University of Milano-Bicocca
AA 2021-2022 - Financial Market Analytics Course

Building portfolios based on the NASDAQ-100 index

Simone Farallo[1], Remo Marconzini[2]

Abstract
The project consists in of the creation of a portfolio on the time series of the past five years of given index, following the principles of the Capital Asset Pricing Model. The goal of the project is to understand the factors underlying the performance of a specific portfolio. To understand this empirically, we need to construct real portfolios that are concentrated/inclined with respect to a specific level and type of risk, in practice it will be necessary to determine the amount of total risk that we take in investing in a specific portfolio, divided then into specific risk and systematic risk. The index chosen was the Nasdaq100: it is a basket of the 100 largest, most actively traded U.S companies listed on the Nasdaq stock exchange. The index includes companies from various industries except for the financial industry, like commercial and investment banks. These non-financial sectors include retail, biotechnology, industrial, technology, health care, and others.

[1,2,3,4] *Dipartimento di Informatica, Sistemistica e Comunicazione, Università degli studi di Milano-Bicocca, Milano, Italia*

Contents

1. Introduction

We constructed and analyzed 10 portfolios regarding the historical series of the last 5 years of the NASDAQ 100 financial index by following the rules of the CAPM(Capital Asset Pricing Model and analyzing the SML(Secure Market Line). The data were obtained from FactSet and were manipulated and prepared to perform a rolling regression, using 180-day samples. Then, the portfolios were rebalanced weekly, and for each portfolio the results obtained were analyzed.

2. Data Acquisition

After choosing the NASDAQ-100 index we moved on to the data acquisition phase, the data was obtained from the FactSet platform which collects economic-financial data from around the world and provides tools for analyzing, comparing, and exporting the data, targeting professionals and scholars in the field of finance. FactSet creates flexible, open data and software solutions for more than 170,000 investment professionals worldwide, providing instant, anytime, anywhere access to the financial data and analytics that investors use to make crucial decisions. The index is constituted according to a modified capitalisation methodology. This method uses individual weights of the included companies based on their market capitalisation. The weights are necessary in order to limit the influence of the largest companies and thus balance the index with all members. The index is revalued annually at the close of trading on the third Friday in December. The eligibility criteria are applied using market data as of the end of October and total shares outstanding as of the end of November. Securities that fulfil the criteria are included in the index until the following year. Additions and removals of securities are made effective after the close of trading on the third Friday in December. In addition, if at any time during the year other than the valuation, a security in the index no longer meets the eligibility criteria or is no longer eligible for continued inclusion in the index, the security is removed from the index. It is therefore necessary to identify the various movements over the past 5 years. Specifically:

- Lists of the companies that make up the index for each year were downloaded.
- A year-by-year comparison was made to identify which companies, if any, left the index and which entered.
- As a result, a list of all the companies that made up the index in the last 5 years was obtained, indicating their presence/non-presence year by year.

Name	Symbol	2017	2018	2019	2020	2021	2022
Cintas Corporation	**CTAS-US**	CTAS-US	CTAS-US	CTAS-US	CTAS-US	CTAS-US	CTAS-US
Fastenal Company	**FAST-US**	FAST-US	FAST-US	FAST-US	FAST-US	FAST-US	FAST-US
Vodafone Group Plc Sponsored ADR	**VOD-US**	VOD-US	VOD-US				
Walgreens Boots Alliance Inc	**WBA-US**	WBA-US	WBA-US	WBA-US	WBA-US	WBA-US	WBA-US
Mattel, Inc.	**MAT-US**	MAT-US					
Autodesk, Inc.	**ADSK-US**	ADSK-US	ADSK-US	ADSK-US	ADSK-US	ADSK-US	ADSK-US
Analog Devices, Inc.	**ADI-US**	ADI-US	ADI-US	ADI-US	ADI-US	ADI-US	ADI-US
Illumina, Inc.	**ILMN-US**	ILMN-US	ILMN-US	ILMN-US	ILMN-US	ILMN-US	ILMN-US
Henry Schein, Inc.	**HSIC-US**	HSIC-US	HSIC-US	HSIC-US			
NortonLifeLock Inc.	**NLOK-US**	NLOK-US	NLOK-US	NLOK-US			
Xilinx, Inc.	**XLNX-US**	XLNX-US	XLNX-US	XLNX-US	XLNX-US	XLNX-US	
Microsoft Corporation	**MSFT-US**	MSFT-US	MSFT-US	MSFT-US	MSFT-US	MSFT-US	MSFT-US
Marriott International, Inc. Class A	**MAR-US**	MAR-US	MAR-US	MAR-US	MAR-US	MAR-US	MAR-US
Starbucks Corporation	**SBUX-US**	SBUX-US	SBUX-US	SBUX-US	SBUX-US	SBUX-US	SBUX-US
Monster Beverage Corporation	**MNST-US**	MNST-US	MNST-US	MNST-US	MNST-US	MNST-US	MNST-US
Dollar Tree, Inc.	**DLTR-US**	DLTR-US	DLTR-US	DLTR-US	DLTR-US	DLTR-US	DLTR-US
Cerner Corporation	**CERN-US**	CERN-US	CERN-US	CERN-US	CERN-US	CERN-US	
Costco Wholesale Corporation	**COST-US**	COST-US	COST-US	COST-US	COST-US	COST-US	COST-US
Fiserv, Inc.	**FISV-US**	FISV-US	FISV-US	FISV-US	FISV-US	FISV-US	FISV-US
Regeneron Pharmaceuticals, Inc.	**REGN-US**	REGN-US	REGN-US	REGN-US	REGN-US	REGN-US	REGN-US
Celgene Corporation	**CELG-US**	CELG-US	CELG-US	CELG-US			
Qualcomm Incorporated	**QCOM-US**	QCOM-US	QCOM-US	QCOM-US	QCOM-US	QCOM-US	QCOM-US
Skyworks Solutions, Inc.	**SWKS-US**	SWKS-US	SWKS-US	SWKS-US	SWKS-US	SWKS-US	SWKS-US
BioMarin Pharmaceutical Inc.	**BMRN-US**	BMRN-US	BMRN-US	BMRN-US	BMRN-US		
Sirius XM Holdings, Inc.	**SIRI-US**	SIRI-US	SIRI-US	SIRI-US	SIRI-US	SIRI-US	SIRI-US
Alexion Pharmaceuticals, Inc.	**ALXN-US**	ALXN-US	ALXN-US	ALXN-US	ALXN-US	ALXN-US	
Tractor Supply Company	**TSCO-US**	TSCO-US					
Automatic Data Processing, Inc.	**ADP-US**	ADP-US	ADP-US	ADP-US	ADP-US	ADP-US	ADP-US
Biogen Inc.	**BIIB-US**	BIIB-US	BIIB-US	BIIB-US	BIIB-US	BIIB-US	BIIB-US
Gilead Sciences, Inc.	**GILD-US**	GILD-US	GILD-US	GILD-US	GILD-US	GILD-US	GILD-US
Akamai Technologies, Inc.	**AKAM-US**	AKAM-US					
Amgen Inc.	**AMGN-US**	AMGN-US	AMGN-US	AMGN-US	AMGN-US	AMGN-US	AMGN-US
Ross Stores, Inc.	**ROST-US**	ROST-US	ROST-US	ROST-US	ROST-US	ROST-US	ROST-US
Hasbro, Inc.	**HAS-US**	HAS-US	HAS-US	HAS-US			

Figure 1. List of the first 34 stocks.

Daily time series for the past 5 years have been downloaded for each company that appeared in the index. Considering the data from when the company actually appeared in the index, we took the third Friday of December as the reference day for entering and leaving the index. Although some companies have been present within the index for the past 5 years, now they are no longer listed due to MA, delisting etc.. Therefore, the companies listed below are those whose historical data could not be obtained. Furthermore, the reasons why they are no longer listed is also provided:

- Xilinx, Inc(acquired in February 2022 by AMD)
- Cerner corporation (acquired by Oracle in December 2021)
- Celgene corporation(The company was acquired by Bristol-Myers Squibb in 2019)
- Alexion Pharmaceuticals, Inc.(AstraZeneca acquired the company in July 2021. Upon completion of the merger, Alexion shareholders will hold about 15 percent of the combined company)
- Maxim Integrated Products, Inc.(On August 26, 2021, the company was acquired by Analog Devices.
- Netease Inc Sponsored ADR (Chinese company no news disclosed)
- Express Scripts Holding Company (On March 7, 2018, it was announced that Cigna would buy Express Scripts in a 67 billion of dollar deal)
- Hologic (In March 2022, merge with Women's Tennis Association and Hologic announced a multi-million, multi-year partnership)
- Shire(Shire was acquired by Takeda Pharmaceutical Companyon January 8, 2019)
- Liberty Global Plc LiLAC Group Class C (In September 2021, Liberty Global announced the sale of its Polish operations to Iliad Group's Play (P4) subsidiary for 1.8 billion of dollars, also held several sales and divisional)
- Discovery, Inc. Class C(Merged with Warner Bros 2021)
- CA (In July 2018 it was acquired for 18.9 billion of dollars in cash by Broadcom)
- Viacom Inc. Class B(On December 4, 2019, the company merged with CBS Corporation to form the new company named ViacomCBS)
- Twenty-First Century Fox, Inc. Class A,Twenty-First Century Fox, Inc. Class B (no information on this)
- Liberty Interactive Corporation Class A and Liberty Interactive Corporation Class A (changed name to Qurate)
- Liberty Global Plc LiLAC Group Class A (sold several divisions to other companies and made several mergers of the sold stocks ,etc.)

Information regarding the considered was taken from Investopedia, Wikipedia, and Factset.

3. Data pre-processing

3.1 Data exploration

As mentioned above, stocks that leave the index after the annual revaluation are not replaced until the following year, consequently there aren't exactly 100 stocks for each day considered. The price level of some securities within the index for the five years considered is shown below:

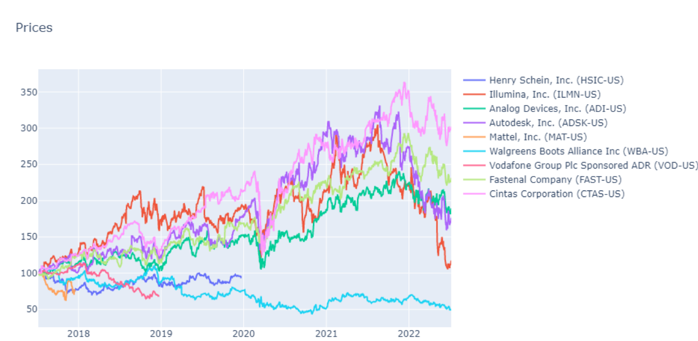

Figure 2. Price level of some securities.

3.2 Data preparation

As a first step, all securities that didn't have a sufficient number of observations for the estimation of the rolling regression were removed. The total number of titles was reduced from 137 to 120.

Next, we calculate the percentage change in share prices for each day. Specifically, log-returns are calculated, as the difference between the natural logarithm of the price on day $t+1$ and the natural logarithm of the price on day t:

$$\ln(p_{t+1}) - \ln(p_t)$$

4. Rolling regression

Rolling regressions are one of the simplest models for analysing changing relationships among variables overtime. They use linear regression but allow the data set used to change over time. In most linear regression models, parameters are assumed to be time-invariant and thus they should not change overtime. Rolling regressions estimate model parameters using a fixed window of time over the entire dataset. A larger sample size, or window, used will result in fewer parameter estimates but it uses more observations. We performed the rolling regression on our dataset following the formula derived from the CAPM:

$$r_i = \alpha_i + \beta_i(R_M) + \varepsilon_i$$

We used 180 days as sample (thus losing the first 180 observations) and we applied the rolling regression on each security, excluding those with an insufficient number of observations. The results (filtered weekly) have been obtained through the python library *statmodels* and have been saved within dictionaries, in order to extract the parameters of interest.

5. Portfolios building

5.0.1 Selecting parameters

Once the model estimates described above were obtained, we moved on to the portfolio construction phase. Several parameters have been selected, which were necessary for the subsequent creation of the portfolios. Specifically, the following parameters and some combinations of them have been considered:

- R^2: computed as the ratio between systematic and specific risk.
- α: the excess return, obtained from intercept of model
- β: the sensitive to movement in the overall market
- $\beta^2\sigma_M^2$ represents the systematic risk
- σ_{ei}^2 represents the specific risk
- σ_i^2: represent the total risk.
- The ratio between excess returns (obtained from log-returns) and total returns
- The product between R^2 and β.

5.0.2 Portfolio construction

It has been decided to construct each portfolio by selecting the top-20-percentile for each parameter considered.
Ultimately, 8 different portfolios have been obtained:

- one for each of the 8 parameters (and their combinations) described above
- one for the market index itself (which will later be used as a benchmark for comparison with the other portfolios)
- one based on the momentum strategy (we performed a top-20% momentum strategy)

6. Analysis of portfolios

In the last step, the performance of the various constructed portfolios has been evaluated in terms of returns and volatility. In particular, we will show the graphs of the various portfolios constructed divided by parameter type: basic parameters, risk type and parameter combination.

6.1 Portfolios returns

6.1.1 Portfolio based on model parameters

In Figure 3 it is possible to see the portfolios constructed based on the model parameters. We note that, for the portfolios based on alpha and beta, the performance in the first part of the time series is similar to that of the market index, while the portfolio based on R2 performs slightly less. In the January-May 2020 window, due to the advent of the epidemic, we note that there is a sharp decline in the value of portfolios, followed by a rapid increase (probably caused by the post-Covid economic recovery and rising inflation). We also note how the value of the various portfolios is more variable after this window than before the pre-Covid period.

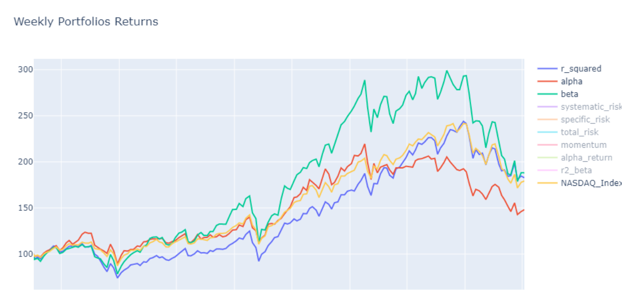

Figure 3. Time series weekly returns of portfolios based on model parameters

6.1.2 Portfolio based on risk types

In Figure 4 it is possible to see the portfolios constructed based on the risk types. Again, we note that for the first part of the time series, all portfolios based on risk types follow the trend of the market index. After the advent of the pandemic, the best-performing portfolio turns out to be the one based on specific risk.

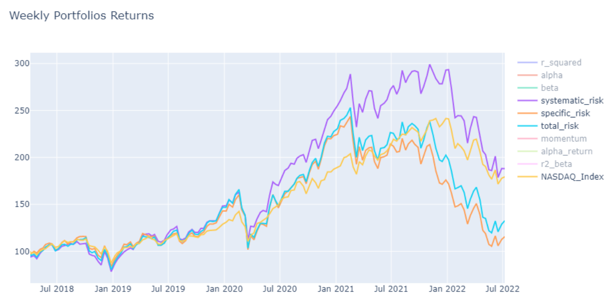

Figure 4. Time series weekly returns of portfolios based on risk types

6.1.3 Portfolios based on the combination of parameters

In Figure 7 we see the portfolio based on the combination of parameters. In particular, it is possible to note that, in the first part of time series, the portfolio based on ratio α/r produces better results than both the market index and the portfolio based on the product between R^2 and β. In the final part, however, the performance of the α/r portfolio declined, falling below the market index.

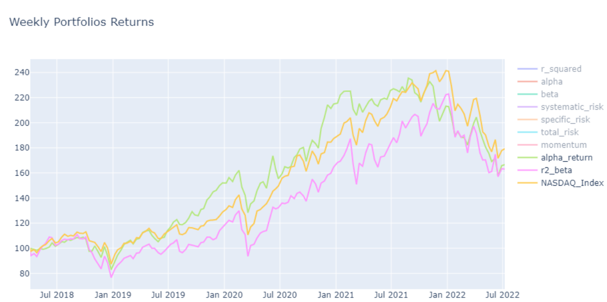

Figure 5. Time series weekly returns of portfolios based on combination of the parameters

6.1.4 Portfolio base on the momentum strategy

In Figure 5 we show the performance of the portfolio based on the momentum strategy, compared with the performance of the market index. This portfolio outperforms the market index for the duration of the time series. However, it has to be considered that this portfolio is only theoretically valid (like all others), as commission costs and taxes due to weekly re-balancing are not taken into account.

6.1.5 Annual return

Finally, the annual return has been calculated, shown in figure 7. The annual return was calculated from the weekly returns, assuming i.i.d. observations, according to this formula:

$$R_A = R_W * 52$$

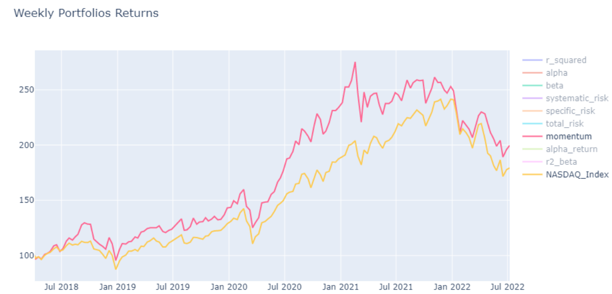

Figure 6. Time series weekly returns of portfolios based on momentum strategy

Looking at the annual average values of returns, the portfolios with the highest returns are those based on: R^2, $beta$ and momentum strategy. While the portfolios based on $alpha$ and risk type have lower values than both the other portfolios and the market index.

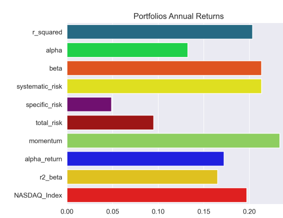

Figure 7. Annual return for each portfolios

6.2 Portfolios volatility

Volatility contributes to the problem of stocks that seem unexpectedly and inexplicably volatile, although sometimes it results in being higher than normal. We will determine the weekly volatility of each portfolio and then we'll approximate the annual volatility. In order to do this assuming that our variables are IIDs (in probability theory and statistics, a set of random variables is independent and identically distributed if each random variable has the same probability distribution as the others and all are mutually independent) we can turn monthly estimates into annual estimates. We'll approximate the annual volatility as: $\sigma_A = \sigma_W * \sqrt{52}$

In Figure 8 as expected, all constructed portfolios have higher volatility than the market index, this is because the market portfolio by including all assets is more diversified and consequently has a lower level of risk. We can observe how the trend of all portfolios changes from week to week, in particular at some times the values approach 0, this meaning that at some times the volatility of assets depended almost solely on the market and not on external events We can note the effects of the Covid 19 pandemic(February 2020 - May 2020) and the war in Ukraine(February 2022 -), the latter, as an event, seems to have a lower impact. But since the data stops at July 2022 and the war has not ended yet, probably the volatility could be higher in the following months because of

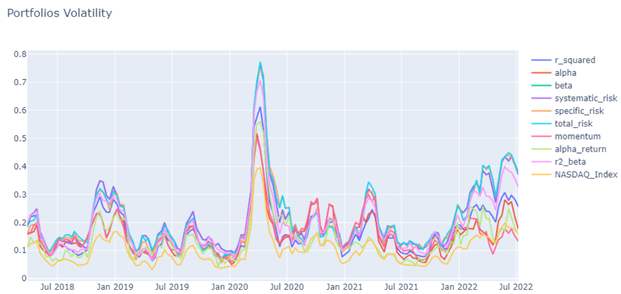

Figure 8. Time series of portfolios volatility

the economic crisis that has been accentuated in August and September. We note that risk-based portfolios have higher values than the others, while portfolios with lower values are those based on α and α_R^2, this is because only assets with significant and positive alpha values belong to this portfolio. Even the portfolio constructed using the momentum strategy, has lower volatility values relative to risk and in line with those constructed using the alpha parameter.

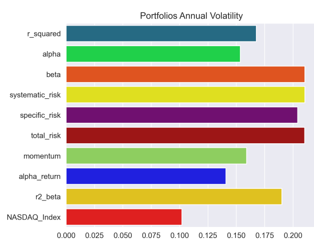

Figure 9. Annual volatility

In figure 9 graph represents the average annual volatility and this confirms our assessments.

7. Conclusions

Finally, we compare the annual values of returns and volatility In conclusion, the results are in line with the CAPM method-

Portfolio	Annual Returns	Annual Volatility
R^2	20.35%	16.77%
α	13.25%	15.35%
β	21.33%	20.05%
Systematic risk ($\beta^2\sigma_M^2$)	21.33%	21.05%
Specific risk (σ_{ei}^2)	4.89%	20.41%
Total risk (σ_i^2)	9.51%	21.03%
Momentum	**23.35%**	15.90%
α/r	17.23%	**14.09%**
$R^2\beta$	16.51%	19.02%
Nasdaq-100	19.71%	10.21%

Table 1. Values annual returns and volatility.

ology and the analyses done previously.Looking at the values of returns, the best portfolio compared with the Nasdaq 100 turns out to be the one constructed with the momentum strategy. As a matter of fact, the values of annual returns are higher than the index, while the values of volatility turn out to be similar to the values of the portfolios obtained from the parameters from the Rolling Regression. However, they are lower than those constructed on risk, the latter having lower returns than the other portfolios, especially the portfolios constructed on specific and total risk. While the portfolio with lower volatility and therefore lower risk turns out to be the one constructed with returns and the α So making a trade off between the two ratios, the best portfolio is the one built with the Momentum strategy, followed by the one built with returns and alpha parameter and the one built on R^2. Knowing that the assumptions of the CAPM model are very strong and difficult to realize in realities, to continue our analysis and obtain results closer to reality it would be appropriate to use other methodologies such as the Black-Litterman Model that it allows that merge the prior distribution (i.e. historical time series from market data) with the posterior distribution, a new set of parameters coming from the subjective views of the investors. A technical analysis could be carried out, that is an approach used by many investors to interpret in a better way the graphs of time series.

References

[1] Nasdaq-100 Index. https://indexes.nasdaq.com/docs/methodologyndx.pdf.

[2] Data acquisition. https://www.factset.com.

[3] CAPM. https://www.kaggle.com/code/gauravduttakiit/capital-asset-pricing-model.

[4] Optimization. https://www.machinelearningplus.com/machine-learning/portfolio-optimization-python-example/.

[5] Statsmodels package. https://www.statsmodels.org.

[6] Nasdaq-Composite-Index. https://www.investopedia.com/terms/n/nasdaqcompositeindex.asp.

[7] Fernando Lasso. https://www.kaggle.com/code/fernandol/will-garch-make-you-rich.

[8] Ludwig Chincarini and Daehwan Kim. *Quantiative Equity Portfolio*. McGRAW-HILL, 2006.