

Financial Market Analytics Project

CERABOLINI AURORA, COMENSOLI PAOLO, TRITELLA MIRKO

839327, 883147, 887196

University of Milano Bicocca

MSc Data Science

Academic Year 2021/2022

July, 2022

All codes, notebook, and the dataset folder can be viewed or downloaded here: [GitHub](#)

[Repository](#)

I. INTRODUCTION

In this group project, we want to understand what are the structural characteristics that risk brings to real investment portfolios. For this reason, we need to build real portfolios that are focused on a specific level and type of risk. A portfolio is a collection of individual assets or securities and investors seek to diversify it instead of investing all of their wealth in a single or a few assets. With diversification, investors decrease the level of risk, in fact they want to maximize their return with less risk on their investment in a portfolio. To investigate the level of risk of individual stocks, we rely on the Single Index (beta) Model (SIM), proposed by William Sharpe in 1963, and then proceed to group the securities based on common characteristics in order to study them.

i. Single Index Model

Efficient portfolio construction with minimal risk for a given expected return can be achieved with the use of the Single Index Model proposed by Sharpe. The Single Index Model is based on the assumption that the fluctuations in the value of stock relative to other stocks do not depend on the characteristics of those securities alone. Relationships between securities occur only through their individual relationships with some indexes of business activity. Our index will be the Nasdaq-100. The equa-

tion of the Single Index Model links the returns of a single stock with the returns of the market index

$$r_i = \alpha_i + \beta_i(R_M) + \epsilon_i \quad (1)$$

In this linear equation we have:

- r_i expected return on security i
- R_M is the return of the market
- α is the intercept of the straight line or alpha coefficient: α is often referred to as "excess return" and it measures the performance of an investment against a market index or benchmark that is believed to represent the movement of the market as a whole.
- β_i is the slope of the straight line or beta coefficient: beta is a measure of the volatility, or systematic risk, of a security or portfolio relative to the market as a whole.
- ϵ_i is the error term with the mean of zero and a standard deviation which is a constant

The variance of the security has two components namely systematic risk, or market risk, and unsystematic risk, or specific risk. The variance explained by the index is referred to as systematic risk and the unexplained variance is called residual variance or unsystematic risk.

$$\text{SystematicRisk} = \beta_i^2 \sigma_M^2 \quad (2)$$

$$\text{UnsystematicRisk} = \text{TotalVariance} - \text{SystematicRisk} \quad (3)$$

ii. Benchmark Index

The Nasdaq-100 (NDX) is one of the world's leading indices by capitalization. It includes 100 of the largest national (US) and international non-financial companies listed on the Nasdaq stock market by market capitalization. Within this index we find some of the major technology companies such as Apple, Amazon, Alphabet (Google), Facebook and Microsoft. The Nasdaq-100 undergoes quarterly reviews that occur in March, June, September and December and become effective with the closing values on the third Friday of the month. To become part of this index, stocks must have a daily trading volume of at least 200,000 pieces, have a total average market capitalization equal to or greater than 0.1% of the average market capitalization of the Nasdaq-100 stocks and typically must have been listed for at least a couple of years - Figure [1].

iii. Dataset

At this time (July 2022) there are 102 stocks within the Nasdaq-100. The complete list can be viewed here: [Nadaq-100](#). For the construc-

tion of the dataset, we downloaded the daily closing prices of these securities and of all those securities that were part of the index starting from 2017 (beginning of our historical series). We then managed the revisions of the securities by finding the incoming securities, the outgoing securities, and the effective date of exchange on the official Nasdaq website. We used the Bloomberg terminal to download the data.

```
closing_prices = pd.read_csv(CLOSING_PRICE_CSV)
closing_prices.head()
```

Output in Figure [2]

When building the dataset we put attention to the survivorship bias. The survivorship bias indicates the tendency to consider only existing or "survivors" or "winners" stocks when measuring the performance of a portfolio. This is a bias linked to the "sample selection" in which observations that have already ceased to exist are not considered. For this reason, a portfolio's return may be overestimated, leading to overly optimistic conclusions. To avoid survivorship bias, we need to make sure that all data is present and that no longer present

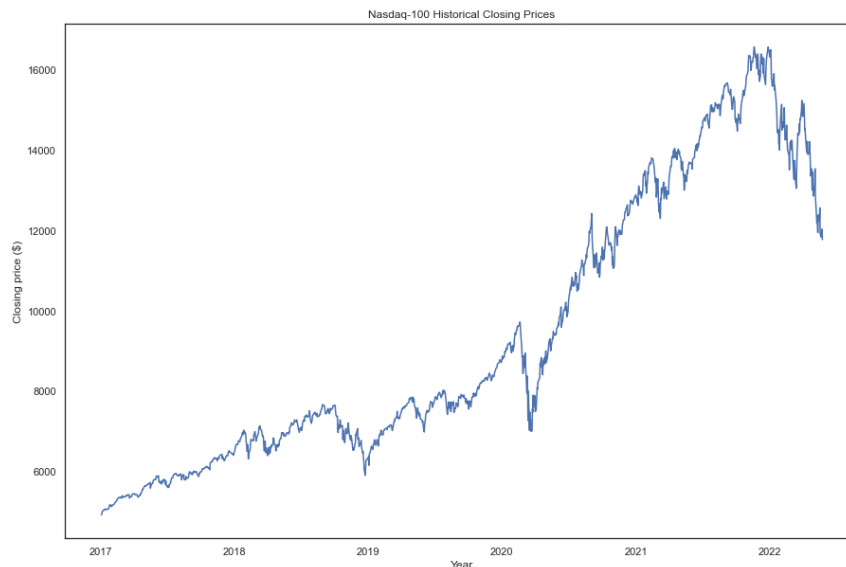


Figure 1: NASDA-100 Closing Prices

	Dates	NDX Index	AAPL	ABNB	ADBE	ADI	ADP	ADSK	AEP	ALGN	...	MXIM	CDW	CERN	CHKP	FOXA	FOX	INCY	TCOM	PTON	XLNX
0	3/1/2017	4911.33	29.038	NaN	103.48	72.505	103.50	76.18	NaN	NaN	...	38.62	NaN	49.40	84.57	NaN	NaN	102.320	40.49	NaN	59.07
1	4/1/2017	4937.21	29.005	NaN	104.14	72.360	103.66	77.52	NaN	NaN	...	39.03	NaN	47.93	84.92	NaN	NaN	102.840	41.34	NaN	58.64
2	5/1/2017	4964.95	29.153	NaN	105.91	71.320	103.04	76.93	NaN	NaN	...	38.44	NaN	47.65	86.66	NaN	NaN	104.810	42.80	NaN	57.93
3	6/1/2017	5007.08	29.478	NaN	108.30	71.600	103.11	79.30	NaN	NaN	...	39.74	NaN	47.55	87.24	NaN	NaN	108.305	42.50	NaN	59.05
4	9/1/2017	5024.90	29.748	NaN	108.57	71.940	102.47	79.59	NaN	NaN	...	40.65	NaN	47.99	87.20	NaN	NaN	118.525	43.24	NaN	59.06

5 rows × 155 columns

Figure 2: Closing Prices

observations have not been omitted. To create a dataset that takes into account the revisions that took place during the analysis period, we also downloaded the historical series of daily prices even for those stocks that are no longer part of the index, but that are still within the U.S. stock market. By checking the date on which a certain change took place, indicated by **X**, between a security **A** (incoming) and **B** (outgoing), we set to #N/A all the cells of **B** starting from **X**, valued the cells of **A** starting from **X** and setting to #N/A all cells of **A** prior to **X**. In these 5 years that we are analyzing, it can happen that a stock was removed from the Nasdaq-100. We also had to look up the prices of those stocks that are no longer listed. In this case we had to consider the sale price or what was obtained from a merge.

iv. Risk-free rate

In order to have somewhat more precise metrics, we decided to also consider the risk-free rate. The risk-free rate of return is the theoretical rate of return of an investment with zero risk. Since we are in the U.S. market we will use the U.S. 3-Month T-Bill as the risk-free rate. This is a useful proxy because the market considers there to be virtually no chance of the U.S. government defaulting on its obligations. For the data we actually used the Coupon Equivalent¹. The Coupon Equivalent, also called the Bond Equivalent, or the Investment Yield, is the bill's yield based on the purchase price, discount, and a 365- or 366-day year.

¹<https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics>

v. Pre-processing

Starting from the closing prices of each title in our dataset and of our benchmark index, we calculate log returns. The log return can be calculated by simply taking a natural logarithm of the current price minus the logarithm of closing price of the previous day. One of the advantages is that the logarithmic returns are symmetric, this implies that logarithmic returns of equal quantity but opposite signs will cancel each other out. This means that if an investment of \$100 has a simple return of 50% followed by a simple return of 50%, the result will be \$75, while an investment of \$100 that give a logarithmic return of 50% followed by a logarithmic return of 50% will come back to \$100. We use logarithmic returns to perform our regression which will allow us to build our portfolios.

```
log_returns = np.log(closing_prices) -
→ np.log(closing_prices.shift(1))
```

	Dates	NDX Index	AAPL	ABNB	ADBE	ADI	ADP	ADSK
1	4/1/2017	0.005256	-0.001137	nan	0.006358	-0.002002	0.001545	0.017437
2	5/1/2017	0.005603	0.005090	nan	0.016854	-0.014477	-0.005999	-0.007640
3	6/1/2017	0.008450	0.011086	nan	0.022315	0.003918	0.000679	0.030342
4	9/1/2017	0.003553	0.009118	nan	0.002490	0.004737	-0.006226	0.003650
5	10/1/2017	0.002042	0.001008	nan	-0.002859	0.005683	-0.002638	0.004888

II. BUILDING PORTFOLIOS

i. Ranking

As we said previously with the Single-Index Model we can do a simple linear regression and the return of a stock i is expressed by:

$$r_i = \alpha_i + \beta_i(R_M) + \epsilon_i \quad (4)$$

By breaking down all the returns of the stocks under consideration into these components, and by using their properties, we can construct different rankings based on them or combinations of them. Through these rankings we can build our portfolios. To keep our tilted portfolios up to date, we will need to reevaluate these rankings weekly. In our study we will also try to see how these components can explain the returns of our portfolios. The following are the selectors used to build the portfolios:

- **Highest level of R^2 .** R-squared measures how closely each change in the price of an asset is correlated to a benchmark
- **Lowest level of R^2 and then high level of σ_i^2 (total risk)**
- **Stock with positive α_i and eventually significant**
- **Stock with positive α_i and also high β_i**
- **Highest and lowest level of $\beta_i^2 \sigma_M^2$ (systematic risk)**
- **Stock with the highest absolute return r_i ,** where absolute return is the return that an asset achieves over a specific period (in our case 180 days)
- **Excess returns over total returns, $\frac{\alpha_i}{r_i}$**

The selection of securities that are part of our portfolios is done by ranking. At the beginning of our procedure we calculate the first ranking on the previous 180 days. Rankings are obtained by going to sort the data according to specific selectors, such as highest r-square values or best absolute returns. We then calculate the current week's returns and re-run the previous procedure on the past 180 days by moving the window from week to week. We continue in this way until we get to the end of our data. To avoid look ahead bias and survivorship bias, the stocks selected by the ranking procedure are based solely on past data relative to the week in which we will need to calculate returns. Even if a stock in that week were to drop out of the index or have completely different performance than in the past this does not matter. Within each ranking procedure what we do are the following steps:

1. We load the 180-day window of stock returns, and only the stocks that are present for all 180 days
2. For each security we perform a linear regression against the index (also with data referring to the same dates)
3. From the model we extract the values of the various parameters
4. Based on the selected selector, the best 10 titles will be chosen

Having obtained the list of securities we will then move on to the calculation of portfolio returns. For example to get the r-squared value:

```
model = sm.OLS(title_returns, ndx_returns)
result = model.fit()

r2 = result.rsquared
```

Regarding linear regression in our study, we used statsmodels^[2]. In addition to the list of titles, the ranking procedure also returns the table containing the values of all parameters. For this reason let's now go into a little more detail and look at the values that are returned by the model with regard to some selectors.

Portfolio based on stocks with the highest level of R-squared

We look at the first iteration of the selector with highest level of R^2 . The higher the R-squared number, the more correlated the asset is to its benchmark. A hypothetical portfolio with an R-squared of 0 has no correlation to its benchmark at all, while a portfolio with an R-squared of 1 matches the performance of its benchmark precisely. Looking at Figure [3] the first thing that jumps out is the fact that in this iteration, but which will remain true in subsequent iterations, is that within the portfolio we will find major American technology companies. This is actually not all that surprising, as they are giants from the standpoint of value and capitalization, and

^[2]<https://www.statsmodels.org/stable/index.html>

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
930	5/5/2020	MSFT	0.938464	0.000049	1.156625	0.000448	0.397175	0.241141	0.334307	0.334356
940	14/5/2020	MSFT	0.936981	0.000051	1.157416	0.000551	0.305699	0.291043	0.334667	0.334718
950	26/5/2020	MSFT	0.935464	0.000052	1.156903	0.000464	0.392364	0.282964	0.334343	0.334395
920	24/4/2020	MSFT	0.934167	0.000052	1.152983	0.000627	0.248603	0.206161	0.332415	0.332467
980	24/6/2020	MSFT	0.934000	0.000055	1.152711	0.000286	0.606374	0.297998	0.331784	0.331838
...

Figure 3: R-squared selector first iteration result

consequently it is plausible that they are the ones driving the economy and the underlying index. In this iteration we obtained stocks with r-square between 0.5 and 0.6. Good values but not outstanding. Over the whole series of this portfolio, the average R-square value is 0.68. In general, almost all stocks have a beta above one, and this means that they tend to be more aggressive than defensive stocks. The specific risk is very low and the alpha is also practically zero. The results proved to be as expected, and this portfolio performed slightly better than the index. As a final note the highest level of R-squared was obtained by Microsoft (MSFT) during 2020 with a maximum value of 0.93 - Figure [4].

Portfolio based on stocks with best absolute return

Look again at the first iteration of our procedure. Figure [5] shows the data of the first ten selected stocks. We can still see a high beta, a very small and nonsignificant alpha and a tendentially low r-square. What immediately jumps out at you, predictably, is the absolute return column. The first stock in the first six months of analysis achieved an absolute return of about 71%. However, as we will see later, this portfolio did not perform exceptionally

well, and underperformed the benchmark. The returns for the first week of this portfolio are:

	Date	Log Return
181	21/9/2017	-0.013666
182	22/9/2017	-0.003156
183	25/9/2017	-0.025036
184	26/9/2017	-0.006764
185	27/9/2017	0.016818
186	28/9/2017	0.003350
187	29/9/2017	0.011621

This is a good example of the fact that if a stock has performed well in the past, it does not necessarily perform well in the future. In fact, if we look at the returns of individual stocks in the first week we find that they were coming off a bearish period - Figure [6].

Portfolio based on stock with positive alpha

The excess return of a title relative to the return of a benchmark index is the title's alpha. In other words it is the part of return of a stock due to firm-specific factors and therefore untied from macroeconomic events that affect the entire market. We know from theory that the CAPM specifies that alphas must be zero and that deviations from zero

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
930	5/5/2020	MSFT	0.938464	0.000049	1.156625	0.000448	0.397175	0.241141	0.334307	0.334356
940	14/5/2020	MSFT	0.936981	0.000051	1.157416	0.000551	0.305699	0.291043	0.334667	0.334718
950	26/5/2020	MSFT	0.935464	0.000052	1.156903	0.000464	0.392364	0.282964	0.334343	0.334395
920	24/4/2020	MSFT	0.934167	0.000052	1.152983	0.000627	0.248603	0.206161	0.332415	0.332467
980	24/6/2020	MSFT	0.934000	0.000055	1.152711	0.000286	0.606374	0.297998	0.331784	0.331838
...

Figure 4: Microsoft (NDX:MSFT) r-squared

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
0	2/10/2017	VRTX	0.153709	0.000553	1.540778	0.002283	0.203283	0.712586	0.592259	0.592812
1	2/10/2017	NVDA	0.241062	0.000547	2.026196	0.001128	0.526473	0.599815	1.024223	1.024770
2	2/10/2017	ATVI	0.242589	0.000272	1.434711	0.001624	0.196800	0.573247	0.513524	0.513796
3	2/10/2017	TSLA	0.145605	0.000441	1.331995	0.001574	0.325138	0.544164	0.442626	0.443067
4	2/10/2017	JD	0.333218	0.000223	1.621815	0.001252	0.271072	0.542991	0.656197	0.656420
5	2/10/2017	ISRG	0.200560	0.000096	0.755998	0.001893	0.012045	0.488818	0.142585	0.142681
6	2/10/2017	LRCX	0.504303	0.000124	1.723359	0.000782	0.356067	0.478151	0.740940	0.741063
7	2/10/2017	PYPL	0.346526	0.000095	1.092873	0.001452	0.052022	0.475269	0.297969	0.298064
8	2/10/2017	MU	0.360894	0.000272	1.906027	0.000521	0.678468	0.466950	0.906337	0.906609
9	2/10/2017	ILMN	0.102850	0.000358	0.984149	0.001360	0.345376	0.437460	0.241631	0.241989

Figure 5: Absolute return selector first iteration result

	VRTX	NVDA	ATVI	TSLA	JD	ISRG	LRCX	PYPL	MU	ILMN
181	-0.004331	-0.027716	-0.013787	-0.020071	-0.033405	-0.015560	-0.004286	-0.001701	-0.001669	-0.014133
182	-0.005275	-0.009784	-0.003281	-0.042901	-0.014294	0.001266	0.025502	0.006939	0.004446	0.005826
183	-0.014383	-0.045722	-0.027929	-0.017527	-0.034577	-0.013837	-0.037825	-0.024577	-0.033835	-0.000149
184	-0.004302	0.005598	-0.006134	0.000753	-0.015016	-0.000690	-0.004299	-0.008858	-0.019986	-0.014707
185	0.004100	0.021687	0.021306	-0.012474	-0.015245	0.007295	0.057063	0.006177	0.081707	-0.003433
186	0.003683	-0.000285	0.007107	-0.004026	-0.027056	0.001756	0.013455	0.009430	0.023186	0.006251
187	0.016045	0.017436	0.015151	0.004407	-0.011711	0.019306	0.017555	0.001563	0.035455	0.001005

Figure 6: First week of return (from day 181 to 187)

are the result of temporary disequilibria. In practice, however, assets may have non-zero alphas. For that reason alpha is also called “abnormal rate of return”, due to the fact that since markets are efficient, there is no way to systematically earn returns that exceed the broad market as a whole. Fund managers use both alpha and beta to evaluate an investment portfolio’s returns, in fact, positive value of alpha imply higher average excess return without the “cost” of any additional exposure to market risk. A value of 1.0 means that the title outperformed its benchmark index by 1% and so the investment has a return in excess of the reward for the assumed risk. A negative number means that it has under-performed, or that it was too risky for the return, and if the alpha is zero, its returns match the benchmark, and so the return was adequate for the risk taken. As you can see from Figure [7], in our execution of the model we could not find stocks with a large alpha value. The r-square value tended to be lower than in other portfolios. In this portfolio the average value of alpha is 0.0018. Adding the condition of selecting positive and significant alpha to the

ranking procedure, however, returns an error because there are not 10 stocks with significant alpha in each period. Going to look within the dataset of this portfolio we find that the range of beta values exceeds 2 in some cases, while in others it is negative (max negative -0.21) and with very low risks, both specific and systematic.

Stocks with the highest systematic risk and highest Idiosyncratic Risk

Systematic Risk refers to the risk inherent to the entire market or market segment while *Unsystematic Risk*, also known as specific risk or idiosyncratic risk, is the risk that is unique to a specific company or industry. Here we try to compare the results of the two portfolios containing the securities with higher systematic risk and higher specific risk - Figure [8] and Figure [9]. The specific risk values are not too high so it is not possible at first glance to see substantial differences. Despite this, the two portfolios performed quite differently. The portfolio with the highest level of systemic risk had an annualized return of 8.63%, while the portfolio with the highest level of specific

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
0	2/10/2017	VRTX	0.153709	0.000553	1.540778	0.002283	0.203283	0.712586	0.592259	0.592812
1	2/10/2017	ISRG	0.200560	0.000096	0.755998	0.001893	0.012045	0.488818	0.142585	0.142681
2	2/10/2017	ATVI	0.242589	0.000272	1.434711	0.001624	0.196800	0.573247	0.513524	0.513796
3	2/10/2017	TSLA	0.145605	0.000441	1.331995	0.001574	0.325138	0.544164	0.442626	0.443067
4	2/10/2017	PYPL	0.346526	0.000095	1.092873	0.001452	0.052022	0.475269	0.297969	0.298064
5	2/10/2017	ILMN	0.102850	0.000358	0.984149	0.001360	0.345376	0.437460	0.241631	0.241989
6	2/10/2017	JD	0.333218	0.000223	1.621815	0.001252	0.271072	0.542991	0.656197	0.656420
7	2/10/2017	NVDA	0.241062	0.000547	2.026196	0.001128	0.526473	0.599815	1.024223	1.024770
8	2/10/2017	EA	0.259101	0.000159	1.144536	0.001081	0.260080	0.418724	0.326807	0.326965
9	2/10/2017	CERN	0.153864	0.000137	0.766338	0.001079	0.226305	0.344250	0.146512	0.146648

Figure 7: Positive alpha selector first iteration result

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
0	2/10/2017	NVDA	0.241062	0.000547	2.026196	0.001128	0.526473	0.599815	1.024223	1.024770
1	2/10/2017	MU	0.360894	0.000272	1.906027	0.000521	0.678468	0.466950	0.906337	0.906609
2	2/10/2017	AMAT	0.587449	0.000097	1.801867	0.000219	0.769658	0.392228	0.809985	0.810081
3	2/10/2017	LRCX	0.504303	0.000124	1.723359	0.000782	0.356067	0.478151	0.740940	0.741063
4	2/10/2017	AVGO	0.506455	0.000113	1.655574	-0.000084	0.917055	0.308999	0.683799	0.683912
5	2/10/2017	NFLX	0.369296	0.000195	1.643572	0.000296	0.780910	0.375071	0.673921	0.674116
6	2/10/2017	JD	0.333218	0.000223	1.621815	0.001252	0.271072	0.542991	0.656197	0.656420
7	2/10/2017	SWKS	0.352364	0.000189	1.557655	0.000084	0.935696	0.320182	0.605305	0.605494
8	2/10/2017	VRTX	0.153709	0.000553	1.540778	0.002283	0.203283	0.712586	0.592259	0.592812
9	2/10/2017	WDC	0.282852	0.000254	1.536931	-0.000459	0.704915	0.218321	0.589306	0.589559

Figure 8: Highest systematic risk selector first iteration result

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
0	2/10/2017	STX	0.127503	0.000639	1.484874	-0.002736	0.155994	-0.201793	0.550061	0.550700
1	2/10/2017	MAT	0.019970	0.000591	0.533200	-0.004138	0.026290	-0.640485	0.070927	0.071518
2	2/10/2017	INCY	0.136559	0.000565	1.452992	-0.000938	0.604323	0.115696	0.526694	0.527259
3	2/10/2017	VRTX	0.153709	0.000553	1.540778	0.002283	0.203283	0.712586	0.592259	0.592812
4	2/10/2017	NVDA	0.241062	0.000547	2.026196	0.001128	0.526473	0.599815	1.024223	1.024770
5	2/10/2017	AKAM	0.063874	0.000482	0.881772	-0.002897	0.084327	-0.348862	0.193974	0.194457
6	2/10/2017	ORLY	0.001172	0.000458	-0.112696	-0.001673	0.305018	-0.323228	0.003168	0.003627
7	2/10/2017	TSLA	0.145605	0.000441	1.331995	0.001574	0.325138	0.544164	0.442626	0.443067
8	2/10/2017	NTES	0.148982	0.000425	1.325811	0.000014	0.992670	0.262195	0.438525	0.438950
9	2/10/2017	CSX	0.091696	0.000412	0.991262	0.001067	0.489842	0.386243	0.245137	0.245549

Figure 9: Highest idiosyncratic risk selector first iteration result

risk had a negative performance of -0.79%. In terms of volatility, on the other hand, the difference is not so marked, the former 38% while the latter 35%, and thus less for the highest specific risk. In general, going by the data in the various iterations, ranking based on maximum specific risk often has negative absolute returns. However, the r-square and total risk are higher in the portfolio with higher systemic risk.

Stocks with the lowest total risk

With this portfolio, we want to minimize the total risk of the components. The total risk is

expressed as:

$$TotalRisk = SystematicRisk + SpecificRisk \quad (5)$$

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{ei}^2 \quad (6)$$

If we continue looking at the first iteration, the results are as shown in Figure [10]. Total risk, as well as systematic and specific risk, of course, are very low. Interestingly, the values of r-square are very low and beta values are also very low, less than one. In fact, the stocks with the lowest risks are those that are defensive. Regarding alphas in this case we find several negative alphas, but still close to zero. Looking at the selected stocks, it is possible to see

	Dates	Title	r2	specific_risk	beta	alpha	alpha_significance	absolute_returns	systematic_risk	total_risk
0	2/10/2017	TSCO	0.001484	0.000277	0.098691	-0.001357	0.285068	-0.224957	0.002430	0.002707
1	2/10/2017	ORLY	0.001172	0.000458	-0.112696	-0.001673	0.305018	-0.323228	0.003168	0.003627
2	2/10/2017	COST	0.020284	0.000143	0.264130	-0.000180	0.843156	0.019345	0.017405	0.017547
3	2/10/2017	ROST	0.015743	0.000188	0.266445	-0.000702	0.501338	-0.074167	0.017711	0.017899
4	2/10/2017	DLTR	0.018114	0.000205	0.299133	0.000087	0.936316	0.074256	0.022323	0.022529
5	2/10/2017	VOD	0.043517	0.000090	0.311569	0.000468	0.518167	0.145182	0.024218	0.024308
6	2/10/2017	MNST	0.025836	0.000188	0.343631	0.000770	0.461535	0.205867	0.029459	0.029647
7	2/10/2017	PAYX	0.072053	0.000083	0.390339	-0.000662	0.340774	-0.042659	0.038012	0.038095
8	2/10/2017	ESRX	0.035739	0.000182	0.399220	-0.001077	0.295046	-0.115660	0.039761	0.039943
9	2/10/2017	DISCA	0.031266	0.000227	0.415813	-0.001974	0.086250	-0.273948	0.043135	0.043362

Figure 10: Minimum total risk selector first iteration result

that the tech giants are not present, but instead there are retail companies, like TSCO (agriculture and garden products), ROST, mass media like Discovery (DISCA) or food and beverage companies like MNST.

ii. How we calculate the portfolios returns

At the end of each weekly iteration of the portfolio construction procedure we obtain a new list containing the ten titles based on some specific ranking. Before making the turnover of the assets inside our portfolios we need to calculate the portfolio return of the week just past. We look at the log returns dataset and we extract the seven rows of the week of the ten titles found within our portfolio. Since we decide to use a 1/N weight scheme we associate the same weight with each return of each individual stock. The daily return of the portfolio will be given by the sum of the returns of individual securities multiplied by their weight. In the first part of this report we talked about the *survivorship bias* and so during the weekly calculation two exceptions may occur:

- The stock was removed from the Nasdaq-100 but is still within the U.S. stock market
- The stock was removed both from the Nasdaq-100 and from the U.S. stock market, for example, because of an acquisition by another company

In the first case we have to go and get the closing price of the missing day (closing_prices_removed_titles dataset),

calculate the return and use that return for the particular day. The second case is more tricky. Based on our rankings, all the stocks that were part of the second exception were acquired or merged with another company and so we had to look up the price that the various shareholders received for each share. We searched for such information and loaded it into the dictionary `title.out_of_the_market`. In this case once the return is calculated, the title cannot be used on the remaining days for that week. This means that theoretically we need to decrease the number of stocks in the few remaining days and re-balance the weights within the portfolio, decreasing N by -1.

iii. Analyzing Returns and Volatility

For each portfolio we calculated the annualized return and annualized volatility. For returns, we multiplied the average by 252, i.e. the average of trading days in a year of the NASDAQ and NYSE. On the other hand for the annualized volatility, we multiplied the standard deviation of returns by the square root of 252. Based on our data, starting from the beginning of 2017 to the first part of 2022, governed by strong uncertainty³ the annualized returns of the NDX is equal to 15.01% with a volatility of 25.22%. As we can see from Figure [11], the only two portfolios that did slightly better than the index are max_r2 and

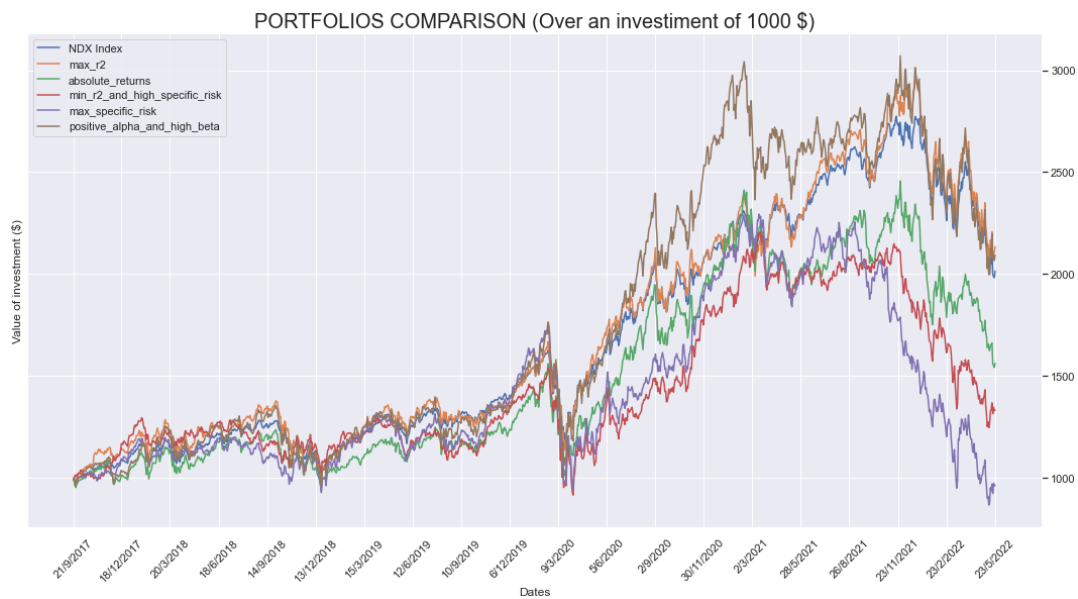
³<https://www.aa.com.tr/en/economy/nasdaq-sees-first-7-week-losing-streak-in-14-years/2594254>

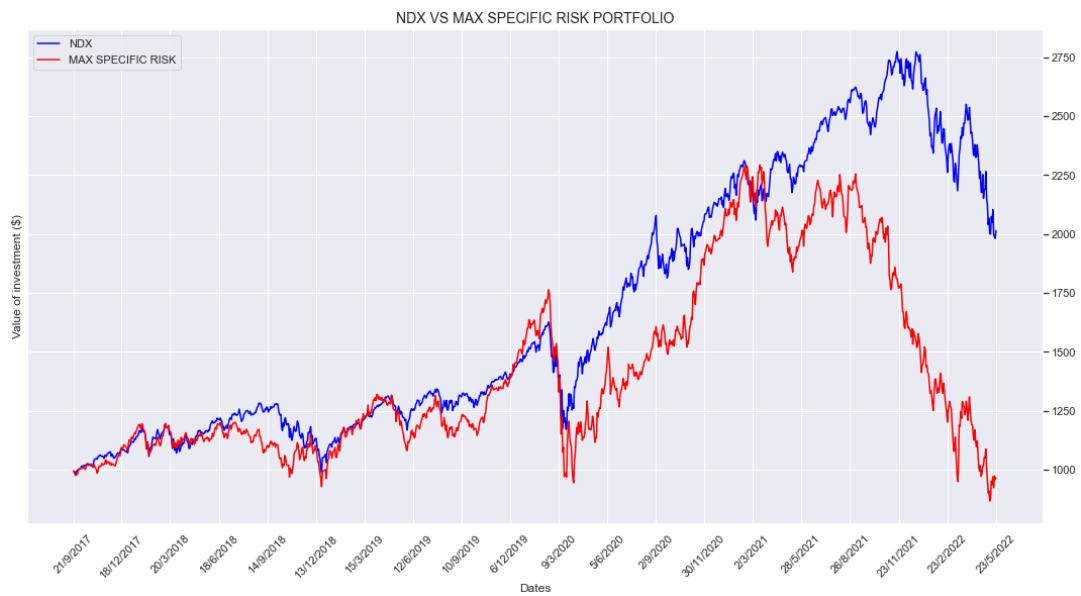
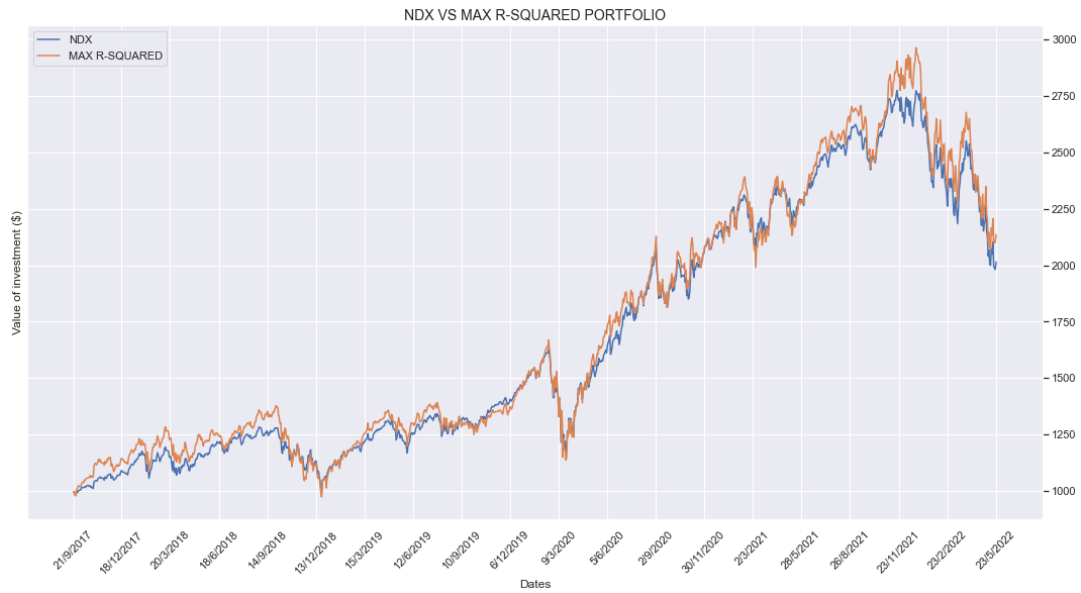
	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_r2	16.26%	30.05%
1	absolute_returns	9.57%	30.53%
2	min_r2_and_high_specific_risk	6.14%	29.2%
3	max_specific_risk	-0.79%	35.25%
4	high_systematic_risk	8.63%	38.46%
5	low_systematic_risk	13.18%	19.25%
6	positive_alpha	5.15%	28.91%
7	positive_alpha_and_high_beta	15.83%	36.2%
8	max_total_risk	8.5%	38.48%
9	min_total_risk	12.95%	19.23%
10	excess_return_over_total	10.21%	25.65%
11	Nasdaq-100	15.01%	25.22%

Figure 11: Annualized return and volatility

positive_alpha_and_high_beta. Both, however, have higher volatility. The portfolio related to titles with the highest specific risk not only under-performed the index but also achieved a negative return of -0.79% with a great level of volatility. The portfolio with the greatest volatility, unsurprisingly, is the one

that collects the securities exposed to the greatest total risk. A volatility of 38.48% compared with a return of 8.5%. In the opposite direction, the portfolio containing the securities with the lowest total risk obtains the lowest volatility over all. To get a clearer picture let us assume that we have invested \$1,000. We can then visualize the price chart:





iv. NDX vs Portfolio with Highest Specific Risk

As we can see from the previous figure, this portfolio has been highly influenced by the events of the last period entailing a giant loss. Among the main causes we can find pandemic, inflation, energy and war. If we try to increase the number of securities within the portfolio, for example from 10 to 20 and then 30 what we see is a reduction in volatility, but still negative returns.

	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_specific_risk	-1.34%	30.54%

	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_specific_risk	-1.32%	28.2%

Of course, by reducing the number of securities, we also reduce diversification and thus achieve even worse performance.

	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_specific_risk	-2.61%	41.91%

III. COMPARISON WITH METRICS

We calculate final statistics to compare the different tilted portfolios. The metrics we take into account are:

- **Sharpe Ratio:** calculates an investment's average returns compared to its potential risks. It is calculated by subtracting the risk-free rate from the expected rate of return of the portfolio, and dividing the result by the standard deviation, otherwise known as the statistical measure of the asset's volatility. The Sharpe ratio helps you determine whether the risk you've taken on has paid off in your returns, compared to the returns you might have seen without taking on risk. A high Sharpe ratio means the risk is paying off in the form of above-average returns. However, a Sharpe ratio greater than zero is typically considered good. In the formula, as a risk-free rate we used the US 3 month T bills: they

are considered nearly free of default risk because they are fully backed by the U.S. government. On 14/07/2022, the sharpe ratio for nasdaq-100 was equal to -0.77. A negative Sharpe ratio means that the risk-free rate is higher than the portfolio's return. This value does not convey any meaningful information.

- **Maximum Drawdown:** expresses in percentage terms, and for a fixed period of time, the maximum loss of value that an investor can experience. In the history of an instrument or a portfolio, it is the distance between the absolute peak and the absolute minimum. A low maximum drawdown is preferred as this indicates that losses from investment were small. If an investment never lost a penny, the maximum drawdown would be zero. The worst possible maximum drawdown would be -100%, meaning the investment is completely worthless.
- **Calmar Ratio:** is a formula that measures the performance of an investment fund compared to its risk. It is a function of the expected annual rate of return and the maximum drawdown over the previous years. It is used to assess the success of various investment funds and make investment decisions. A high ratio suggests that the return of the investment was not at risk of significant drawdowns. On the other hand, a low ratio indicates that the risk of drawdown is greater.
- **Value at Risk:** it is a statistic that quantifies the extent of possible financial losses within a firm, portfolio, or position over a specific time frame. Risk managers use VaR to measure and control the level of risk exposure. One can apply VaR calculations to specific positions or whole portfolios or use them to measure firm-wide risk exposure. One measures VaR by assessing the amount of potential loss, the probability of occurrence for the amount of loss, and the timeframe.
- **Information Ratio:** provides the amount of excess return of the portfolio with re-

spect to the reference benchmark for each unit of relative risk (represented by the tracking error) and allows to evaluate the manager's ability to outperform the benchmark in relation to the risk assumed (represented by the deviation from the benchmark). If the Information Ratio is positive, it means that the difference between profits and losses of these operations is positive, if the Information Ratio is negative, it means that the difference between profits and losses of the operations is negative.

- **Modigliani Ratio:** measures the returns of the portfolio, adjusted for the risk of the portfolio relative to that of some benchmark. To calculate the M2 ratio, we first calculate the Sharpe ratio and then multiply it by the annualized standard deviation of a chosen benchmark. We then add the risk-free rate to the derived value to give M2 ratio.

i. The results

The table in Figure [12] shows the values of the previously described metrics, calculated for each portfolio. Looking at the Sharpe Ratio values, the portfolio that obtains the highest return per unit of risk is the one obtained by taking into consideration the 10 titles with the lowest Systematic Risk. This portfolio has a Sharpe Ratio value of 0.63, and is the highest value among the other portfolios. The other portfolio that has a high Sharpe Ratio value is the portfolio obtained considering the lowest

values of total risk (SR = 0.619).

Portfolio that has a Sharpe value close to zero (SR = 0.051, negative but really close to zero) is the one obtained considering the highest values of the specific risk: in these portfolios the investment's excess return is zero.

Considering the Maximum Drawdown values, the portfolio that has the maximum observed loss from a peak to a trough, before a new peak is attained, is the one built with the highest specific risk values. The MDD value of this portfolio is -0.622, this means that the maximum loss experienced by this portfolio was 62%.

The portfolios that have the lowest value of the MDD are those obtained with the minimum values of the systematic risk and the lowest values of total risk: both portfolios have a value of MDD equal to -0.245.

The highest Calmar ratio value is obtained from the portfolio built taking into account the minimum values of the systematic risk and the value is equal to 0.537. This value is considered good and this means that the portfolio is not at risk of significant drawdown. The portfolio that is most at risk of drawdown is the one obtained with the positive alpha coefficient values: the Calmar Ratio value is 0.136.

The portfolio obtained with the highest values of specific risk has a negative Calmar ratio value but values below zero do not convey any meaningful information. Observing the values of the Value at Risk calculated with a confidence level of 90%, the portfolio that has the

	Portfolio Title	Sharpe Ratio	MDD	CL	Var 90	Var 95	Var 99	IR	M2
0	max_r2	0.50644	-0.319009	0.509824	-0.0236161	-0.0304939	-0.0433954	0.147124	0.13814
1	absolute_returns	0.279216	-0.37224	0.257058	-0.0242679	-0.0312552	-0.0443621	-0.374972	0.0808451
2	min_r2_and_high_specific_risk	0.174381	-0.434591	0.141185	-0.023329	-0.0300114	-0.0425466	-0.425863	0.0544106
3	max_specific_risk	-0.0519942	-0.622202	-0.0126751	-0.0284863	-0.0365529	-0.0516845	-0.660405	-0.00267051
4	high_systematic_risk	0.197286	-0.433131	0.199285	-0.0307064	-0.0395083	-0.0560193	-0.325048	0.0601861
5	low_systematic_risk	0.630407	-0.245205	0.537489	-0.0150178	-0.0194234	-0.0276875	-0.0987685	0.169399
6	positive_alpha	0.142176	-0.378936	0.136007	-0.0231317	-0.0297472	-0.0421568	-0.735389	0.04629
7	positive_alpha_and_high_beta	0.408441	-0.398539	0.397201	-0.028597	-0.036882	-0.0524232	0.0519947	0.113429
8	max_total_risk	0.193723	-0.433131	0.196196	-0.0307253	-0.039531	-0.0560492	-0.331627	0.0592876
9	min_total_risk	0.619095	-0.245205	0.528119	-0.0150112	-0.0194124	-0.0276682	-0.111512	0.166546
10	excess_return_over_total	0.357313	-0.357034	0.285965	-0.0203041	-0.0261748	-0.0371875	-0.4256	0.100537

Figure 12: Advanced Metrics

highest value of this metric compared to other portfolios is the one obtained with the highest values of the total risk: this portfolio has the risk of losing more than 3% in only 10. If we observe the values of the Value at Risk calculated with a confidence level of 95% and with a confidence level of 99%, we can see that the portfolio that has the risk of having a greater loss than the other portfolios is always the one obtained with the highest values of the total risk.

Looking at the values of the Information Ratio, only the portfolio obtained considering the highest values of R-squared and the portfolio with highest alpha and beta coefficients have obtained a positive value, and this means that investment management is active and efficient. In fact, the higher a value of the IR, the greater the manager's ability to achieve higher returns than the Benchmark Index. In this case, portfolio with highest values of R-squared has a higher IR value than portfolio with highest alpha and beta coefficients.

Considering the Modigliani Ratio, the portfolio that has the highest value ($M2 = 0.169$) is the one built considering the lowest values of the systematic risk. Compared to other portfolios, with the amount of risk taken, this portfolio is rewarding the investor more, in relation to the Benchmark Index and the risk-free rate of return.

IV. NON COMPULSORY TASK

i. Momentum Strategy

Momentum strategy is a system of buying stocks or other securities that have had high returns over the past months, and selling those that have had poor returns over the same period. In our project, to build the portfolios we consider the titles with a positive momentum value. To implement this strategy, we consider the sample of 180 days (6 month momentum) and calculate the momentum of the stocks present in this window, we then select 1/3 of the stocks with the highest positive momentum value. After that, for each of the titles

we perform the rolling regression and build the portfolios based on the previously described selectors, in the same way we did before without the momentum calculation. As we did previously, we re-balance the portfolio weekly and, each time we move the window forward seven-days, we calculate the momentum of the new stocks inside the window and take 1/3 of these stocks with the highest positive momentum value. After selecting the stocks with the highest momentum value, we identify a new stock ranking based on selectors.

There are several ways to implement a momentum strategy. In our study we decide to implement the so-called Clenow Momentum⁴. Clenow proposes an exponential regression in which we run a regression of log prices on time. However, in addition to the trend also the goodness-of-fit are important. In this way momentum is calculated by multiplying the annualized exponential regression slope of the past 180 days by the R-squared coefficient of the regression calculation:

```
def get_clenow_momentum(title, days):
    prices = closing_prices[title].iloc[days:
        → days + ROLLING_WINDOW_SIZE]
    prices_log = np.log(prices)
    x = np.arange(len(prices_log))
    slope, _, rvalue, _, _ = linregress(x,
        → prices_log)
    m = ((1 + slope) ** 252) * (rvalue ** 2)
    return m
```

After running the code, the result is as shown in the Figure [13] and [14]. Depending on the type of selector, returns increased or decreased, more or less significantly. Variability, on the other hand, remained very similar. The most important change certainly concerns portfolios with the riskiest securities. The portfolio with the highest specific risk stocks went from negative to positive return, even better than others. And the portfolios with maximum total risk stocks and better absolute returns went from underperforming the index to overperforming

<https://www.amazon.it/Stocks-Move-Beating-Momentum-Strategies/dp/1511466146/>

	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_r2	13.8%	30.67%
1	absolute_returns	15.18%	30.19%
2	min_r2_and_high_specific_risk	7.15%	21.71%
3	max_specific_risk	11.84%	29.63%
4	high_systematic_risk	17.02%	34.98%
5	low_systematic_risk	3.45%	19.62%
6	positive_alpha	10.29%	28.43%
7	positive_alpha_and_high_beta	15.84%	34.31%
8	max_total_risk	17.08%	34.99%
9	min_total_risk	3.45%	19.62%
10	excess_return_over_total	9.44%	27.57%

Figure 13: Annualized return and volatility with Momentum

	Portfolio Title	Sharpe Ratio	MDD	CL	Var 90	Var 95	Var 99	IR	M2
0	max_r2	0.415989	-0.324446	0.425407	-0.0242118	-0.0312308	-0.0443973	-0.112278	0.115333
1	absolute_returns	0.468196	-0.339194	0.447477	-0.0237689	-0.0306779	-0.0436378	0.0118272	0.128497
2	min_r2_and_high_specific_risk	0.28113	-0.260577	0.274246	-0.0172398	-0.0222074	-0.0315258	-0.519758	0.0813276
3	max_specific_risk	0.364244	-0.379188	0.312158	-0.0234508	-0.030232	-0.0429522	-0.221401	0.102285
4	high_systematic_risk	0.456785	-0.384213	0.4431	-0.0275676	-0.0355741	-0.0505931	0.133443	0.125619
5	low_systematic_risk	0.122868	-0.235603	0.146623	-0.015701	-0.0201908	-0.028613	-0.781894	0.0414215
6	positive_alpha	0.325353	-0.327148	0.314663	-0.022544	-0.0290507	-0.0412562	-0.353882	0.0924786
7	positive_alpha_and_high_beta	0.431187	-0.380265	0.416466	-0.0270677	-0.0349191	-0.0496471	0.0553922	0.119165
8	max_total_risk	0.458434	-0.384213	0.444641	-0.027568	-0.0355753	-0.0505957	0.137339	0.126035
9	min_total_risk	0.122868	-0.235603	0.146623	-0.015701	-0.0201908	-0.028613	-0.781894	0.0414215
10	excess_return_over_total	0.304666	-0.366548	0.257637	-0.0218825	-0.0281921	-0.0400279	-0.512999	0.0872623

Figure 14: Advanced Metric with Momentum

it⁵ Looking at the Sharpe Ratio values, in this case the portfolio that obtains the highest return per unit of risk is the one obtained by taking into consideration the highest absolute returns. Portfolios that have the lowest Sharpe value are those obtained considering the lowest systematic risk and the lowest values of the total risk. Without the momentum calculation, the portfolio built with the lowest systematic risk was the one with the highest value of Sharpe Ratio. Considering the Maximum Drawdown values, the portfolios that have the maximum observed loss from a peak to a trough, before a new peak is attained, are those built with the highest systematic risk values and the highest total risk. The portfolios that have the lowest value of the MDD are the same obtained without the momentum calculation. The highest Calmar ratio value is obtained from the portfolio built taking into account the highest absolute returns. The port-

folios that are most at risk of drawdown are those obtained with the lowest systematic risk and the lowest values of the total risk. Without the momentum calculation, the portfolio built with the lowest systematic risk was the one with the highest value of Calmar Ratio. Unlike before, no portfolio has a negative value of the Calmar Ratio. Observing the values of the Value at Risk calculated with a confidence level of 90%, with a confidence level of 95% and with a confidence level of 99%, the portfolio that has the risk of having a greater loss than the other portfolios is always the one built with the highest values of the total risk and it is the same portfolio obtained without the momentum calculation. Looking at the values of the Information Ratio, there are four portfolios with positive values and the one with the highest value is the one obtained considering the highest values of total risk. When we built the portfolios without the momentum calculation, the only portfolios with a positive value of IR were those built with highest alpha and beta

⁵The NDX annualized return was 15.01% and annualized volatility of 25.22%

coefficients and with highest R-squared. Here, the portfolio obtained with R-squared has a negative value. Considering the Modigliani Ratio, the portfolio that has the highest value is the one built considering the highest values of absolute returns, unlike when we built the portfolios without the momentum calculation where the portfolio with the highest value was the one with the lowest values of the systematic risk. Now, the portfolio built with the lowest values of the systematic risk is the one with the lowest value of M2.

ii. Inverse Volatility Weights

In inverse volatility strategy the risk is measured with volatility, and assets are weighted in inverse proportion to their risk. An **inverse volatility weighted portfolio** is one in which highly volatile assets are assigned smaller weights and low volatile assets are allotted larger weights. The weights are given by,

$$W_i = \frac{1}{\sigma_i} \frac{1}{\sum_j \frac{1}{\sigma_j}} \quad (7)$$

In Figure [15] we can see the results. From these results we can see how depending on the type of portfolio the annual return either improved or slightly worsened. But the important

thing you see is that in general volatility has decreased for everyone to a greater or lesser degree.

iii. Simple returns vs Log returns

We have implemented a strategy where the return of our portfolios over any time period is the weighted sum of all the returns from each of the security. However we need to remember that log returns is a continuous compounded rate over time. When you add log returns, you compound. Across different stocks within the same time period, there are no compounding elements here, and so the sum of these returns represents only an approximation. So for computing portfolio return across contributions from securities within the same time period, simple returns might be a better choice. Let's look at an example:

	Weight	Simple return	Log return
AAPL	0.20	45.00%	37.16%
GOOGL	0.20	12.00%	11.33%
TSLA	0.20	15.00%	13.98%
AMZN	0.20	16.00%	14.84%
NFLX	0.20	-30.00%	-35.67%
Portfolio	1.00	11.60%	8.33%

To analyze the results with simple returns, we converted the log returns to simple returns before calculating the daily portfolio return.

	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_r2	17.06%	29.48%
1	absolute_returns	10.13%	28.94%
2	min_r2_and_high_specific_risk	7.94%	27.62%
3	max_specific_risk	-0.44%	34.09%
4	high_systematic_risk	8.16%	38.02%
5	low_systematic_risk	12.48%	18.57%
6	positive_alpha	5.07%	26.98%
7	positive_alpha_and_high_beta	15.67%	35.48%
8	max_total_risk	8.12%	38.04%
9	min_total_risk	12.24%	18.55%
10	excess_return_over_total	11.12%	24.7%

Figure 15: Inverse Volatility Portfolios

	Portfolio Title	Annualized Returns	Annualized Volatility
0	max_r2	22.8%	29.92%
1	absolute_returns	19.37%	30.32%
2	min_r2_and_high_specific_risk	17.95%	29.25%
3	max_specific_risk	14.09%	35.49%
4	high_systematic_risk	21.45%	38.34%
5	low_systematic_risk	18.01%	19.2%
6	positive_alpha	14.29%	28.7%
7	positive_alpha_and_high_beta	26.6%	35.99%
8	max_total_risk	21.33%	38.36%
9	min_total_risk	17.77%	19.18%
10	excess_return_over_total	16.91%	25.49%
11	Nasdaq-100	18.18%	25.11%

Figure 16: Simple Returns

The formula for conversion is as follows:

$$r = e^{\ln(1+r)-1} \quad (8)$$

And in the end we got the results in Figure [16]

REFERENCES

- [1] <https://www.investopedia.com/ask/answers/102714/whats-difference-between-alpha-and-beta.asp>
- [2] <https://www.borsaitaliana.it/notizie/sotto-la-lente/nasdaq.htm>
- [3] <https://investmentcache.com/magic-of-log-returns-concept-part-1/>
- [4] <https://chrischow.github.io/dataandstuff/2018-11-10-how-not-to-invest-clenow-momentum/>
- [5] <https://www.quant-investing.com/blog/this-easy-to-use-adjusted-slope-momentum-strategy-performed-7-times-better-than-the-market>