FE507 TERM PROJECT

Abstract

For the term project of FE507, Different financial data sets had been analyzed

Mehmetcan Caniklioglu

Table of Contents

Introduction

For the term project of FE507, Different financial data sets had been analyzed and for different time intervals with different time units. Main decision variables under consideration were return data derived from either index values or price data. For different parts of the project, different return statistics were examined such as daily, weekly, monthly, yearly returns. The analysis conducted during the project was applied to the data with Python programming language which employs a rich library of analysis tools and a high ability of developing algorithms for preparing data from raw format with daily index and price values to unable format with weekly, monthly return values.

- Important notice for reader; for BISTALL and BIST100 indexes FX refers to Turkish Lira

Data Preparation

As mentioned in the introduction part, python programming language was used to conduct the study during the project. Data preparation section can be divided into four different sub-sections; Getting Packages and Data, Preparing Weekly Data, Preparing Monthly Data and Merge to Build Master Data Frames and Clear Intermediate Data frames

Getting Packages and Data

This section involves importing essential packages to the working environment (Jupyter Notebook) and afterwards; getting the data from excel files to the working environment as well. Packages used in this project their brief explanations are given below:

- pandas: For data preparation
- matplotlib: For visual exploration
- datetime: to work with time series data
- numpy: to make advanced matrix operations
- math: to make advanced mathematical calculations
- seaborn: For visual exploration
- scipy: to calculate advanced statistical analysis
- statsmodels: to calculate advanced statistical analysis
- statistics: to calculate advanced statistical analysis

After initializing the necessary packages phase of importing data was conducted and the raw data was imported as python pandas' data frames.

```
df_SP500=pd.read_excel('DATA FE507 2021.xlsx',sheet_name='SP500')
df_BIST100=pd.read_excel('DATA FE507 2021.xlsx',sheet_name='BIST100')
df_BISTAll=pd.read_excel('DATA FE507 2021.xlsx',sheet_name='BIST All')
df_BTC_ETH=pd.read_excel('DATA FE507 2021.xlsx', sheet_name='Bitcoin and Ethereum')
df_Gold=pd.read_excel('DATA FE507 2021.xlsx', sheet_name='Gold')
df_TLUSD=pd.read_excel('DATA FE507 2021.xlsx',sheet_name='TL USD')
df_TRInf=pd.read_excel('DATA FE507 2021.xlsx', sheet_name='TR Inflation')
df_TRRF=pd.read_excel('DATA FE507 2021.xlsx', sheet_name='TR Deposit Rate')
```

Imported raw data frames are as follows:

- df_SP500: Data frame containing daily index and market cap values of SP500.
- df_BIST100: Data frame containing daily index and market cap values of BIST100.
- df_BISTAll: Data frame containing daily index and market cap values of BISTALL.
- df_BTC_ETH: Data frame containing daily price values of BTC and ETH.
- df_Gold: Data frame containing daily price values of Gold.
- df_TLUSD: Daily FX rate.
- df_TRInf: Monthly Inflation rate in Turkey.
- df_TRRF: Quarterly nominal interest rate in Turkey.

```
In [7]: df_BTC_ETH
```

Out[7]:

	Date	Bitcoin	Ethereum
0	2014-11-04	324.467934	NaN
1	2014-11-05	328.644408	NaN
2	2014-11-06	337.921358	NaN
3	2014-11-07	348.992860	NaN
4	2014-11-08	341.459753	NaN
...	...	...	...
2591	2021-12-08	50638.162863	4310.180119
2592	2021-12-09	50512.038512	4440.333251
2593	2021-12-10	47594.381934	4106.330000
2594	2021-12-11	47162.324050	3901.851456
2595	2021-12-12	49353.419314	4082.900000

Figure 1 Example of Raw Data Frame

Preparing Weekly Data

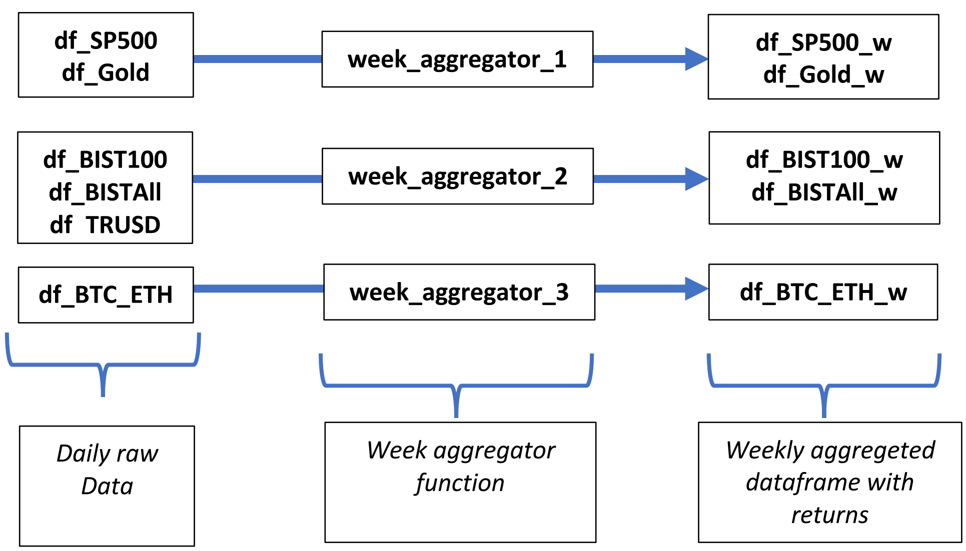
To prepare new data frames representing weekly returns, three different functions were built according to the different dynamics of datasets. The Important issue here was to develop different algorithms to determine the closing and opening days of each week since different investment options have different trading days due to geographical differences (BIST vs SP500) and structural differences of markets (BTC/ETH vs. SP500).

First a function (week_aggregator_1) finding and tagging the closing days of weeks for USD based 5-Days traded time series was built and used for Gold and SP500 datasets.

Secondly, a function (week_aggregator_2) finding and tagging the closing days of weeks for FX based 5-Days traded time series was built and used to determine weekly returns of BISTALL and BIST100 indexes.

Lastly, a function (week_aggregator_3) finding and tagging the closing days of weeks for USD based 7-Days traded time series (ETH/BTC) and used to build a weekly return data frame for weekly BT and ETH returns.

Functions and data frames built and used in this section are as follows:



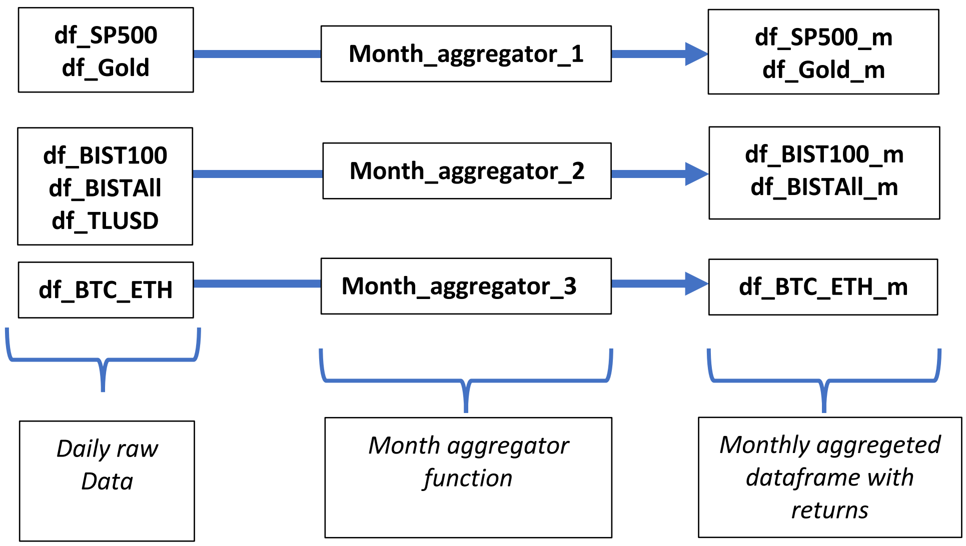
df_BTC_ETH_w

	Bitcoin Price	Ethereum Price	Week_ID	BTC Weekly Return	ETC Weekly Return
0	344.745289	NaN	2014-11-03	NaN	NaN
1	374.983975	NaN	2014-11-10	0.084077	NaN
2	352.080105	NaN	2014-11-17	-0.063025	NaN
3	375.964613	NaN	2014-11-24	0.065636	NaN
4	375.097528	NaN	2014-12-01	-0.002309	NaN
...	...	...	...	...	...

Figure 2 Example of weekly Aggregated Data frames

Prepare Monthly Data

Similar process to the process applied in the preparation of weekly data, was developed to obtain data frames containing monthly return data for different investment options. The main difference was the algorithm built to determine the opening and closing days of each month. Functions and data frames built and used in this section are as follows:




```
In [370]: df_BIST100_m.head(100)
Out[370]:
```

	BIST100 Index	BIST100 Market_cap (m TL)	TL/USD	Month_ID	BIST100 Monthly Return in FX	BIST100 Monthly Return in USD
0	0.09	2	0.00111	19881	NaN	NaN
1	0.07	2	0.00118	19882	-0.251314	-0.312469
2	0.06	2	0.00122	19883	-0.154151	-0.187487
3	0.06	2	0.00126	19884	0.000000	-0.032261
4	0.06	2	0.00132	19885	0.000000	-0.046520
...	...	...	...	...	...	...

Figure 3 Example of monthly Aggregated Data frames

Furthermore, in some analysis yearly and daily data were used as well in this report, but their data preparation processes were done under corresponding analysis sections.

Merge to Build Master Data Frames and Clear Intermediate Data Frames

After building weekly and monthly data frames for each individual asset, master tables were generated by renaming, merging and re-indexing operations applied to data frames. Functions and data frames built and used in this section are as follows:


```
In [371]: df_master_w
Out[371]:
```

	SP500 Index	SP500 Market_cap (m$)	Year_ID	Week_ID	SP500 Weekly Return	Gold Price ($/t oz)	Gold Weekly Return	Bitcoin Price	Ethereum Price	BTC Weekly Return	ETC Weekly Return	BIST100 Index	BIST100 Market_cap (m TL)	TL/USD
0	79.94	357907	1964	1964-03-30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	79.85	357907	1964	1964-04-06	-0.001126	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	80.55	357907	1964	1964-04-13	0.008728	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	79.75	357907	1964	1964-04-20	-0.009981	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	80.17	363019	1964	1964-04-27	0.005253	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3006	4682.85	39739140	2021	2021-11-08	-0.003130	1862.36	0.028258	64414.909083	4646.859294	0.045716	0.027196	1638.50	1297522.0	9.98715
3007	4697.96	39866160	2021	2021-11-15	0.003221	1859.66	-0.001451	59752.144235	4415.109310	-0.075140	-0.051159	1737.33	1375482.0	11.23000

Figure 4 Weekly Master Data Frame

Analysis

This section of the report contains both visual and analytical findings about the random nature of datasets given.

Graphical Analysis of BIST100 and SP500

For 1990-1999 time period we see that an upward trend had been affective specially after 1995 for SP500 index and carried the index value from 500s to over 1200s. While on the other hand we see that similar upward trend, even steeper, had been observed in BIST100 after 1997. Though, it will be sensible to highlight a shock can be observed in BIST100 time series, while we see that its effect had recovered quickly. In addition, we may spot this shock in SP500 time series as well.

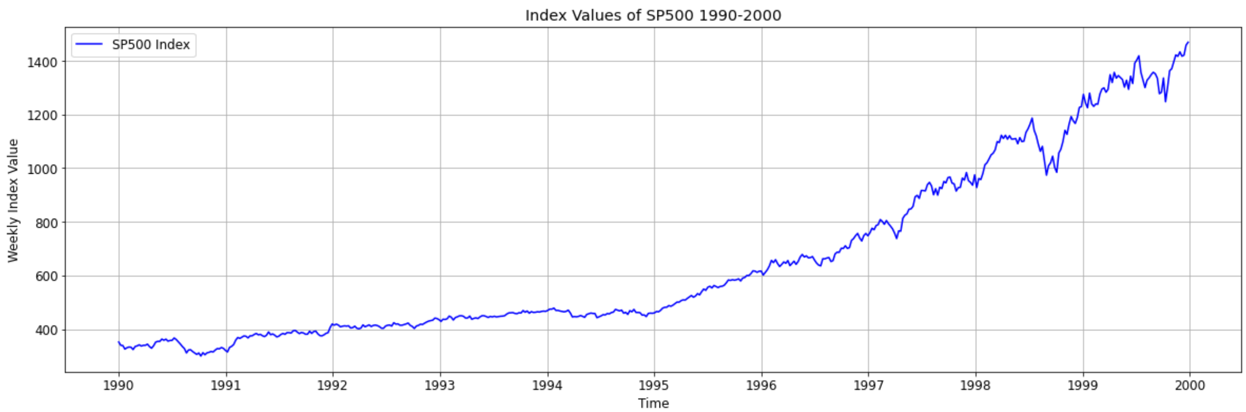


Figure 5 Index Values of SP500 1990-2000

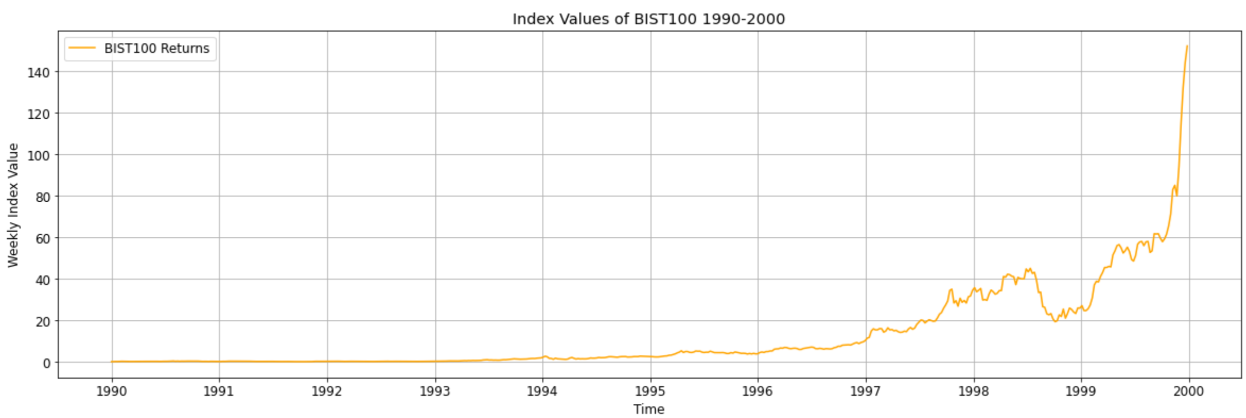


Figure 6 Index Values of BIST100 1990-2000

When we checked the return numbers for this time period, first thing that we noticed that both series seems to act like stationary series with mean near 0. Another key finding that can be derived from the figures given below is that Turkey observed a shock in 1994 where variance, therefore risk, had increased sharply for a while.

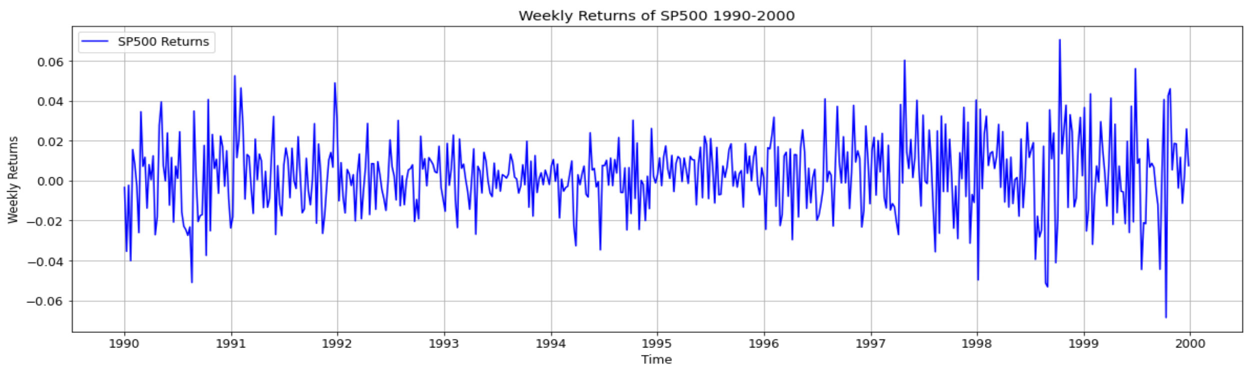


Figure 7 Weekly Returns of SP500 1990-2000

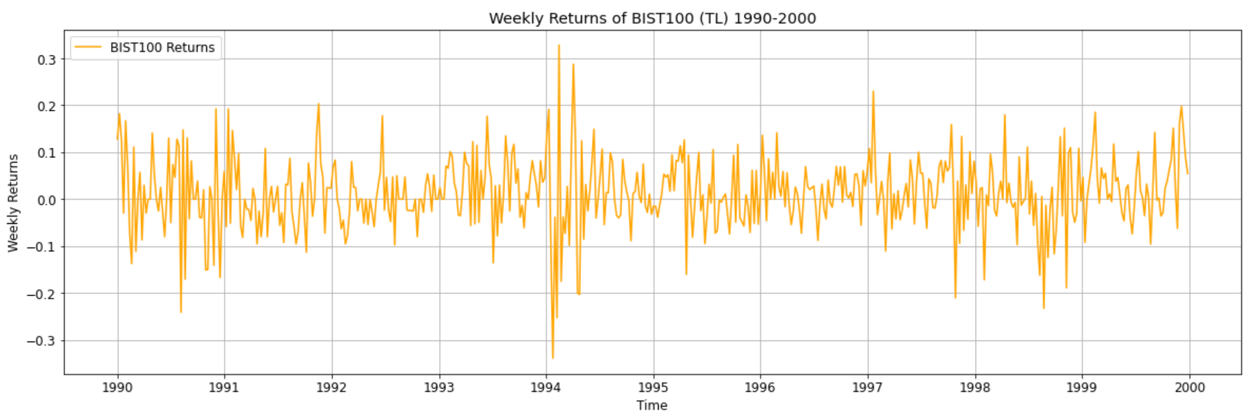


Figure 8 Weekly Returns of BIST100 (TL) 1990-2000

Also, the figures show that before 1995 risk behavior of two asset acted similarly though after 1995, we see that variance of SP500 increased significantly more than BIST100. Though when we checked BIST100 returns in form of USD we see that the risk behavior acted similarly for both assets.

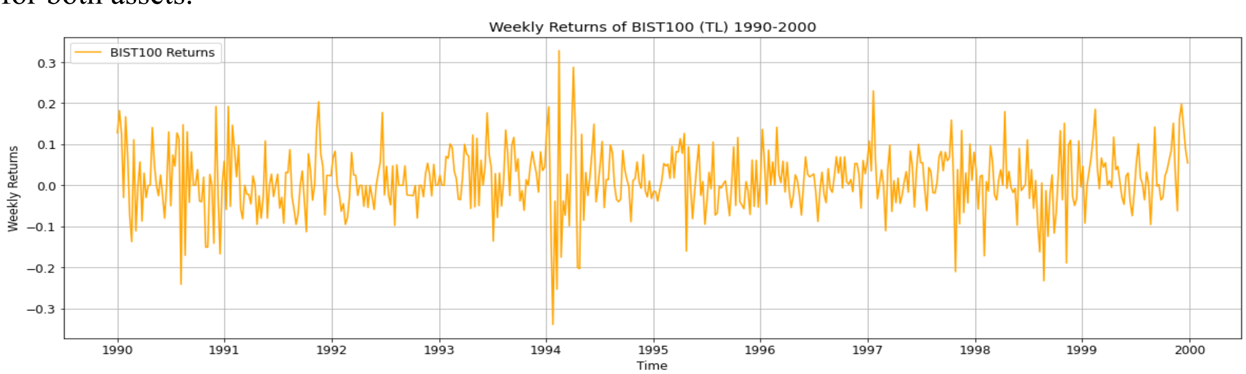


Figure 9 Weekly Returns of BIST100 (USD) 1990-2000

When we checked the index values for time period 200-2009, we see a downward trend for both assets lasted almost 3 years from 2000 to 2003 where SP500 index lost more than 33% of its value and BIST100 lost similar proportion of its value. After 2003 wee see upward trends for both series and in 2008 we noticed that both series had experienced sharp declines that can be associated with global financial crisis.

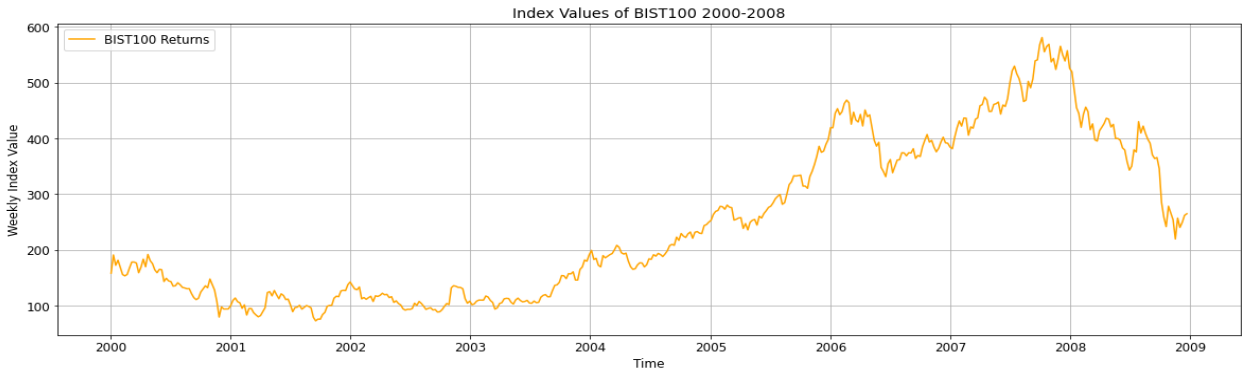


Figure 10 Index Values of BIST100 2000-2008

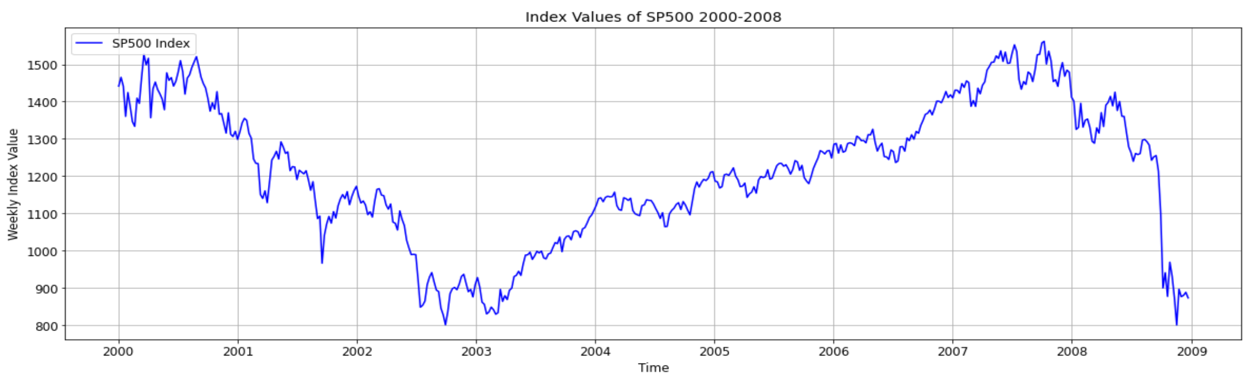


Figure 11 Index Values of SP500 2000-2008

When we analyze the weekly return data for our two assets, SP500 and BIST100, we see that the risk in both markets had been increasing since now we observe higher variance in both assets than their correspondents in 1990-1999, specially for SP500. While in addition for both figures we notice extreme points have extreme values, 2001 for BIST100 and 2008 for SP500, that can be related to economic phenomena occurred at that time.

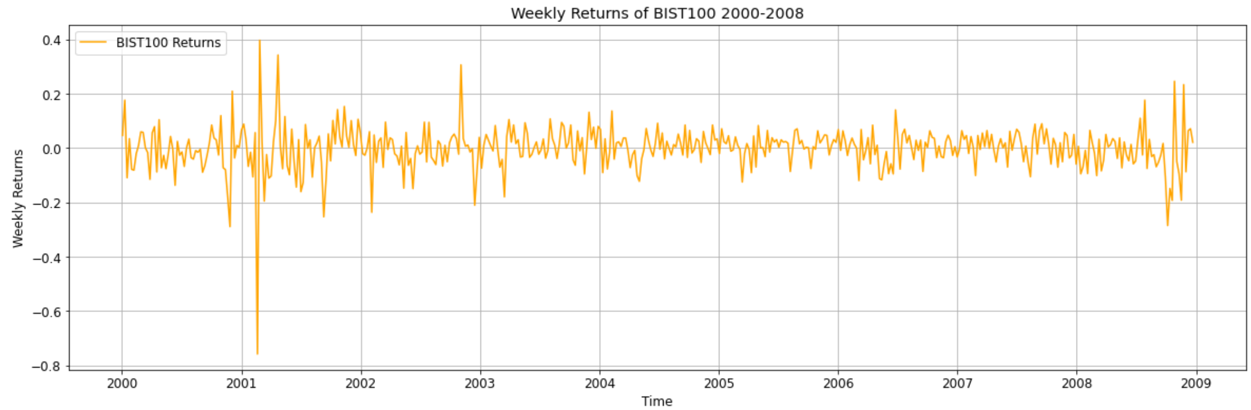


Figure 12 Weekly Returns of BIST100 2000-2008

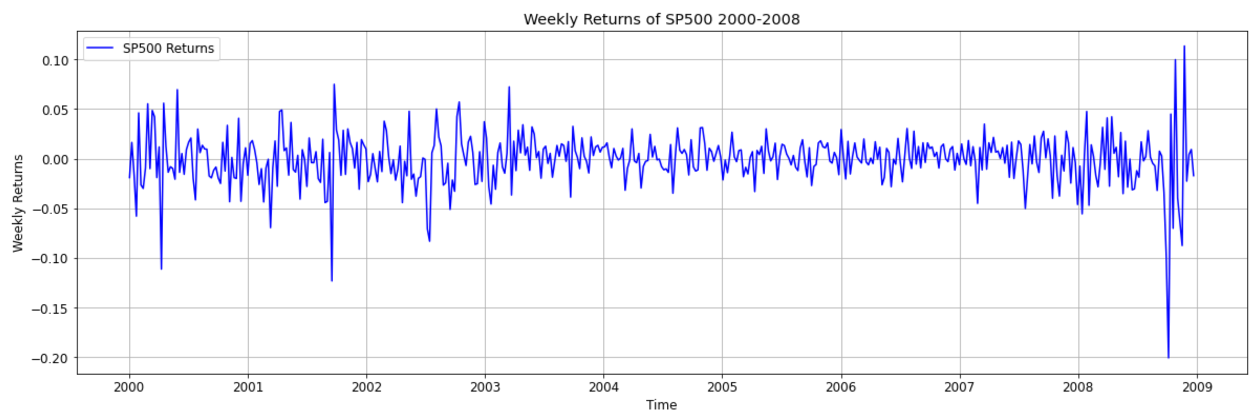


Figure 13 Weekly Returns of SP500 2000-2008

Other than extreme events and behavior of variance, we could visually claim that the series acts similar to stationary series.

For time period 2009-2021 we, again, observe an upward trend in general. While SP 500 Index shows a more consistent behavior, BIST100 index experienced a more Fluctuated increase. When we checked overall increase, we notice that both indexes multiplied their value around 8-9 times over 11 years which is the most improvement we observed among tree different time period. Another interesting finding is that the Covid 19 pandemic's affect over indexes is less than the affect that we observed in 2008 financial crisis.

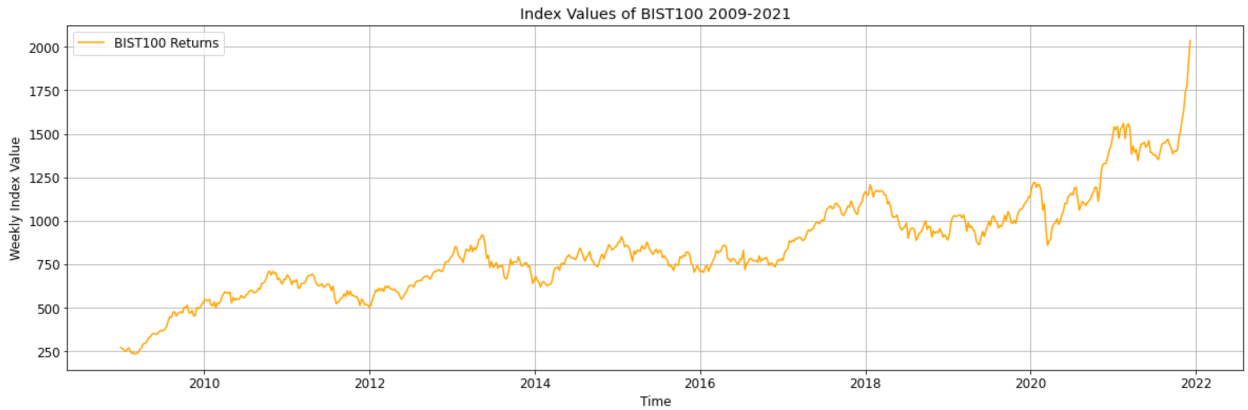


Figure 14 Index Values of BIST100 2009-2021

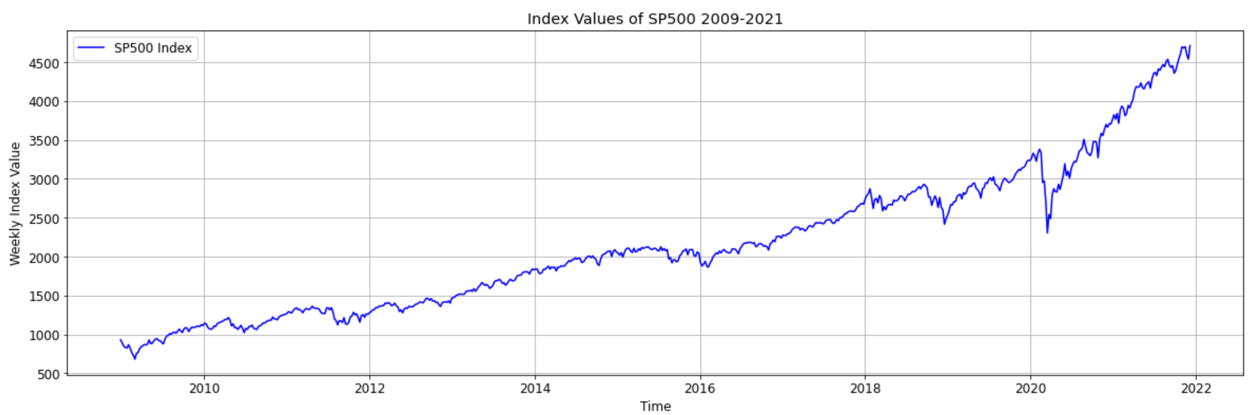


Figure 15 Index Values of SP500 2009-2021

When we checked the weekly return values of Indexes for our third time period, we observed that BIST100 index's variance had increased slightly while SP500 index's variance had declined in a small amount. While this divergence can be explained by the increasing social and political environment in Turkey specially after 2012, it might be a resourceful finding for investors.

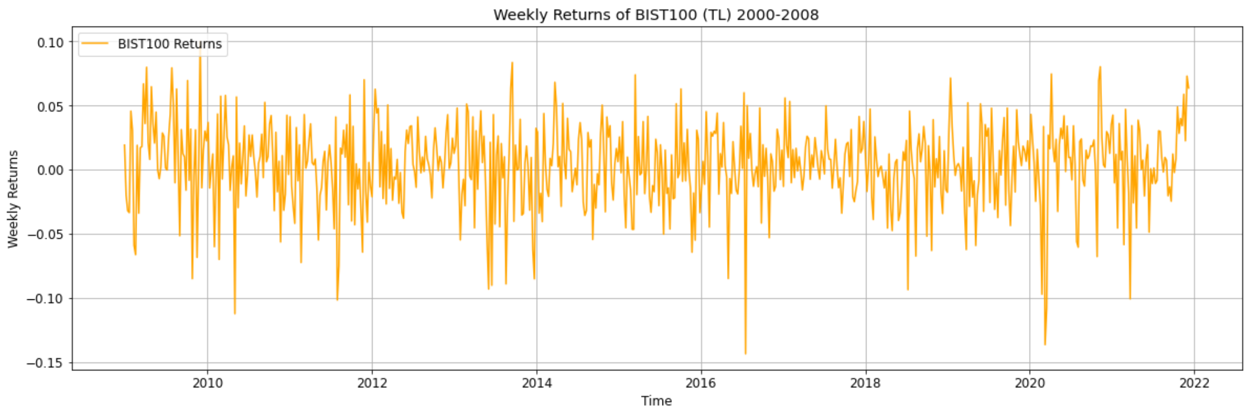


Figure 16 Weekly Returns of BIST100 (TL) 2000-2008

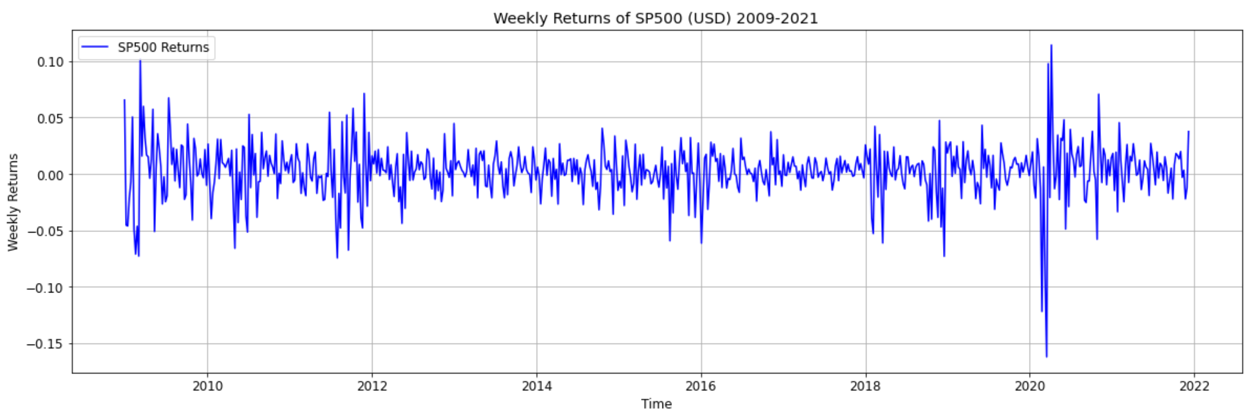


Figure 17 Weekly Returns of SP500 (USD) 2009-2021

To conclude with this section, we can claim that the returns seem like behaving in a stationary manner while the mean is around 0. Another important issue here is that the indexes had shown an increasing trend over last 30 years while the variance of returns raised with the index values as well.

Statistical Analysis of BIST100, BISTALL, Gold, BTC and SP500

In this section different statistics related to 5 different random variables (weekly returns of SP500, BIST100, BISTALL, Gold, BTC) are examined for time period 2015-2021.

Random Variable	mean	Variance	Standart Deviation	minimum	maximum	Interquartile Range	Skewness	kurtosis	First Order Autocorrelation
SP500 Weekly Return	0.22%	0.06%	2.35%	-16.23%	11.42%	0.022	-1.325	10.948	-0.096
BIST100 Weekly Return in TL	0.24%	0.10%	3.17%	-14.37%	8.00%	0.037	-0.875	2.481	-0.008
BISTALL Weekly Return in TL	0.27%	0.10%	3.11%	-15.36%	7.88%	0.036	-1.033	3.089	-0.003
Gold Weekly Return	0.11%	0.04%	1.97%	-8.57%	9.00%	0.024	-0.097	2.183	-0.004
BTC Weekly Return	1.39%	1.18%	10.88%	-54.39%	36.19%	0.105	-0.392	2.441	-0.013

When we observe the sample statistics of 5 different random variable, first thing to notice that BTC had a significantly higher weekly return rate though this return rate is followed with the higher risk that can be derived from its high variance, standard deviation and interquartile range. Generally, we can claim that the assets acted as we would expect that the higher returns come with higher dispersion and therefore higher risk. One interesting finding here is that BISTALL shows a higher return rate than BIST100 with less variance therefore we may say that BISTALL dominates BIST100 Index under the decision making with mean variance portfolio selection criteria.

When we checked skewness statistics, we see that all the random variables are negatively skewed implying that negative extreme events, great lost, are more likely to happen in all investment options. Although we notice that the indexes are more left skewed than Gold and BTC and this outcome also can be associated with the symmetry, way too empirical approach, of maximum and minimum values of each random variable. Again, another statistic that can be associated with skewness is the kurtosis statistics of assets where we see that SP500 showed the most centric distribution and BTC showed the least centric distribution.

Lastly, if were to analyze the first order auto correlation coefficient estimates of each time series, we would see that all coefficients are negative which represent the "retracement" behavior of investors in market.

	SP500 Weekly Return	BIST100 Weekly Return in FX	BISTALL Weekly Return in FX	Gold Weekly Return	BTC Weekly Return
SP500 Weekly Return	1.000000	0.436467	0.449521	0.142453	0.140780
BIST100 Weekly Return in FX	0.436467	1.000000	0.995505	0.176752	0.135227
BISTALL Weekly Return in FX	0.449521	0.995505	1.000000	0.185621	0.147998
Gold Weekly Return	0.142453	0.176752	0.185621	1.000000	0.067437
BTC Weekly Return	0.140780	0.135227	0.147998	0.067437	1.000000

Figure 18 Correlation Matrix

	SP500 Weekly Return	BIST100 Weekly Return in FX	BISTALL Weekly Return in FX	Gold Weekly Return	BTC Weekly Return
SP500 Weekly Return	0.000550	0.000325	0.000328	0.000066	0.000359
BIST100 Weekly Return in FX	0.000325	0.001005	0.000983	0.000110	0.000467
BISTALL Weekly Return in FX	0.000328	0.000983	0.000970	0.000114	0.000502
Gold Weekly Return	0.000066	0.000110	0.000114	0.000388	0.000144
BTC Weekly Return	0.000359	0.000467	0.000502	0.000144	0.011842

Figure 19 Covariance Matrix

When we examined the pair relationships between investment options, we see that there are no negatively correlated options, which is not an ideal situation. The highest correlation is associated between BIST100 and BISTALL as predicted but interestingly the least correlation is associated between Gold and BIST100.

Geometric Mean of SP500 Weekly Return = 0.0
Geometric Mean of BIST100 Weekly Return in TL= 0.0
Geometric Mean of BISTALL Weekly Return in TL= 0.0
Geometric Mean of Gold Weekly Return = 0.0
Geometric Mean of BTC Weekly Return = 0.0

Lastly when we checked the geometric means of time series, we noticed that the geometric means converge to 0. This situation might occur due to the fact that return values are between 1 and -1 and there are 363 of them for the time period therefore their multiplication converges to 0. Another possible reason of this result is that there might be 0 return weeks which directly makes the geometric mean 0.

Moving Sample Analysis of BIST100

To compute 52 week moving average of BIST100 time series a basic function "MA52_finder_w" was developed and used. The resulting moving average statistics are represented in the *figure-20*.

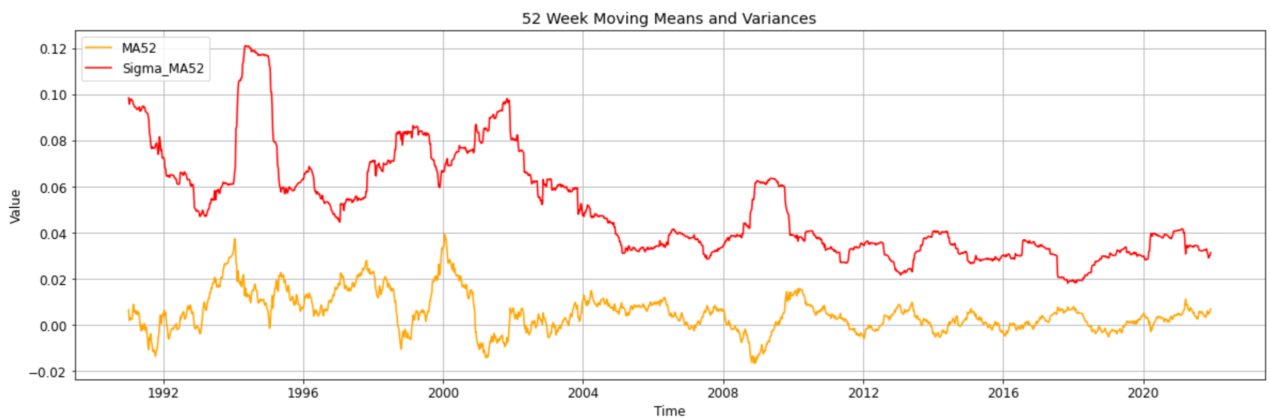


Figure 20 52-Week Moving Means and Variances of BIST100 for 1990-2021

What can be seen in the first sight is that moving sample variances peaks as the moving sample averages observe shocks as expected. Furthermore, an important finding is that we see generally, the MA52 of weekly return are likely to be over 0 most of the time.

We see that MA52 of means peaked in 2000 where it is logical since in question 1 we observed that there was a steep increase in the index value in the late 90's, but it is insightful to observe that even though there were even steeper increases in the index value in 2000's the Moving average could not yield that much indicating that the distribution of weekly returns in sample size 52 became more symmetric over time.

Empirical Distribution of Daily Returns for 2001-2021

To check the empirical distribution of the random variables I first plotted box graph and histogram to visually investigate the random nature of series. *(Figures 21-22-23-24)*

Also, daily returns had to be calculated and for the easiness of interacting with python's plotting libraries, numpy array was my choice to store daily return data.

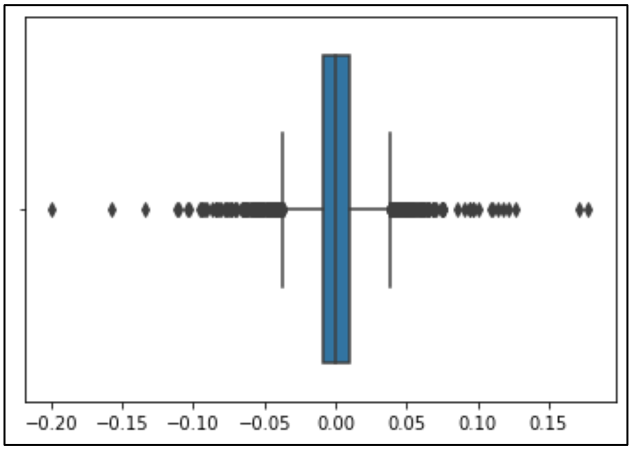


Figure 21 Box Plot of BIST100 Daily Returns for 2000-2021

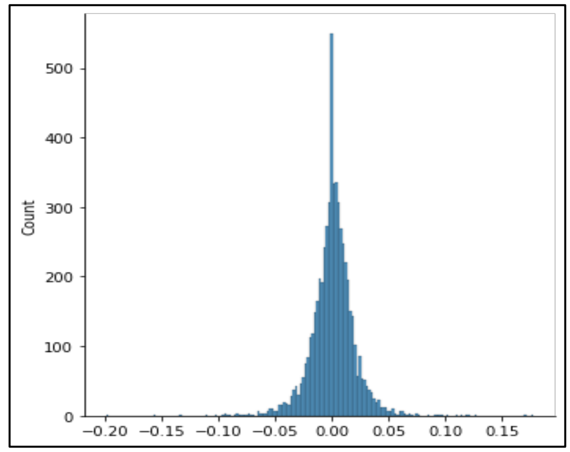


Figure 22 Histogram of Daily Returns for 2000-2021

Even though generally both plots have similar shapes of normal distribution, we notice that both plots indicate left skewness since left extreme points were occurred. Furthermore, when I applied Kolmogorov-Smirnov test to data I obtained a statistic of 0.469 and P-value of 0 indicating that the Daily returns were not normally distributed.

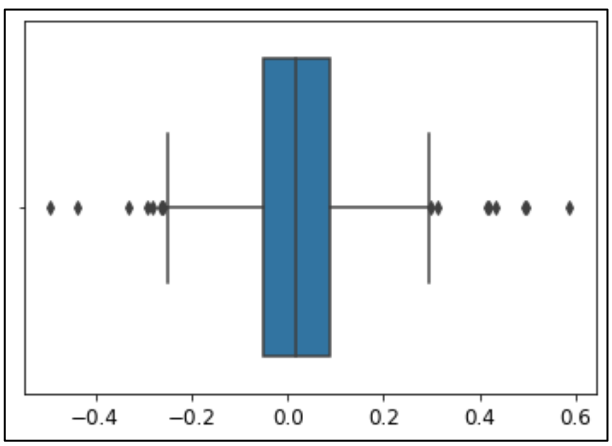


Figure 23 Box Plot of BIST100Monthly Returns for 2000-2021

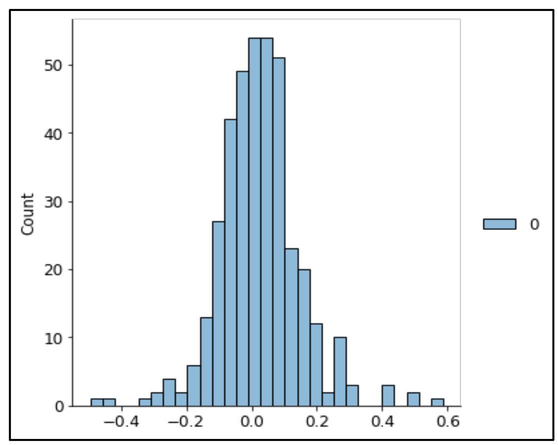


Figure 24 Histogram of Monthly Returns for 2000-2021

When I checked for monthly data, I observed that unlike Daily data monthly data looks like righter skewed though it can be easily stated that monthly return data is not likely have normal distribution as underlying probability distribution and Smirnov-Kolmogorov test supports this claim with statistic of 0.721 and P-value of 0.

Autocorrelation Tests

Up to 30 lags, auto correlation functions were computed for BIST100 and SP500 daily returns in this sub-section, results are as follows:

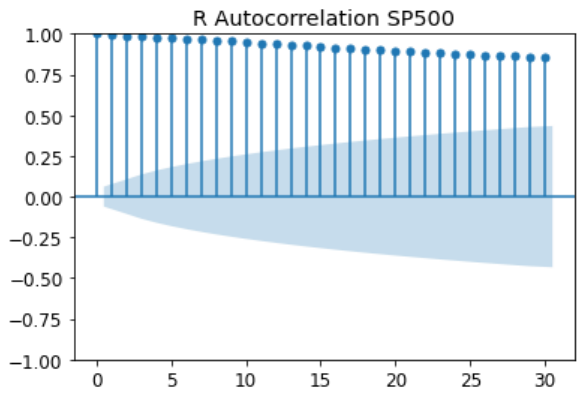


Figure 25 Auto correlation Function for BIST100 Daily Returns 2018-21

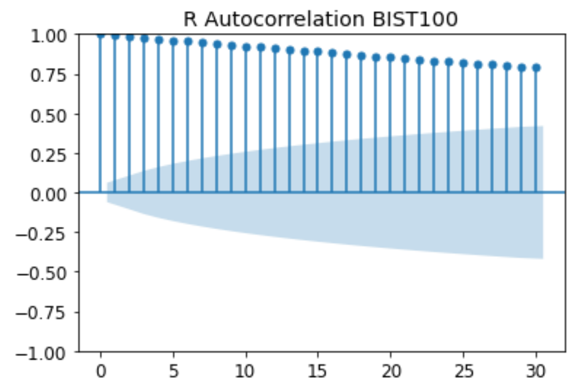


Figure 26 Auto correlation Function for SP500 Daily Returns 2018-21

The outputs of autocorrelation functions indicates that up to 30 days, the residuals are significantly dependent to lagged residuals.

Normality Test

When I analyzed the shock occurred on 17 December 2021, the probability computed by the normality assumption was 0.000585 while the empirical probability was 0.001829. This indicates that extreme negative returns, or losses, may be misjudged by the assumption of normality.

Value Calculations

Risk Premiums

To calculate the yearly risk premiums, I first aggregated the Turkish risk-free return rate by averaging each interest rate for each year. Then this aggregated data frame was merged to the yearly return data frame of BIST00 index. Risk premiums for each year were calculated by the output data frame and then displayed by a line graph which is given below:

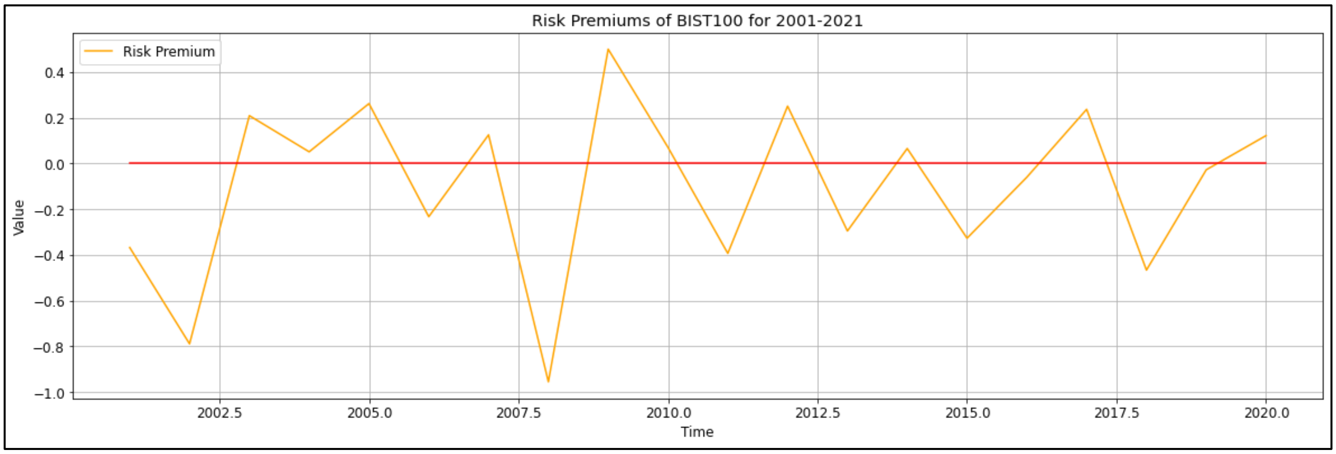


Figure 27 Risk Premiums of BIST100 for 2001-2021

Risk premium graph shows that the return rates of BIST100 had been generally failing to exceed risk-free rate of return in Turkey. The average risk premium was calculated as -0.1014 which. sadly. indicates that taking risk cannot provide effective extra income.

Portfolio Scenario

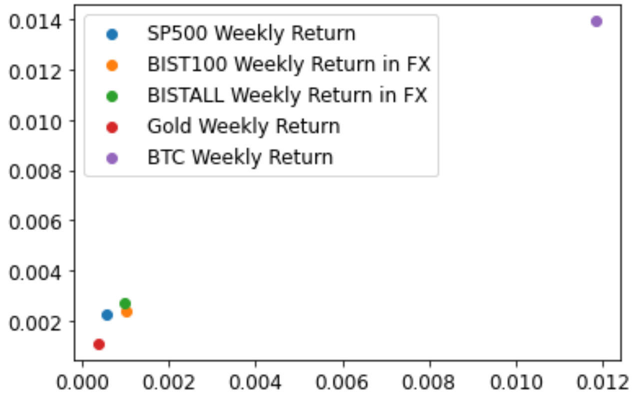


Figure 28 Mean-Variance Graph for 5 Assets 2015-2021

While selecting portfolio as mentioned in question 2, selecting BISTALL over BIST100 is more sensible so I will demonstrate a bundle consist of BISTALL, SP500, gold and BTC.

To estimate an investment plan in beginning of 2021, I checked last 4 years of weekly returns' mean and came up with following results:

SP500 Weekly Return	0.002245
BISTALL Weekly Return in FX	0.002925
Gold Weekly Return	0.002137
BTC Weekly Return	0.015902

So, by distributing $1000 equally to these 4 assets, I am expecting to have $ 1290.1052. On the other hand, my portfolio would have performed to reach only $1271.7067.

Conclusion

In this project I tried to explain some of the randomness of 5 different financial assets and try to understand some of the dynamics of the randomness lying behind their price action. Both visual and statistical studies conducted in this report complemented each other and enable both the writer and, hopefully, the reader to understand financial assets dynamics a bit deeper