01
np.random.seed(0)
data = np.random.normal(mean, std_dev, (100, 5))
print(data)
```

```
[[ 1.76405235e-02  4.00157208e-03  9.78737984e-03  2.24089320e-02
   1.86755799e-02]
 [-9.77277880e-03  9.50088418e-03 -1.51357208e-03 -1.03218852e-03
   4.10598502e-03]
 [ 1.44043571e-03  1.45427351e-02  7.61037725e-03  1.21675016e-03
   4.43863233e-03]
 [ 3.33674327e-03  1.49407907e-02 -2.05158264e-03  3.13067702e-03
  -8.54095739e-03]
 [-2.55298982e-02  6.53618595e-03  8.64436199e-03 -7.42165020e-03
   2.26975462e-02]
 [-1.45436567e-02  4.57585173e-04 -1.87183850e-03  1.53277921e-02
   1.46935877e-02]
 [ 1.54947426e-03  3.78162520e-03 -8.87785748e-03 -1.98079647e-02
  -3.47912149e-03]
```

```
[ 1.56348969e-03  1.23029068e-02  1.20237985e-02 -3.87326817e-03
 -3.02302751e-03]
[-1.04855297e-02 -1.42001794e-02 -1.70627019e-02  1.95077540e-02
 -5.09652182e-03]
[-4.38074302e-03 -1.25279536e-02  7.77490356e-03 -1.61389785e-02
 -2.12740280e-03]
[-8.95466561e-03  3.86902498e-03 -5.10805138e-03 -1.18063218e-02
 -2.81822283e-04]
[ 4.28331871e-03  6.65172224e-04  3.02471898e-03 -6.34322094e-03
 -3.62741166e-03]
[-6.72460448e-03 -3.59553162e-03 -8.13146282e-03 -1.72628260e-02
  1.77426142e-03]
[-4.01780936e-03 -1.63019835e-02  4.62782256e-03 -9.07298364e-03
  5.19453958e-04]
[ 7.29090562e-03  1.28982911e-03  1.13940068e-02 -1.23482582e-02
  4.02341641e-03]
[-6.84810091e-03 -8.70797149e-03 -5.78849665e-03 -3.11552532e-03
  5.61653422e-04]
[-1.16514984e-02  9.00826487e-03  4.65662440e-03 -1.53624369e-02
  1.48825219e-02]
[ 1.89588918e-02  1.17877957e-02 -1.79924836e-03 -1.07075262e-02
  1.05445173e-02]
[-4.03176947e-03  1.22244507e-02  2.08274978e-03  9.76639036e-03
  3.56366397e-03]
[ 7.06573168e-03  1.05000207e-04  1.78587049e-02  1.26912093e-03
  4.01989363e-03]
[ 1.88315070e-02 -1.34775906e-02 -1.27048500e-02  9.69396708e-03
 -1.17312341e-02]
[ 1.94362119e-02 -4.13618981e-03 -7.47454811e-03  1.92294203e-02
  1.48051479e-02]
[ 1.86755896e-02  9.06044658e-03 -8.61225685e-03  1.91006495e-02
 -2.68003371e-03]
[ 8.02456396e-03  9.47251968e-03 -1.55010093e-03  6.14079370e-03
  9.22206672e-03]
[ 3.76425531e-03 -1.09940079e-02  2.98238174e-03  1.32638590e-02
 -6.94567860e-03]
[-1.49634540e-03 -4.35153552e-03  1.84926373e-02  6.72294757e-03
  4.07461836e-03]
[-7.69916074e-03  5.39249191e-03 -6.74332661e-03  3.18305583e-04
 -6.35846078e-03]
[ 6.76433295e-03  5.76590817e-03 -2.08298756e-03  3.96006713e-03
 -1.09306151e-02]
[-1.49125759e-02  4.39391701e-03  1.66673495e-03  6.35031437e-03
  2.38314477e-02]
[ 9.44479487e-03 -9.12822225e-03  1.11701629e-02 -1.31590741e-02
 -4.61584605e-03]
[-6.82416053e-04  1.71334272e-02 -7.44754822e-03 -8.26438539e-03
 -9.84525244e-04]
```

```
[-6.63478286e-03  1.12663592e-02 -1.07993151e-02 -1.14746865e-02
 -4.37820045e-03]
[-4.98032451e-03  1.92953205e-02  9.49420807e-03  8.75512414e-04
 -1.22543552e-02]
[ 8.44362976e-03 -1.00021535e-02 -1.54477110e-02  1.18802979e-02
  3.16942612e-03]
[ 9.20858824e-03  3.18727653e-03  8.56830612e-03 -6.51025593e-03
 -1.03424284e-02]
[ 6.81594518e-03 -8.03409664e-03 -6.89549778e-03 -4.55532504e-03
  1.74791590e-04]
[-3.53993911e-03 -1.37495129e-02 -6.43618403e-03 -2.22340315e-02
  6.25231451e-03]
[-1.60205766e-02 -1.10438334e-02  5.21650793e-04 -7.39562996e-03
  1.54301460e-02]
[-1.29285691e-02  2.67050869e-03 -3.92828182e-04 -1.16809350e-02
  5.23276661e-03]
[-1.71546331e-03  7.71790551e-03  8.23504154e-03  2.16323595e-02
  1.33652795e-02]
[-3.69181838e-03 -2.39379178e-03  1.09965960e-02  6.55263731e-03
  6.40131526e-03]
[-1.61695604e-02 -2.43261244e-04 -7.38030909e-03  2.79924599e-03
 -9.81503896e-04]
[ 9.10178908e-03  3.17218215e-03  7.86327962e-03 -4.66419097e-03
 -9.44446256e-03]
[-4.10049693e-03 -1.70204139e-04  3.79151736e-03  2.25930895e-02
 -4.22571517e-04]
[-9.55945000e-03 -3.45981776e-03 -4.63595975e-03  4.81481474e-03
 -1.54079701e-02]
[ 6.32619942e-04  1.56506538e-03  2.32181036e-03 -5.97316069e-03
 -2.37921730e-03]
[-1.42406091e-02 -4.93319883e-03 -5.42861476e-03  4.16050046e-03
 -1.15618243e-02]
[ 7.81198102e-03  1.49448454e-02 -2.06998503e-02  4.26258731e-03
  6.76908035e-03]
[-6.37437026e-03 -3.97271814e-03 -1.32880578e-03 -2.97790879e-03
 -3.09012969e-03]
[-1.67600381e-02  1.15233156e-02  1.07961859e-02 -8.13364259e-03
 -1.46642433e-02]
[ 5.21064876e-03 -5.75787970e-03  1.41953163e-03 -3.19328417e-03
  6.91538751e-03]
[ 6.94749144e-03 -7.25597378e-03 -1.38336396e-02 -1.58293840e-02
  6.10379379e-03]
[-1.18885926e-02 -5.06816354e-03 -5.96314038e-03 -5.25672963e-04
 -1.93627981e-02]
[ 1.88778597e-03  5.23891024e-03  8.84220870e-04 -3.10886172e-03
  9.74001663e-04]
[ 3.99046346e-03 -2.77259276e-02  1.95591231e-02  3.90093323e-03
 -6.52408582e-03]
```

```
[-3.90953375e-03  4.93741777e-03 -1.16103939e-03 -2.03068447e-02
  2.06449286e-02]
[-1.10540657e-03  1.02017271e-02 -6.92049848e-03  1.53637705e-02
  2.86343689e-03]
[ 6.08843834e-03 -1.04525337e-02  1.21114529e-02  6.89818165e-03
  1.30184623e-02]
[-6.28087560e-03 -4.81027118e-03  2.30391670e-02 -1.06001582e-02
 -1.35949701e-03]
[ 1.13689136e-02  9.77249677e-04  5.82953680e-03 -3.99449029e-03
  3.70055888e-03]
[-1.30652685e-02  1.65813068e-02 -1.18164045e-03 -6.80178204e-03
  6.66383082e-03]
[-4.60719787e-03 -1.33425847e-02 -1.34671751e-02  6.93773153e-03
 -1.59573438e-03]
[-1.33701560e-03  1.07774381e-02 -1.12682581e-02 -7.30677753e-03
 -3.84879809e-03]
[ 9.43515893e-04 -4.21714513e-04 -2.86887192e-03 -6.16264021e-04
 -1.07305276e-03]
[-7.19604389e-03 -8.12992989e-03  2.74516358e-03 -8.90915083e-03
 -1.15735526e-02]
[-3.12292251e-03 -1.57667016e-03  2.25672350e-02 -7.04700276e-03
  9.43260725e-03]
[ 7.47188334e-03 -1.18894496e-02  7.73252977e-03 -1.18388064e-02
 -2.65917224e-02]
[ 6.06319524e-03 -1.75589058e-02  4.50934462e-03 -6.84010898e-03
  1.65955080e-02]
[ 1.06850940e-02 -4.53385804e-03 -6.87837611e-03 -1.21407740e-02
 -4.40922632e-03]
[-2.80355495e-03 -3.64693544e-03  1.56703855e-03  5.78521498e-03
  3.49654457e-03]
[-7.64143924e-03 -1.43779147e-02  1.36453185e-02 -6.89449185e-03
 -6.52293600e-03]
[-5.21189312e-03 -1.84306955e-02 -4.77974004e-03 -4.79655814e-03
  6.20358298e-03]
[ 6.98457149e-03  3.77088909e-05  9.31848374e-03  3.39964984e-03
 -1.56821116e-04]
[ 1.60928168e-03 -1.90653494e-03 -3.94849514e-03 -2.67733537e-03
 -1.12801133e-02]
[ 2.80441705e-03 -9.93123611e-03  8.41631264e-03 -2.49458580e-03
  4.94949817e-04]
[ 4.93836776e-03  6.43314465e-03 -1.57062341e-02 -2.06903676e-03
  8.80178912e-03]
[-1.69810582e-02  3.87280475e-03 -2.25556423e-02 -1.02250684e-02
  3.86305518e-04]
[-1.65671510e-02 -9.85510738e-03 -1.47183501e-02  1.64813493e-02
  1.64227755e-03]
[ 5.67290278e-03 -2.22675101e-03 -3.53431749e-03 -1.61647419e-02
 -2.91837363e-03]
```

```
[-7.61492212e-03  8.57923924e-03  1.14110187e-02  1.46657872e-02
  8.52551939e-03]
[-5.98653937e-03 -1.11589699e-02  7.66663182e-03  3.56292817e-03
 -1.76853845e-02]
[ 3.55481793e-03  8.14519822e-03  5.89255892e-04 -1.85053671e-03
 -8.07648488e-03]
[-1.44653470e-02  8.00297949e-03 -3.09114445e-03 -2.33466662e-03
  1.73272119e-02]
[ 6.84501107e-03  3.70825001e-03  1.42061805e-03  1.51999486e-02
  1.71958931e-02]
[ 9.29505111e-03  5.82224591e-03 -2.09460307e-02  1.23721914e-03
 -1.30106954e-03]
[ 9.39532294e-04  9.43046087e-03 -2.73967717e-02 -5.69312053e-03
  2.69904355e-03]
[-4.66845546e-03 -1.41690611e-02  8.68963487e-03  2.76871906e-03
 -9.71104570e-03]
[ 3.14817205e-03  8.21585712e-03  5.29264630e-05  8.00564803e-03
  7.82601752e-04]
[-3.95228983e-03 -1.15942052e-02 -8.59307670e-04  1.94292938e-03
  8.75832762e-03]
[-1.15107468e-03  4.57415606e-03 -9.64612014e-03 -7.82629156e-03
 -1.10389299e-03]
[-1.05462846e-02  8.20247837e-03  4.63130329e-03  2.79095764e-03
  3.38904125e-03]
[ 2.02104356e-02 -4.68864188e-03 -2.20144129e-02  1.99300197e-03
 -5.06035410e-04]
[-5.17519043e-03 -9.78829859e-03 -4.39189522e-03  1.81338429e-03
 -5.02816701e-03]
[ 2.41245368e-02 -9.60504382e-03 -7.93117363e-03 -2.28862004e-02
  2.51484415e-03]
[-2.01640663e-02 -5.39454633e-03 -2.75670535e-03 -7.09727966e-03
  1.73887268e-02]
[ 9.94394391e-03  1.31913688e-02 -8.82418819e-03  1.12859406e-02
  4.96000946e-03]
[ 7.71405949e-03  1.02943883e-02 -9.08763246e-03 -4.24317621e-03
  8.62596011e-03]
[-2.65561909e-02  1.51332808e-02  5.53132064e-03 -4.57039607e-04
  2.20507656e-03]
[-1.02993528e-02 -3.49943365e-03  1.10028434e-02  1.29802197e-02
  2.69622405e-02]
[-7.39246663e-04 -6.58552967e-03 -5.14233966e-03 -1.01804188e-02
 -7.78547559e-04]]
```

(b) Run a Pricipal Components(PCA)

We perform PCA using the correlation matrix since the variables are uncorrelated.

[2]:
```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA

# Standardize the data
scaler = StandardScaler()
data_standardized = scaler.fit_transform(data)

# Run PCA
pca = PCA()
pca.fit(data_standardized)
```

[2]:
```
PCA()
```

(c) Explain the Variance of Each component

The explained variance ratio tells us how much variance in the data is explained by each principal component

[10]:
```
# c. Explained variance by each component
explained_variance = pca.explained_variance_ratio_
print(f"Explained variance by Component 1: {explained_variance[0]:.2f}")
print(f"Explained variance by Component 2: {explained_variance[1]:.2f}")
print(f"Explained variance by Component 3: {explained_variance[2]:.2f}")
```

```
Explained variance by Component 1: 0.25
Explained variance by Component 2: 0.23
Explained variance by Component 3: 0.21
```

Explanation:

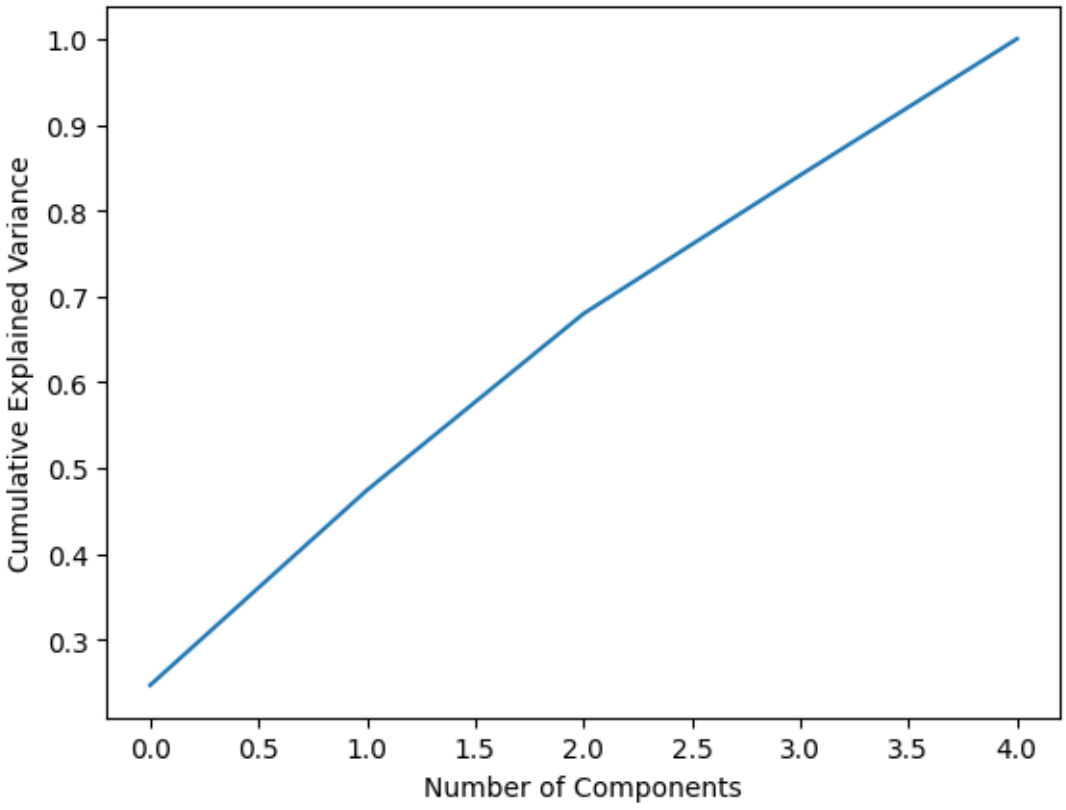
The first principal component explains the highest amount of variance, followed by the second and third components. The cumulative explained variance gives an idea of how much total variance is captured by the first few components.

(d) Produce a screeplot

A scree plot helps to visually assess the importance of each principal component.

[3]:
```
import matplotlib.pyplot as plt

# Scree plot
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('Number of Components')
plt.ylabel('Cumulative Explained Variance')
plt.savefig("figure/scree_plot_random_data.png")
plt.show()
```



1.2 Real data Analysis

(e)Collect aily closing yields for 5 government securities

We collect daily closing yields for 5 US government securities over the last 6 months using the FRED API.

[5]:
```
import pandas as pd
from fredapi import Fred

# FRED API key
fred = Fred(api_key="95eb212842318d85c6198945d6514bf4")

# List of FRED series IDs for 5 US government securities
series_ids = ['DGS1MO',  'DGS3MO',  'DGS6MO',  'DGS1', 'DGS2']


#Define the date range for the last 6 months
end_date = pd.Timestamp.today().strftime('%Y-%m-%d')
start_date = (pd.Timestamp.today() - pd.DateOffset(months=6)).
 ↪strftime('%Y-%m-%d')
```

```
# Collect the data
data = {}
for series_id in series_ids:
    data[series_id] = fred.get_series(series_id, start_date, end_date)


# Convert to a DataFrame
df = pd.DataFrame(data)

# Display the DataFrame
print(df)

# Optionally, save the DataFrame to a CSV file
df.to_csv('data/us_government_securities_yields.csv')
```

```
            DGS1MO  DGS3MO  DGS6MO  DGS1  DGS2
2024-07-22    5.49    5.43    5.24  4.88  4.50
2024-07-23    5.49    5.41    5.22  4.85  4.40
2024-07-24    5.50    5.40    5.19  4.82  4.37
2024-07-25    5.49    5.39    5.19  4.83  4.41
2024-07-26    5.49    5.38    5.18  4.79  4.36
…              …       …       …     …     …
2025-01-13    4.42    4.37    4.30  4.24  4.40
2025-01-14    4.42    4.36    4.29  4.22  4.37
2025-01-15    4.40    4.35    4.26  4.19  4.27
2025-01-16    4.43    4.34    4.26  4.18  4.23
2025-01-17    4.43    4.34    4.28  4.21  4.27

[130 rows x 5 columns]
```

(f) Compute the daily yield changes

[6]:
```
daily_changes = df.diff()
print(daily_changes)
```

```
            DGS1MO  DGS3MO  DGS6MO  DGS1  DGS2
2024-07-22     NaN     NaN     NaN   NaN   NaN
2024-07-23    0.00   -0.02   -0.02 -0.03 -0.10
2024-07-24    0.01   -0.01   -0.03 -0.03 -0.03
2024-07-25   -0.01   -0.01    0.00  0.01  0.04
2024-07-26    0.00   -0.01   -0.01 -0.04 -0.05
…              …       …       …     …     …
2025-01-13    0.00    0.01    0.03 -0.01  0.00
2025-01-14    0.00   -0.01   -0.01 -0.02 -0.03
2025-01-15   -0.02   -0.01   -0.03 -0.03 -0.10
2025-01-16    0.03   -0.01    0.00 -0.01 -0.04
2025-01-17    0.00    0.00    0.02  0.03  0.04

[130 rows x 5 columns]
```

(g) Re-run PCA using the correlation Matrix

[7]:
```
# Calculate the correlation matrix
corr_matrix = df.corr()

# Perform PCA
pca = PCA(n_components=len(series_ids))
pca.fit(corr_matrix)
```

[7]:
```
PCA(n_components=5)
```

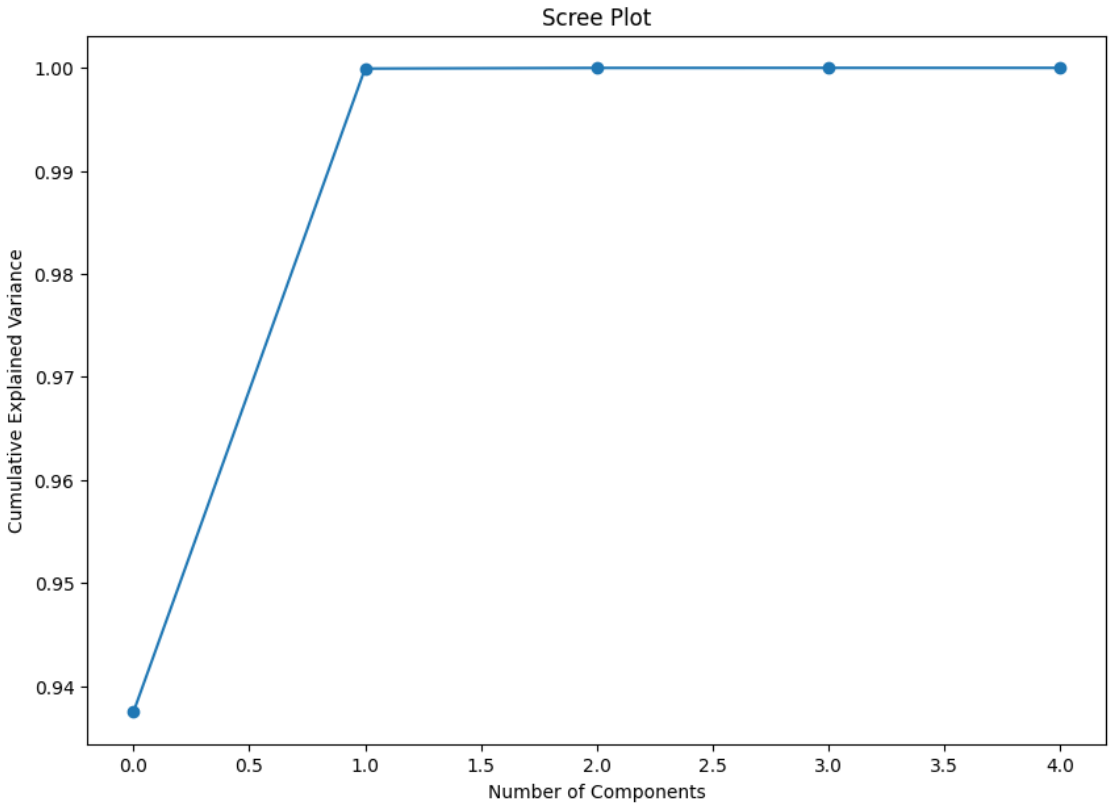
(h) Explain the variance of each component

[8]:
```
explained_variance = pca.explained_variance_ratio_
print("Explained variance by each component:", explained_variance)
```

```
Explained variance by each component: [9.37487603e-01 6.24465507e-02
5.99912996e-05 5.85461747e-06
 1.18932054e-33]
```

(i) Produce a scree plot

[9]:
```
plt.figure(figsize=(10, 7))
plt.plot(np.cumsum(explained_variance), marker='o')
plt.xlabel('Number of Components')
plt.ylabel('Cumulative Explained Variance')
plt.title('Scree Plot')
plt.savefig("figure/scree_plot_real_data.png")
plt.show()
```



(j) Compare Scree Plots

The scree plot from the uncorrelated data shows a more even distribution of explained variance across the components, indicating that no single component dominates. In contrast, the scree plot from the government data likely shows a steeper decline, indicating that the first few components explain a significant portion of the variance. This difference highlights the structured nature of real-world financial data compared to random, uncorrelated data.

References

- Federal Reserve Economic Data (FRED). (n.d.). Federal Reserve Bank of St. Louis. Retrieved from https://fred.stlouisfed.org/
- Scree plot. (n.d.). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Scree_plot