Functional data analysis of stocks during COVID-19

Antonio De Patto, Danial Yntykbay, Jackie Islam

Vilnius University - Functional Data Analyses project

May 27, 2025

Table of contents

Goal

- To study the variability of returns of stocks belonging to different industries during COVID-19 period.
- To study stocks and industries that were affected by COVID-19 and the ones that were not

Table of contents

Data Presentation

- Stocks data from 2020/01/01 until 2022/12/31
- 8 main industrial sectors were considered:
- automobile
- fashion and clothing
- food and beverage
- healthcare
- tech
- logistic
- oil and gas
- travel and tourism
- For each sector 8 stocks were considered
- The closing price for each end of week was considered
- Total of 156 weeks

The stocks considered

- The main stocks were:
  - **Automobile**: VW - Ferrari - Stellantis - Renault - Mercedes - BMW - Tesla - Toyota
  - **Fashion and clothing**: Kering - Capri - Hermès - LVMH - Richemont - Adidas - Nike - Puma
  - **Food and beverage**: Nestlè - Uniliver - Danone - Bonduelle - Pepsi - McDonalds - Kellogs
  - **Healthcare**: Sanofi - Novartis - Bayer - AstraZeneca - UCB - Merck - Argenx - GSK
  - **Tech**: Spotify - Netflix - Nvidia - Meta - Apple - IBM - Microsoft - Google
  - **Logistic**: Zalando - UPS - Amazon - DHL - FedEx - Maersk - Walmart - SF express
  - **Oil and Gas**: Shell - Eni - Enel - Engie - Orsted - Chevron - Repsol
  - **Travel and tourism**: Trivago - Booking - Ryanair - Lyft - Trip.com - Tripadvisor - Hilton - Uber

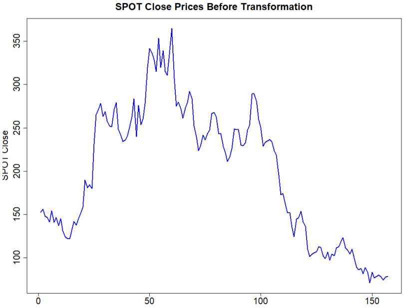
Data transformation

Logarithmic return:

$$R = 100 \cdot \log\left(\frac{P_t}{P_{t-1}}\right)$$

- R is the log return
- P_t is the stock price at time t
- P_{t-1} is the stock price at the previous time period

Before vs. After Transformations



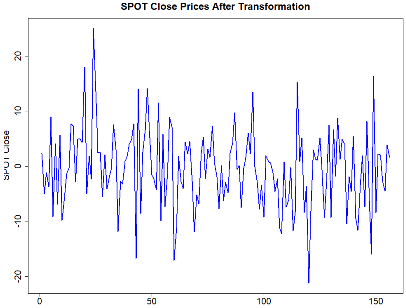


Figure: Before vs. After Transformations.

Table of contents

Euclidean depth

- Euclidean depth of the deepest point in the dataset: 0.00002917722

```
> st[mdE,]
     ZAL.DE.Close         UPS.Close        AMZN.Close       DHL.DE.Close
         47.25000         207.17999         146.81750           42.44000
        FDX.Close       AMKBY.Close         WMT.Close  X002352.SZ.Close
        219.28000          15.27000          47.54333           53.69000
       SPOT.Close        NFLX.Close        NVDA.Close         META.Close
        148.91000         380.14999          25.93100          231.84000
       AAPL.Close         IBM.Close        MSFT.Close         GOOG.Close
        175.06000         128.89000         310.88000          141.06300
    VOW3.DE.Close     RACE.MI.Close    STLAM.MI.Close       RNO.PA.Close
        151.00000         177.89999          14.15600           22.99000
     MBG.DE.Close      BMW.DE.Close        TSLA.Close           TM.Close
         62.14000          75.32000         267.29666          165.92999
     KER.PA.Close        CPRI.Close      RMS.PA.Close        MC.PA.Close
        557.20001          50.16000        1113.50000          590.00000
     CFR.SW.Close      ADS.DE.Close      NKE.DE.Close       PUM.DE.Close
        104.00000         203.70000         108.44000           71.86000
     SAN.PA.Close     NOVN.SW.Close     BAYN.DE.Close          AZN.Close
         93.46938          75.47266          55.94000           61.37000
     UCB.BR.Close      MRK.DE.Close        ARGX.Close        GSK.L.Close
        100.50000         179.20000         286.47000         1581.55164
    NESN.SW.Close          UL.Close       BN.PA.Close       BON.PA.Close
        115.74000          44.39000          52.76000           17.18000
        PEP.Close         MCD.Close           K.Close          KHC.Close
        159.00000         232.57001          57.57747           37.82000
       SHEL.Close      ENI.MI.Close     ENEL.MI.Close      ENGI.PA.Close
         50.35000          12.92400           5.82600           11.53400
  ORSTED.CO.Close       CHV.F.Close      REP.MC.Close       TTE.PA.Close
        829.59998         144.72000          11.46800           45.75500
     SHEL.Close.1    ENI.MI.Close.1   ENEL.MI.Close.1    ENGI.PA.Close.1
         50.35000          12.92400           5.82600           11.53400
ORSTED.CO.Close.1     CHV.F.Close.1    REP.MC.Close.1     TTE.PA.Close.1
        829.59998         144.72000          11.46800           45.75500
```

```
> apply(st,2,median)
     ZAL.DE.Close         UPS.Close        AMZN.Close       DHL.DE.Close
         66.55000         175.50500         155.11150           40.22500
        FDX.Close       AMKBY.Close         WMT.Close  X002352.SZ.Close
        229.71000          11.60500          46.21500           61.58500
       SPOT.Close        NFLX.Close        NVDA.Close         META.Close
        223.00000         480.54001          14.61325          249.07000
       AAPL.Close         IBM.Close        MSFT.Close         GOOG.Close
        135.38000         128.15134         245.14500          104.76375
    VOW3.DE.Close     RACE.MI.Close    STLAM.MI.Close       RNO.PA.Close
        151.89000         177.42500          13.50200           30.13000
     MBG.DE.Close      BMW.DE.Close        TSLA.Close           TM.Close
         57.85313          76.27000         225.39667          153.31499
     KER.PA.Close        CPRI.Close      RMS.PA.Close        MC.PA.Close
        564.75000          46.55000        1066.25000          601.10001
     CFR.SW.Close      ADS.DE.Close      NKE.DE.Close       PUM.DE.Close
         97.14000         256.25000         110.13000           77.19500
     SAN.PA.Close     NOVN.SW.Close     BAYN.DE.Close          AZN.Close
         86.79584          76.67636          53.35000           56.31500
     UCB.BR.Close      MRK.DE.Close        ARGX.Close        GSK.L.Close
         87.80000         155.32500         293.18500         1503.27991
    NESN.SW.Close          UL.Close       BN.PA.Close       BON.PA.Close
        109.47158          54.25500          56.19500           20.20000
        PEP.Close         MCD.Close           K.Close          KHC.Close
        148.05500         233.28500          60.92958           36.13000
       SHEL.Close      ENI.MI.Close     ENEL.MI.Close      ENGI.PA.Close
         41.60500          10.52400           7.22800           12.13100
  ORSTED.CO.Close       CHV.F.Close      REP.MC.Close       TTE.PA.Close
        842.39999          89.35000          10.63039           39.74750
     SHEL.Close.1    ENI.MI.Close.1   ENEL.MI.Close.1    ENGI.PA.Close.1
         41.60500          10.52400           7.22800           12.13100
ORSTED.CO.Close.1     CHV.F.Close.1    REP.MC.Close.1     TTE.PA.Close.1
        842.39999          89.35000          10.63039           39.74750
```

Comparison of the deepest point vs. columns-wise medians

- some values (e.g., UPS.Close, MSFT.Close, GOOG.Close) in the deepest point differ significantly from the median. This may suggest that the dataset contain outliers affecting the distribution

Euclidean depth

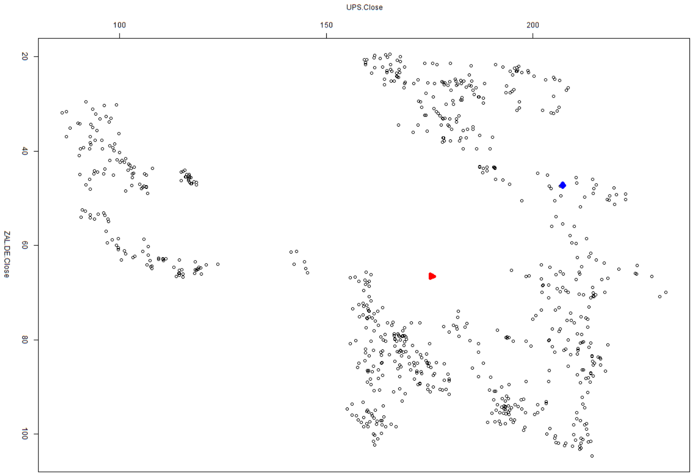


Figure: Comparison between the deepest point(in blue) and the median point(in red) in a scatter plot

Minimum Volume Depth (MBD)

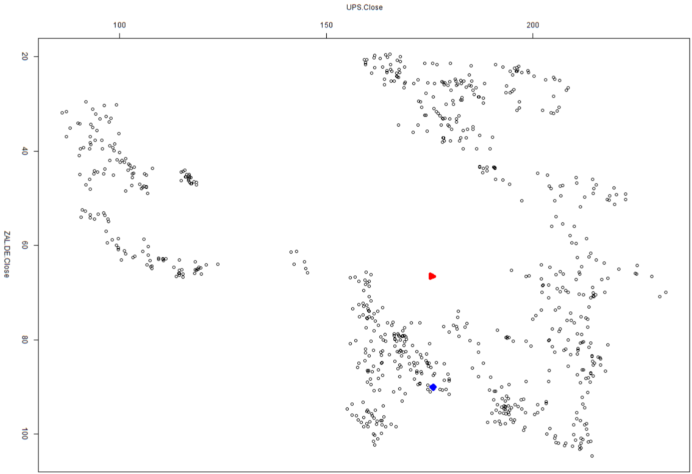


Figure: Comparison between the deepest point(in blue) and the median point(in red) in a scatter plot

Frainman-Muniz depth

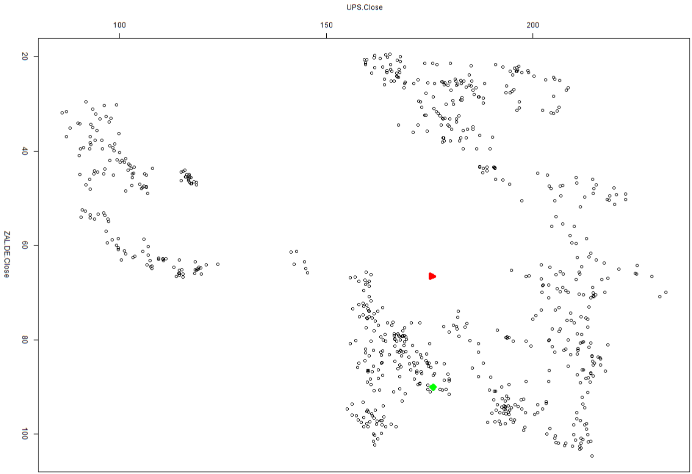


Figure: Comparison between the deepest point(in green) and the median point(in red) in a scatter plot

Table of contents

Smoothing methods

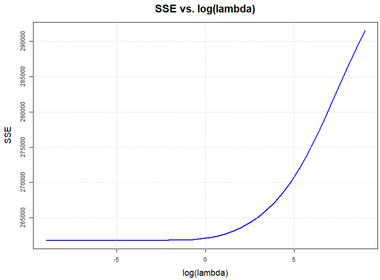
Methods:

- B-splines
- Local regression kernel
- Nadaraya-Watson kernel
- Normal kernel
- Triweight kernel
- Uniform kernel

B-splines results

Smoothing with penalty:

- Generalized cross-validation
  - Optimal lambda ~ 371
  - Number of basis ~ 7
  - GCV ~ 2230
  - SSE ~ 307923



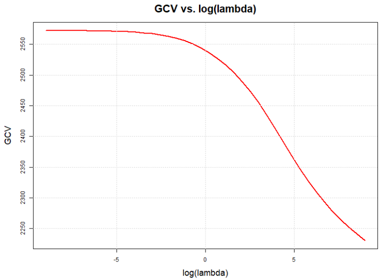


Figure: Comparison of SSE and GCV with respect to lambda.

Kernel smoothing results

Generalized cross-validation

- Sequence of bandwidth between 3 and 70

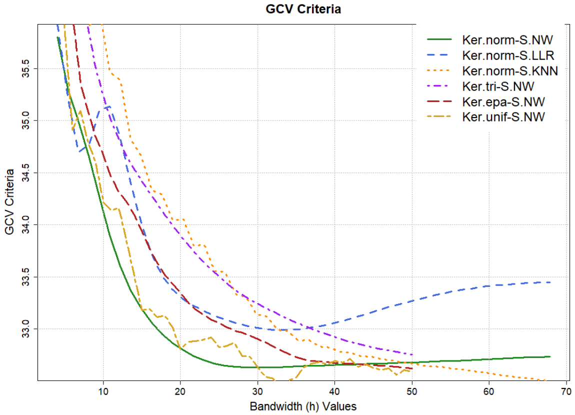


Figure: GCV criteria respect to bandwidth

Comparing optimal GCV and SSE

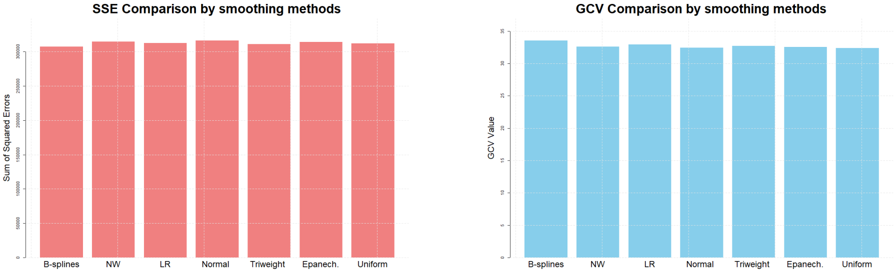


Figure: Comparison of optimal GCV and SSE among smoothing methods.

B-splines vs. Normal kernel

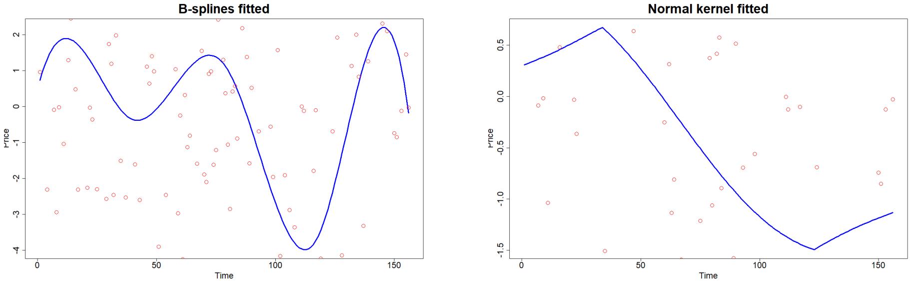


Figure: Comparison of fitted and actual values for B-splines and Normal kernel.

B-splines vs. Triweight kernel

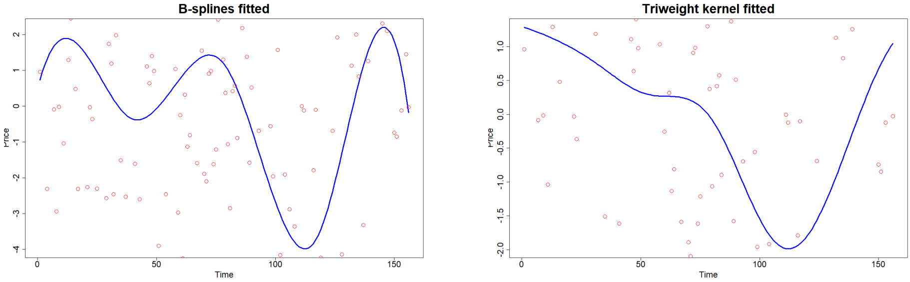


Figure: Comparison of optimal GCV and SSE among smoothing methods.

B-splines fitted

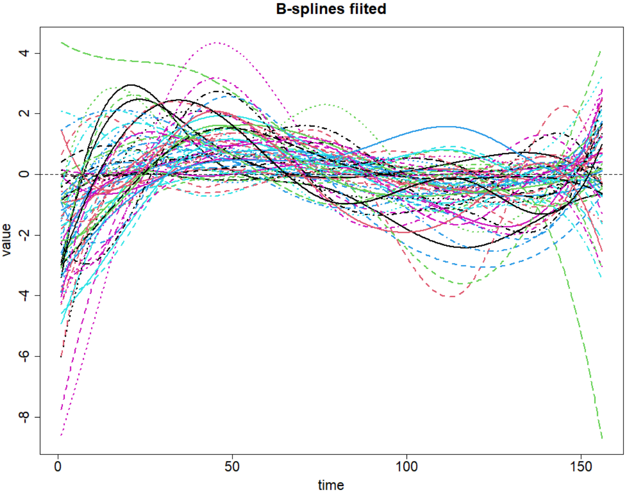


Figure: B-splines fitted for all stocks

Table of contents

Derivatives

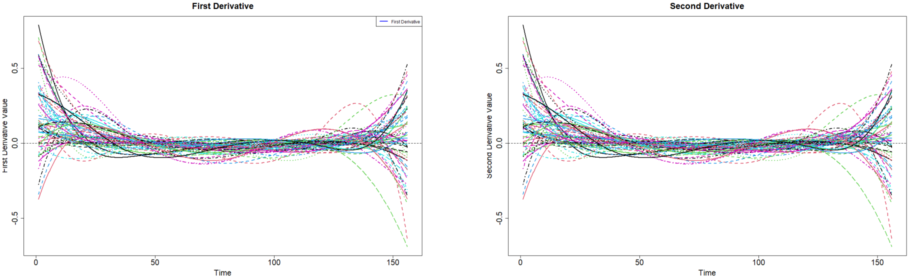


Figure: 1st and 2nd derivatives of B-splines smoothed

Outliers detection

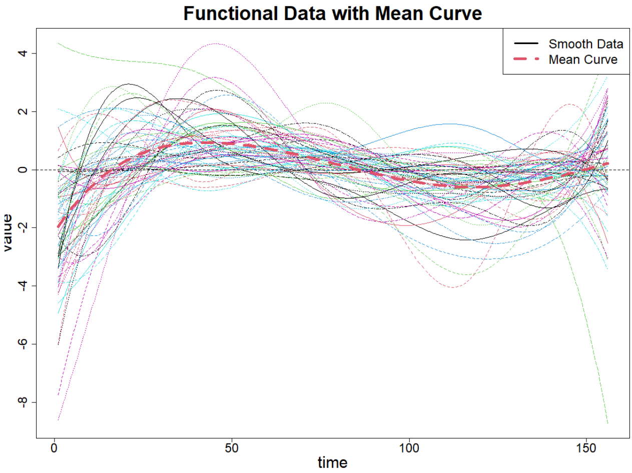


Figure: Mean curve

Euclidean depth

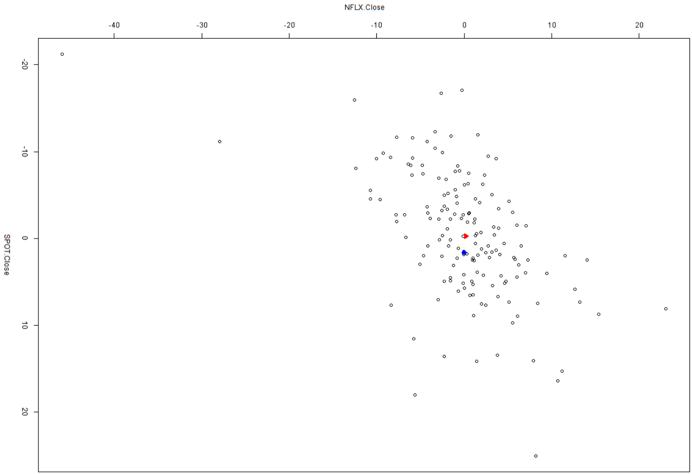


Figure: Comparison between the deepest point(in blue) and the median point(in red) in a scatter plot

Frainman-Muniz and The Most Central Point

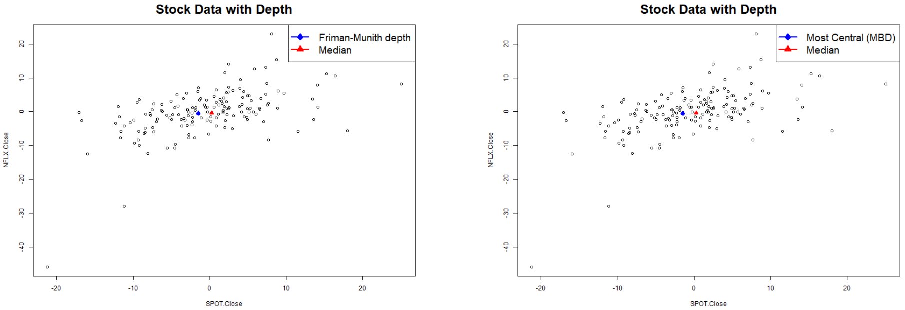


Figure: Friman-Munith depth and MBD

Bivariate Covariance Function

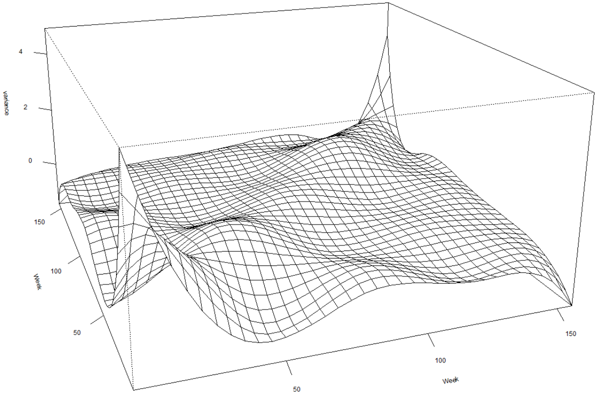


Figure: 3D visualization of the variance function for smoothed stock returns

Contour plot

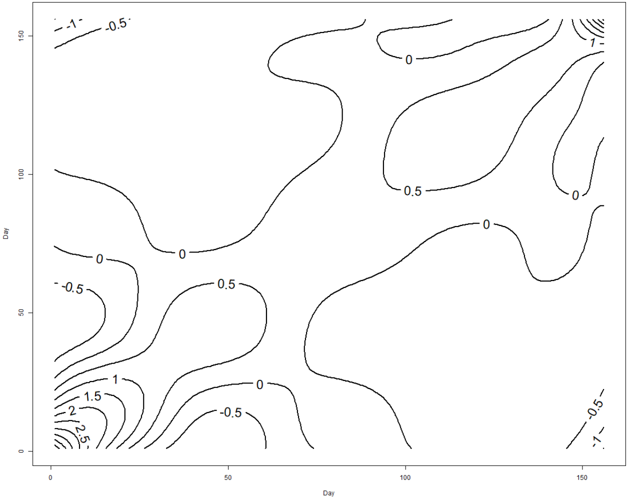


Figure: Contour plot

Table of contents

Principal Component Analysis

FPCA Component	**Percentage Contribution**
PCA Function 1	34%
PCA Function 2	24%
PCA Function 3	15%
PCA Function 4	9%
Total	**82%**

Table: FPCA Component Contributions

VARIMAX Rotation

PCA Function (Varimax Rotation)	**Percentage Contribution**
PCA Function 1	27.8%
PCA Function 2	18.3%
PCA Function 3	13.0%
PCA Function 4	20.7%
Total	**79.8%**

Table: PCA Function Contributions after Varimax Rotation

VARIMAX Rotation

- PCA 4 captured 20.7% and showed strong variability from the mean among stocks till the 70th week(could explain starting phase)
- PCA 1 may capture the overall market variability across different phases in response to specific events possibly COVID-19 related.

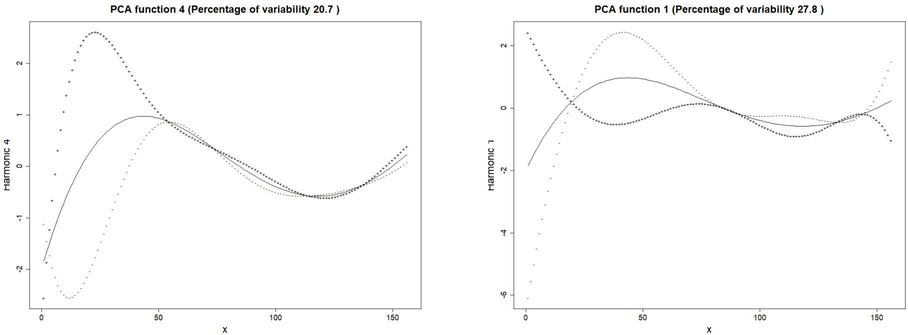


Figure: VARIMAX rotation of FPCA 1 and FPCA 4

VARIMAX Rotation

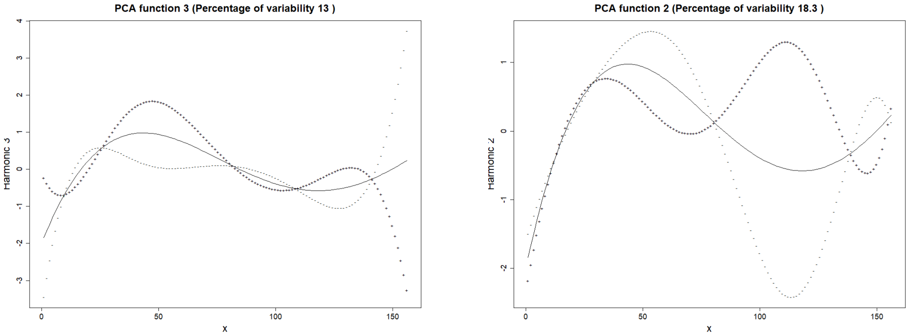


Figure: VARIMAX rotation of FPCA 2 and FPCA 3

- Volatility in stock prices in beginning, mid and end of pandemic
- PCA Functions captured the most volatility in different time periods

Table of contents

The choice of the best number of clusters

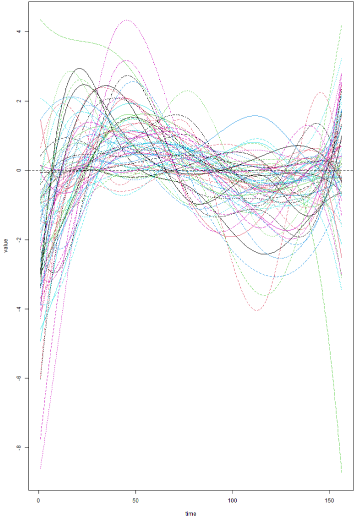
```
>> K = 2
Error in .fstep(fd, T, lambda) : One cluster is almost empty!
>> K = 3
AkjBk   :        bic = -350051.4
>> K = 4
AkjBk   :        bic = -318623.1
>> K = 5
AkjBk   :        bic = -198106.2
>> K = 6
Error in .fstep(fd, T, lambda) : One cluster is almost empty!
>> K = 7
Error in .fstep(fd, T, lambda) : One cluster is almost empty!
>> K = 8
Error in .fstep(fd, T, lambda) : One cluster is almost empty!
>> K = 9
Error in .fstep(fd, T, lambda) : One cluster is almost empty!
>> K = 10
Error in .fstep(fd, T, lambda) : One cluster is almost empty!
The best model is AkjBk with K = 5 ( bic = -198106.2 )
```

	K	model	bic	aic	icl	nbprm	ll
1	2	AkjBk	NA	NA	NA	NA	NA
2	3	AkjBk	-350051.4	-350021.2	-350051.4	28	-349993.2
3	4	AkjBk	-318623.1	-318573.5	-318623.1	46	-318527.5
4	5	AkjBk	-198106.2	-198033.9	-198106.2	67	-197966.9
5	6	AkjBk	NA	NA	NA	NA	NA
6	7	AkjBk	NA	NA	NA	NA	NA
7	8	AkjBk	NA	NA	NA	NA	NA
8	9	AkjBk	NA	NA	NA	NA	NA
9	10	AkjBk	NA	NA	NA	NA	NA

Figure: The choice of the best number of clusters to consider

- The best clustering model selected is AkjBk with $K = 5$ clusters, based on the highest BIC score (-198106.2)
- Models with $K > 5$ resulted in empty clusters.

From original data to clusters



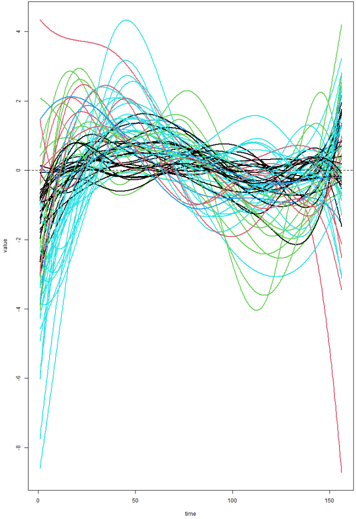


Figure: Comparison between original smoothed data and clusters

Centroids

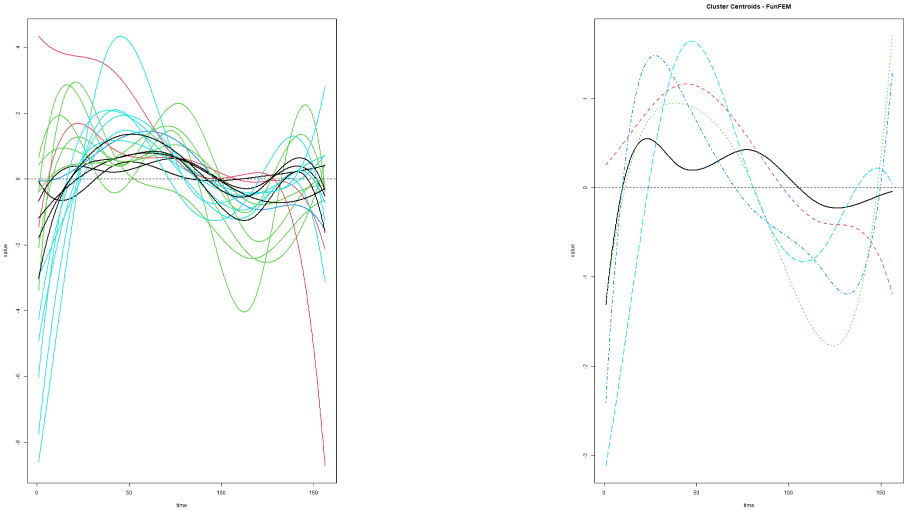
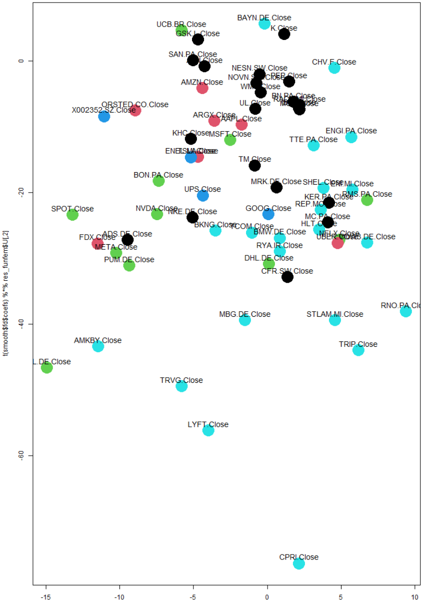
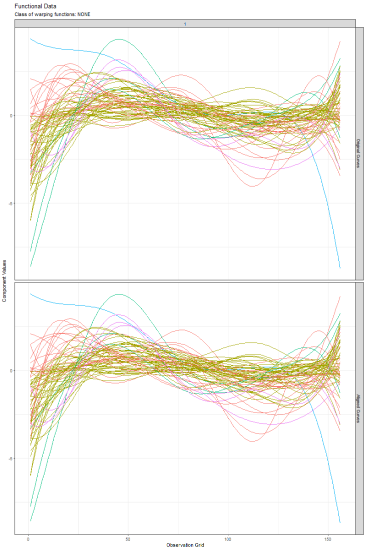


Figure: A more in-depth analyses comparing the first 20 curves(on the left) and the mean of each cluster(on the right)

Discriminative Space Plot



Hierarchical Clustering (HCLUST)



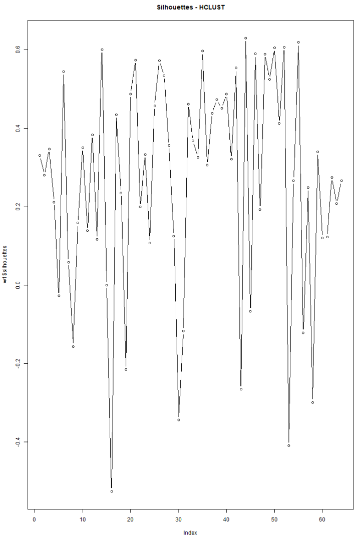


Figure: Hclust with 5 clusters and complete linkage method

HCLUST interpretation

Silhouette Plot for HCLUST Evaluation: assesses the quality of our hierarchical clustering results.

Coefficient Interpretation:

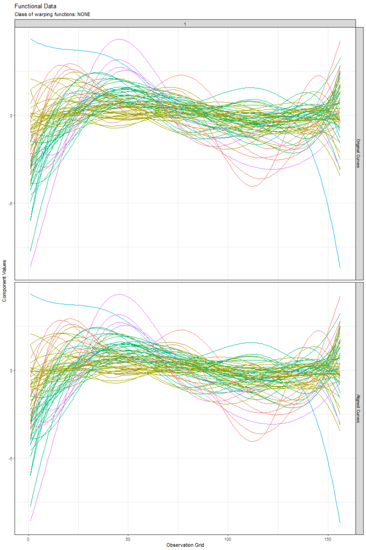
- Values close to $+1$: Well-clustered data point.
- Values around 0: Data point near the cluster boundary.
- Values close to -1: Potential misclassification.

Key Observations:

- **Variable Coefficients:** Significant differences in silhouette values across data points.
- **Negative Values Present:** Some data points (e.g., around indices 17 and 58) show $s(i)<0$, indicating likely misclassification.
- **Values Near Zero:** Many points have s(i)0, suggesting borderline assignments.
- **Well-Defined Clusters:** Some groups exhibit higher positive s(i) (e.g., indices 45-55), indicating good cluster structure.

Overall Conclusion: The silhouette plot suggests that while some clusters are well-formed while other need further investigations

K-means Clustering



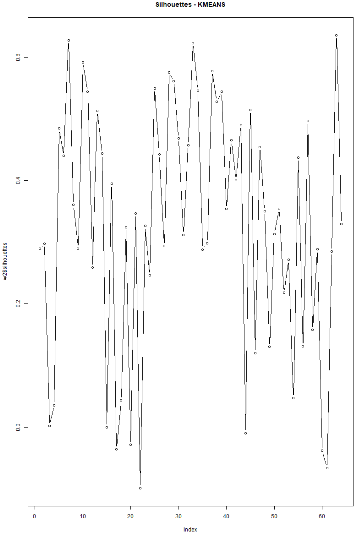


Figure: K-means clustering with 5 clusters and mean used for centroid computation

KMEANS interpretation

Observations from the Plot:

- Predominantly positive silhouette coefficients suggest a reasonable clustering structure.
- Variability in coefficients indicates that some points fit their clusters better than others.
- Several coefficients near zero warrant attention as these points are close to cluster boundaries.
- The absence of strongly negative coefficients implies minimal severe misclassification.

Overall Conclusion: The silhouette plot indicates a moderately good K-Means clustering result. A higher average silhouette score (which can be calculated from these individual values) would signify better overall cluster quality.

DBSCAN Clustering

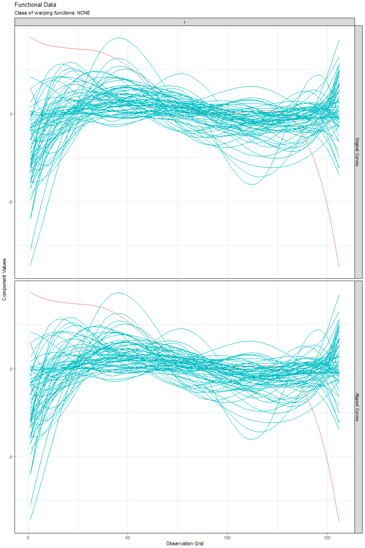


Figure: Results from DBSCAN clustering methods. Red function is Tesla

Interpretations

```
> dt_w[dt_w$FunFEM == 1,]
       StockName FunFEM Hclust Kmeans Dbscan
6      IBM.Close      1      2      2      1
10 RACE.MI.Close      1      2      2      1
16      TM.Close      1      1      2      1
17  KER.PA.Close      1      2      3      1
20   MC.PA.Close      1      2      3      1
21  CFR.SW.Close      1      2      3      1
22  ADS.DE.Close      1      2      3      1
23  NKE.DE.Close      1      2      2      1
25  SAN.PA.Close      1      2      2      1
26 NOVN.SW.Close      1      2      2      1
28     AZN.Close      1      2      2      1
30  MRK.DE.Close      1      1      2      1
32   GSK.L.Close      1      2      2      1
33 NESN.SW.Close      1      2      2      1
34      UL.Close      1      2      2      1
35   BN.PA.Close      1      2      2      1
37     PEP.Close      1      2      2      1
38     MCD.Close      1      2      2      1
39       K.Close      1      2      2      1
40     KHC.Close      1      2      2      1
63     WMT.Close      1      2      2      1
```

Figure: Stocks collected in cluster 1 according to the FunFEM method

Interpretations

```
> dt_w[dt_w$FunFEM == 2,]
         StockName FunFEM Hclust Kmeans Dbscan
5       AAPL.Close      2      1      2      1
15      TSLA.Close      2      4      4      0
31      ARGX.Close      2      1      2      1
45 ORSTED.CO.Close      2      1      2      1
56      UBER.Close      2      1      2      1
59      AMZN.Close      2      1      2      1
61       FDX.Close      2      2      3      1
> dt_w[dt_w$FunFEM == 3,]
      StockName FunFEM Hclust Kmeans Dbscan
1    SPOT.Close      3      1      1      1
2    NFLX.Close      3      1      1      1
3    NVDA.Close      3      1      1      1
4    META.Close      3      1      1      1
7    MSFT.Close      3      1      2      1
19 RMS.PA.Close      3      1      2      1
24 PUM.DE.Close      3      2      2      1
29 UCB.BR.Close      3      1      2      1
36 BON.PA.Close      3      2      2      1
57 ZAL.DE.Close      3      1      1      1
60 DHL.DE.Close      3      2      3      1
> dt_w[dt_w$FunFEM == 4,]
          StockName FunFEM Hclust Kmeans Dbscan
8         GOOG.Close      4      1      2      1
43     ENEL.MI.Close      4      1      2      1
58          UPS.Close      4      1      2      1
64 X002352.SZ.Close      4      1      2      1
```

Figure: Stocks collected in cluster 2-3-4 according to the FunFEM method

Interpretations

```
> dt_w[dt_w$FunFEM == 5,]
        StockName FunFEM Hclust Kmeans Dbscan
9   VOW3.DE.Close      5      2      3      1
11 STLAM.MI.Close      5      2      3      1
12    RNO.PA.Close     5      3      3      1
13    MBG.DE.Close     5      2      3      1
14    BMW.DE.Close     5      2      3      1
18      CPRI.Close     5      3      5      1
27   BAYN.DE.Close     5      2      2      1
41      SHEL.Close     5      2      3      1
42    ENI.MI.Close     5      2      3      1
44   ENGI.PA.Close     5      2      3      1
46     CHV.F.Close     5      2      3      1
47    REP.MC.Close     5      2      3      1
48    TTE.PA.Close     5      2      3      1
49      TRVG.Close     5      5      5      1
50      BKNG.Close     5      2      3      1
51    RYA.IR.Close     5      2      3      1
52      LYFT.Close     5      5      5      1
53      TCOM.Close     5      3      3      1
54      TRIP.Close     5      5      5      1
55       HLT.Close     5      2      3      1
62     AMKBY.Close     5      2      3      1
```

Figure: Stocks collected in cluster 5 according to the FunFEM method

Interpretations

- Applied FunFEM, hierarchical, k-means, and DBSCAN clustering on smoothed weekly stock price functions.
- **Cluster 1:** Stable, mature companies (e.g., IBM, WMT, MCD) — likely low-volatility, defensive stocks - related to different economic sectors
- **Cluster 2:** High-growth, tech-oriented firms (e.g., AAPL, TSLA, AMZN) — showing dynamic, possibly volatile trends.
- **Cluster 3:** Digital and streaming-focused stocks (e.g., SPOT, NVDA, META) — potentially similar usage trends or momentum patterns.
- **Cluster 4:** Smaller, mixed group — includes GOOG and UPS, may indicate intermediate or hybrid behaviors.
- **Cluster 5:** Automotive, energy, and travel sectors (e.g., BMW, ENI, TRIP) — exhibiting cyclical, macro-sensitive patterns.
- Consistency observed between FunFEM, k-means, and hierarchical clustering, especially for Clusters 1 and 5.
- DBSCAN detected one outlier(Tesla), indicating strong internal cohesion in the smoothed data.

Table of contents

Mean Trajectories

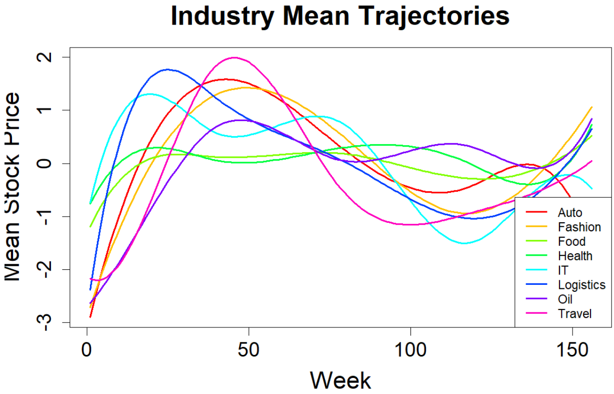


Figure: Mean Trajectories

ANOVA Pointwise

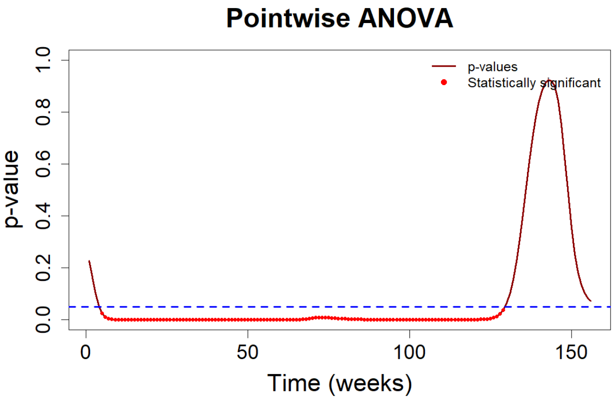


Figure: Pointwise ANOVA: Industry Differences Over Time

Post-Hoc Test 1

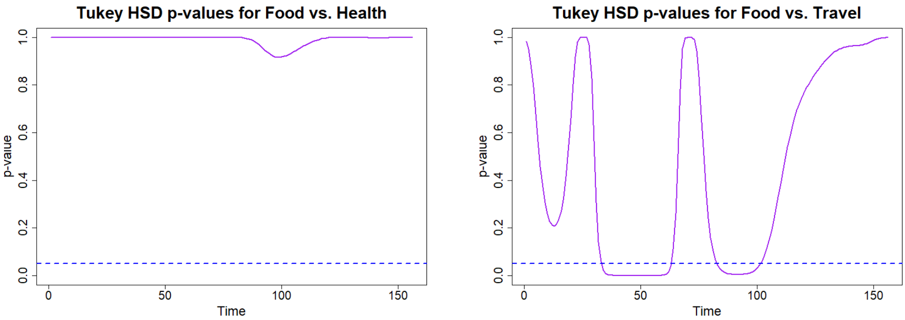


Figure: Tukey p-values

Post-Hoc Test 2

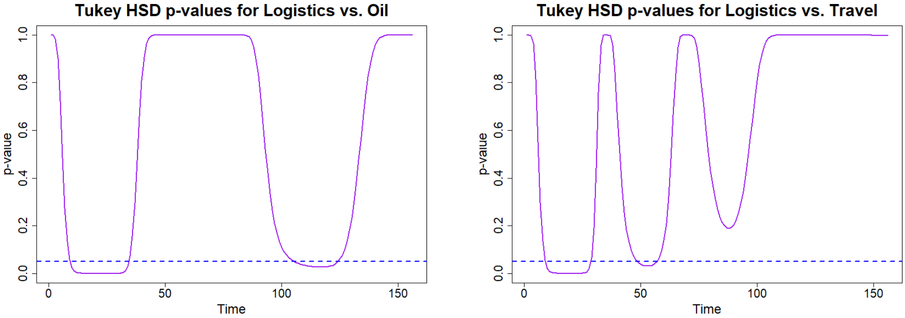
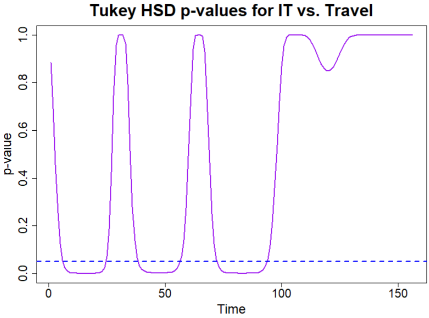


Figure: Tukey p-values

Post-Hoc Test 3



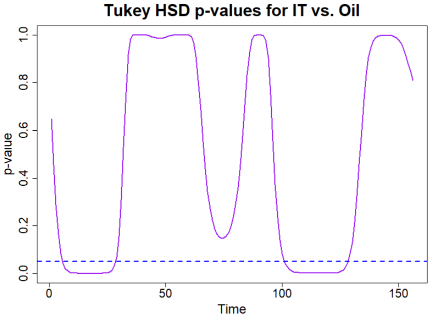


Figure: Tukey p-values

Post-Hoc Test 4

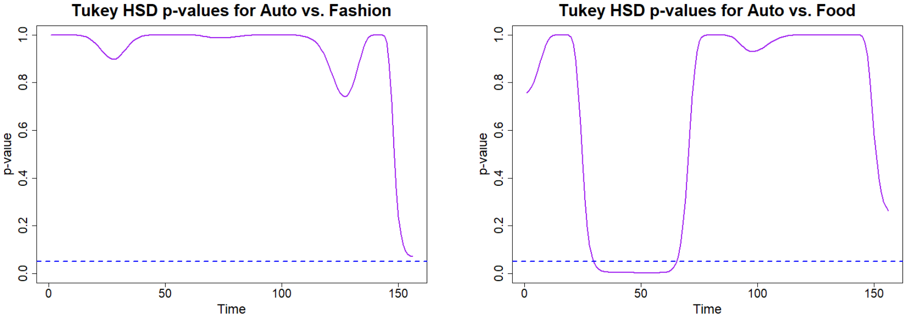


Figure: Tukey p-values

Post-Hoc Test 5

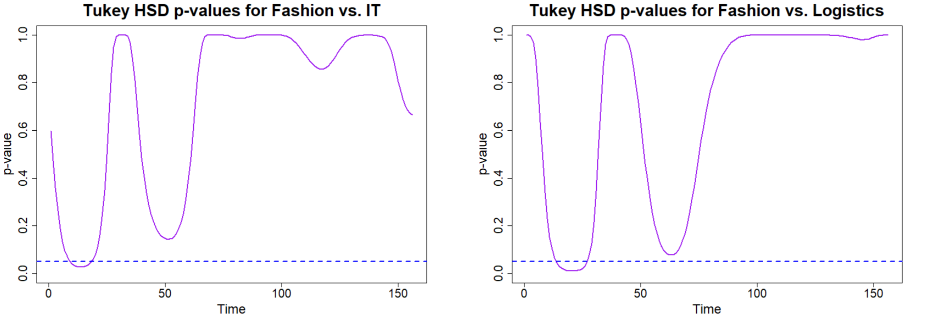
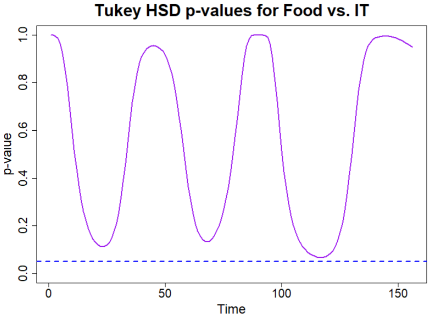


Figure: Tukey p-values

Post-Hoc Test 6



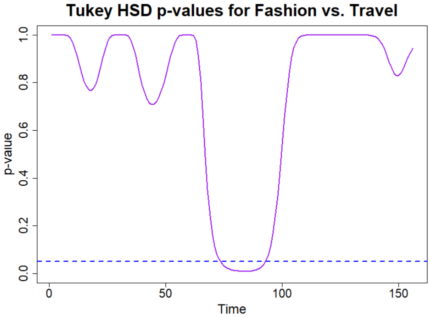


Figure: Tukey p-values

Table of contents

FAR Data Format

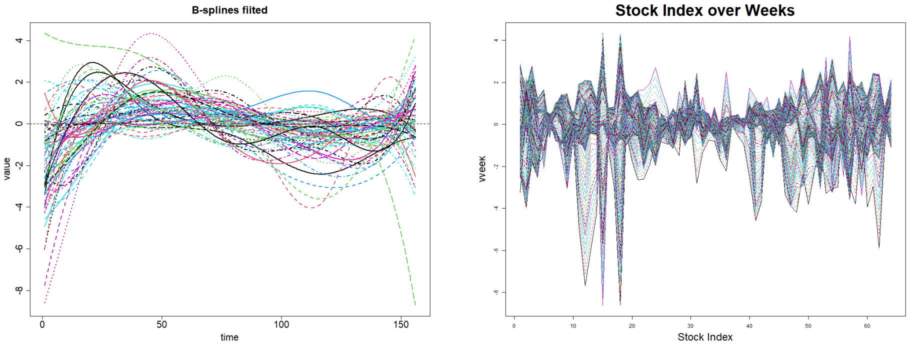


Figure: Change of Data Format

Parameters FAR(1)

Parameters	
Train set	146 weeks
Test set	10 weeks
Optimal K-NN by GCV	4
Past returns order	1

Table: Parameters

FAR(1) Fitted and Residuals

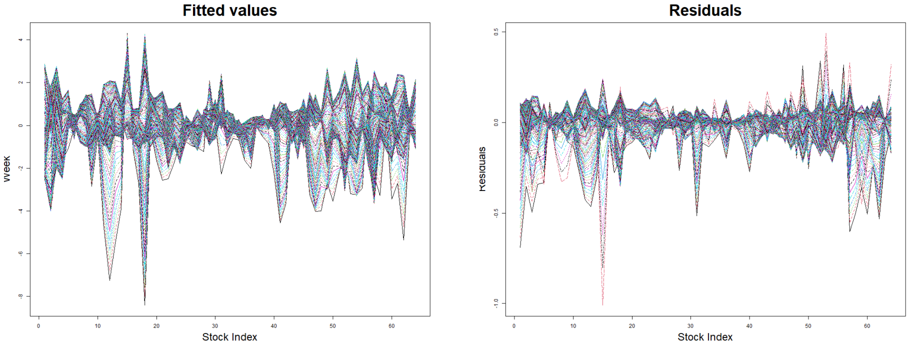


Figure: Fitted and Residuals

Forecasting on Test set

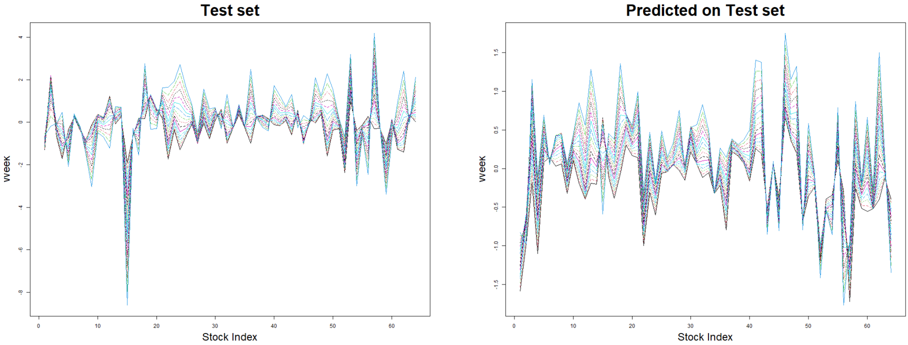


Figure: Prediction on Test set

FPCA + FAR(3)

PCA Function	**Percentage Contribution**
PCA Function 1	57%
PCA Function 2	26%
PCA Function 3	6%
Total	**89%**

Table: PCA Function Contributions to FAR

Fitted and Residuals

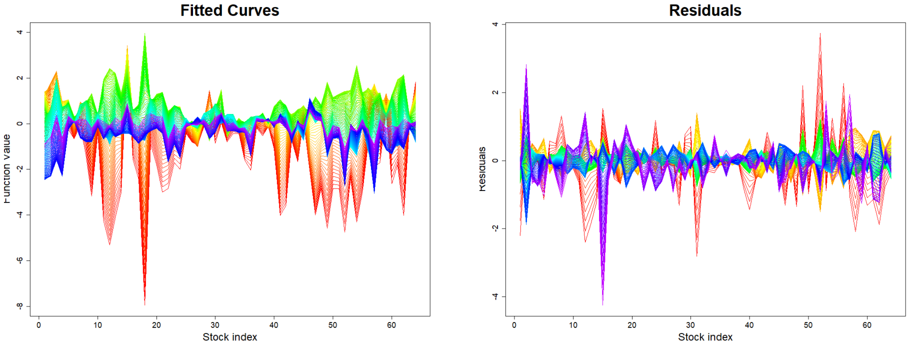


Figure: Fitted and Residuals

Forecasting on Test set

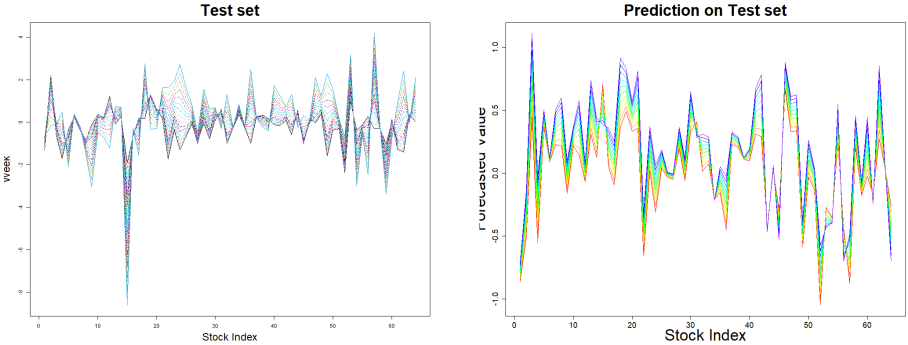


Figure: Prediction on Test set

Results

Model	MSE
FAR(1)	0.05
FPCA + FAR(3)	0.16

Table: MSE Results on Train set

Model	MSE
FAR(1)	1.38
FPCA + FAR(3)	1.82

Table: MSE Results on Test set

Table of contents

Conclusion

- No evident clusters belonging to 8 sectors
- PCA and First Derivative showed clear variation in the beginning and ending of Pandemic
- First constant derivative after the start of Covid-19 and at the end of this indicates how the pandemic has influenced the price movement of stocks
- From Hypothesis only Health and Food industries showed the same pattern
- Both models showed weak predictive accuracy, FAR(1) showed lower MSE than FPCA

Thank you for your attention