

WORLD_LMC_PRICE	Nan	WORLD_SE_NET	Nan
	Nan		Nan
WORLD_LMC_NET	6942.543866	WORLD_LMC_GROSS	10075.100091
	7235.726632		10504.126591

# Data Science on MSCI World Index Data

Author: Christopher Grießhaber, Stuttgart (Germany) 10.02.2021

## Inhalt

.....	1
Data Science on MSCI World Index Data.....	1
A. Introduction .....	3
A.1. Scope .....	3
A.2. Description & Discussion of the Background .....	3
MSCI World indices .....	3
ETFs (Exchange-traded funds) .....	5
TER (Total expense ratio) .....	5
Concept of rebalancing.....	5
A.3. Description of the problem .....	5
Development of profit:.....	6
Market timing:.....	6
Systematic effects on the index performance: .....	6
Rebalancing: .....	6
B. Methodology .....	6
B.1. Analytic approach and Data Requirements .....	6
B.2. Data Description & Collection .....	7
B.3. Used tools & methods.....	8
B.4. Data Understanding & Preparation.....	9
C. Results .....	11
C.1. Evaluation & Modeling.....	11
General Evaluation .....	11
Regression & simple Model building.....	16
Curve_fit regression .....	19
Annual Performance Triangles .....	22

World LMC   12months step   start from 12/2000 .....	23
EM LMC NET   12months step   start from 12/2000.....	24
WORLD SC NET   12months step   start from 12/2000.....	25
World LMC   36months step   start from 12/1969 .....	26
C.2. Investment Strategies - Market Timing.....	27
WORLD LMC NET   36months observation step   start from 12/1969 .....	28
WORLD LMC NET   12months observation step   start from 12/2000 .....	29
EM LMC NET   12months observation step   start from 12/2000 .....	30
WORLD SC NET   12months observation step   start from 12/2000.....	31
C.3. Investment Strategies - Rebalancing Effects.....	32
D. Discussion.....	38
Regression & simple Model building.....	38
Annual Performance Triangle.....	38
Investment Strategies - Market Timing.....	38
Investment Strategies - Rebalancing.....	38
E. Conclusion .....	39
F. References .....	40

# A. Introduction

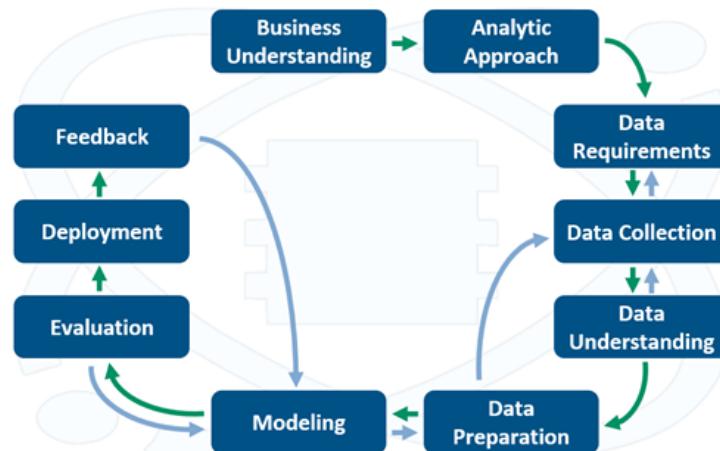
The introduction will cover the project scope, explain the basics of ETF investing and describe the investors' problem, which will be analyzed.

## A.1. Scope

This project is used for my IBM Capstone Review project. In contrast to many of the other participants, I decided not to analyze geographical phenomena, but to get insights for selected areas of the stock markets. Therefor there will be no use of the Foursquare API, which had to be used for previous submissions anyway.

Within this specific article some questions within the finance market will be analyzed. In detail, the data for the different MSCI World indices will be gathered and used to answer some questions, which (especially private) investors might have. Several performance data, investment strategies and investment behaviors will be evaluated. The reader should keep in mind, that the underlying data are historical data and all the insights will apply for the timespan of the available data only. None of the findings are meant as recommendations for the future.

For this research I used the data science methodology as shown in the graph below:



The next chapters explain core concepts, which are relevant for this kind of analyses. If you are already trained in the basics of stock investments, you might skip those explanations and continue with chapter A.3.

## A.2. Description & Discussion of the Background

### MSCI World indices

The MSCI World index is one of the most important and well-known indices within the stock markets. Several products use this index as reference for their performances. The MSCI World in detail is a market cap weighted stock market index. There are different variants of this index, which focus on different countries or company caps. The important index MSCI

World Large and Mid-Cap contains approximately 1,600 stocks from companies throughout the world (23 countries with developed markets) [1].

There are many different indices, which can be used. Within this article the focus will be on 3 MSCI indices:

- WORLD Standard (Large + Mid Cap)
- EM (EMERGING MARKETS) Standard (Large + Mid Cap)
- WORLD SMALL CAP Small Cap

You can find another variety of indices on the official webpage of MSCI. [2]

As you can see, several “Market caps” are used for the World index. The market caps define the range of the market capitalization and vary for different index solution provider. The market cap is used to set the companies into different clusters, which are approximately in a range like shown:

- Mega-cap       $\geq 200$  billion USD
- Large-cap       $\geq 10$  billion USD      and       $< 200$  billion USD
- Mid-cap       $\geq 2$  billion USD      and       $< 10$  billion USD
- Small-cap       $\geq 0.3$  billion USD      and       $< 2$  billion USD
- Micro-cap       $\geq 50$  million USD      and       $< 0.3$  billion USD
- Nano-cap     $< 50$  million USD

Mega-, Large- and Mid-caps cover nearly 85% of all capital in the markets and therefore include the most important companies. Small caps cover approximately 14%, Micro and Nano caps the last 1%.

	Region		Share of market capitalization
	World	Emerging Markets	
Market Cap	Large + Mid Cap	~1.650 shares	~85%
	Small Cap	~4.400 shares	~14%

As shown in the table the analyses will not consider the emerging markets small cap index. But still a huge range of the market capitalization will be considered properly.

## **ETFs (Exchange-traded funds)**

For private investors the most relevant use case is the purchase of ETFs. An investor can buy an ETF share, which refers to the MSCI World index and therefor participate on the performance of all included stocks (in this index 1,600). Technically there can be a "tracking error" (difference of the ETF performance in comparison to the index performance), which can be caused by the mechanism of the replication strategy and some other cost effects. [3]

The benefit for the investor is, that you do not have to buy all single 1,600 stocks, but only one ETF share, which decreases the investment costs a lot. Also, you decrease your company and region related risk (lower volatility) since you distribute your investment over different companies and countries [4].

## **TER (Total expense ratio)**

ETFs try to mirror the referred index by different methods and at minimum costs. The TER value shows to the investor how high the estimated annual costs relative to its value will be. Typical values are in a range of 0.15% to 0.5% per year. Consequently, the ETF will never reach the performance of the index, it refers to, because this index does not contain these costs [5].

## **Concept of rebalancing**

If an investor builds up an portfolio, in many cases more than one position is included.

Example 1:

- 100% of MSCI World Large + Mid Cap

Example 2:

- 50% of MSCI World Large + Mid Cap
- 30% of MSCI Emerging Markets Large + Mid Cap
- 20% of MSCI World Small Cap

In Example 1 obviously no rebalancing is necessary, because your desired position distribution of your portfolio cannot change. In example 2 rebalancing might be an option. It can be expected that the three mentioned positions develop differently and therefor after some time the value of your positions are not reflecting to your initial planned 50/30/20% any longer. The portfolio might be shifted to 40/33/27%, for example. In this case the question is valid, whether and when rebalancing to the originally planned 50/30/20% (40 + 10 / 33 - 3 / 27 - 7) of your portfolio should be done. This obviously causes buying and selling actions (transactions) by the investor, which might cause additional costs. Rebalancing in general shall adjust the risk of the portfolio [6].

## **A.3. Description of the problem**

For an investor several questions are important at this point:

### **Development of profit:**

First the investor should check whether there is any profit possible at all. Within this article the following additional questions will be checked and answered:

1. How is the performance for the indices over time?
2. Can you prognose the performance for the future?
3. Are there situations, in which the performance is negative?

### **Market timing:**

Especially private investors often ask for the right timing to invest. Subjectively the markets are always overpriced, someone always forecasts the next crash and therefore many private investors wait with the investment, sometimes forever. Due to this the following questions will be addressed:

1. Can you increase your profit by market timing?
2. What is better? Investing all the capital at the start or with equal distributed small investments over a long period?
3. Does it make sense to wait for a crash / minor crash at the stock markets to invest?

### **Systematic effects on the index performance:**

Is it possible to detect systematic effects on the index performance, which might help with market timing or to prognose the possible profit in the future?

1. How high is the performance based on the month?
2. In which months did the last crashes happen?
3. Is it possible to do regressions to forecast the performance?

### **Rebalancing:**

Does it help to rebalance your portfolio based on the three mentioned indices above?

1. How often should you rebalance?
2. What effects will the rebalancing costs have on the portfolio performance?

## **B. Methodology**

Within this section the data, used methods and tools will be explained. Also, first preparation and data understanding steps will be shown.

### **B.1. Analytic approach and Data Requirements**

To answer the questions for the investors' problem some data is needed. To address the performance topics the chart history for the analyzed index is crucial. Hereby it helps to have a maximized timespan of data as well as a maximized sampling rate (e.g. 1 value / day, 1 value / week etc.).

The data should contain crashes and be representative for its period.

To check the effects on performance based on the actual month, the course history must have this value implemented, ideally the date is recorded.

For the model building it would help, if the different indices could be parameterized (e.g. number of companies in specific branches or countries, etc.), so that the performance of the index can be analyzed based on those parameters.

It can be summarized, that the requirements on the data are low for the initial questions asked above. If sophisticated model building shall be used, this can get by far more complex.

## B.2. Data Description & Collection

MSCI provides historical index data on its webpage:

<https://www.msci.com/end-of-day-data-search>

For these analyses, the data for

- "Developed Markets Standard (Large+Mid Cap)" (in this post it will be called **WORLD LMC** from now on),
- "Developed Markets (Small Cap)" (**WORLD SC**)
- and "Emerging Markets Standard (Large+Mid Cap)" (**EM LMC**)

are used in three available variants (Price / Gross / Net). All indices are available as a downloadable spreadsheet file (.xls).

A	B	C	D	E
1				
2				
3	Index Level :	Price		
4	Currency :	USD		
5				
6				
7	Date	WORLD Standard (Large+Mid Cap)		
8	Dez 31, 1969	100,000		
9	Jan 30, 1970	94,249		
10	Feb 27, 1970	96,979		
11	Mrz 31, 1970	97,075		
12	Apr 30, 1970	87,800		
13	Mai 29, 1970	82,060		
14	Jun 30, 1970	79,837		

In total there must be downloaded nine spreadsheet files from the MSCI website. Since the files get provided by a Java-Script, the effort to download those files automatically is high. For this reason, the files just got downloaded for these analyses and all files were copied into one spreadsheet file. This new spreadsheet file looks like this:

	A	B	C	D	E	F	G	H
1	Index Level :	Net	Gross	Price	Net	Gross	Price	Net
2	Currency :	USD	USD	USD	USD	USD	USD	USD
3	Date	WORLD_LMC_Net	WORLD_LMC_Gross	WORLD_LMC_Price	EM_LMC_Net	EM_LMC_Gross	EM_LMC_Price	WORLD_SC_N
4	31.12.1969	100,000	100,000	100,000				
5	30.01.1970	94,455	94,528	94,249				
6	27.02.1970	97,405	97,558	96,979				
7	31.03.1970	97,708	97,947	97,075				

As you can see, some fields are empty. For those fields no data was available at the recorded time due to varying starting dates of the indices.

The starting dates for the indices vary like this:

- 31.12.1969: WORLD LMC Net, Gross and Price
- 31.12.1987: EM LMC Gross and Price
- 31.12.1992: WORLD SC Price
- 29.12.2000: EM LMC Net + WORLD SC Net and Gross

All used indices contain data until 31.12.2020. The data is recorded in monthly intervals.

All the following steps are performed within Python.

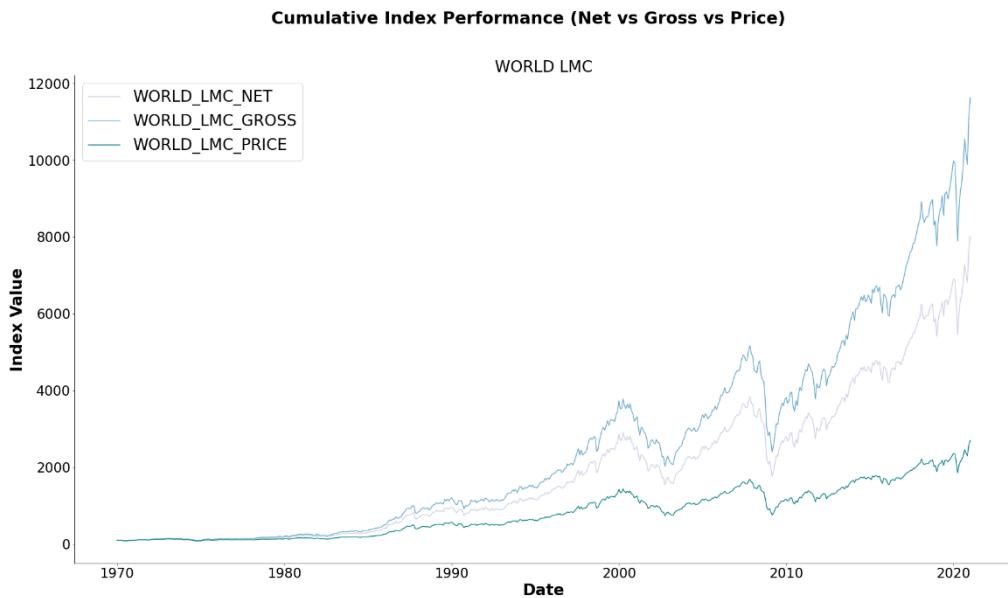
### B.3. Used tools & methods

The most important methods and tools for this project are introduced here:

- **Spyder 3.8** [7]: The used development environment is Spyder 3.8, but the code also works fine with JupyterLab and Jupyter Notebooks etc. of course. Spyder is used due to better debugging options.
- **GitHub** [8]: Repository, which is used to share the code and the data of this project
- **NumPy** [9]: Used for many scientific computing steps
- **Pandas** [10]: Primary data structure library (e.g. for the usage of dataframes)
- **Seaborn** [11] and **Matplotlib** [12]: Seaborn is used for various plots. Matplotlib is underlying to Seaborn and is used to format the plots in more detail.
- **SciPy** [13]: Here the **curve\_fit** function of the **optimize** Toolbox is used for exponential regressions steps.

## B.4. Data Understanding & Preparation

In a first step, all data are loaded into a dataframe. Afterwards some plots are created to understand the behavior and the normalization of data.



The shown index is the World LMC. In the plot the variants of this index for "Net", "Gross" and "Price" can be seen. These indices cover the developed markets in 23 developed countries. Find more details regarding this index on the official page of MSCI:

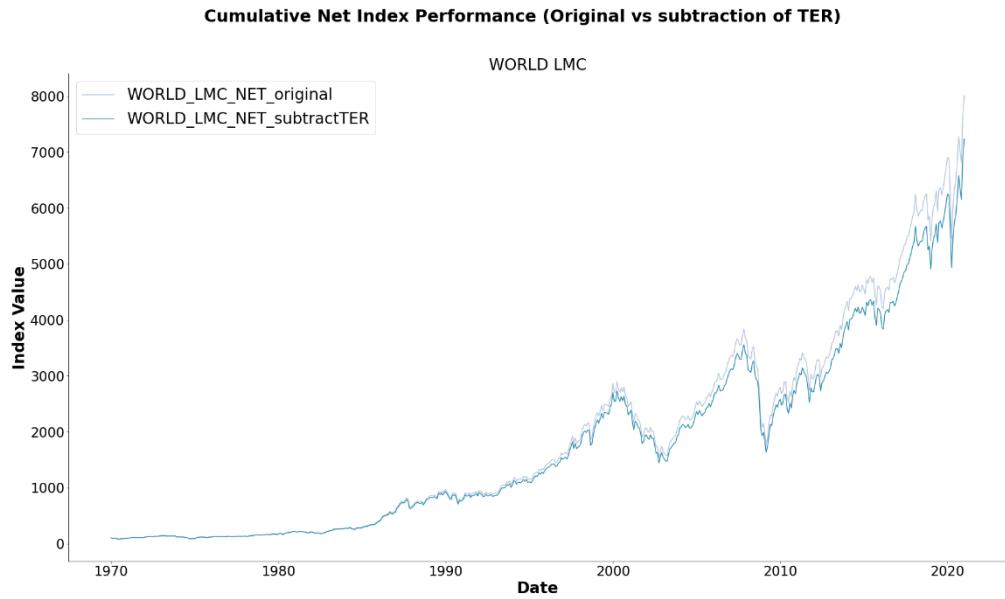
<https://www.msci.com/developed-markets>

**Price:** the price index shows the normal index value. In this index all dividend payments will not be reinvested in the index. The investor might use this money for consumption.

**Gross:** in this index the dividends will be reinvested in the index completely. This money will increase the shares of the investor. For the gross index no taxes are considered, therefor this index cannot be achieved for the investor.

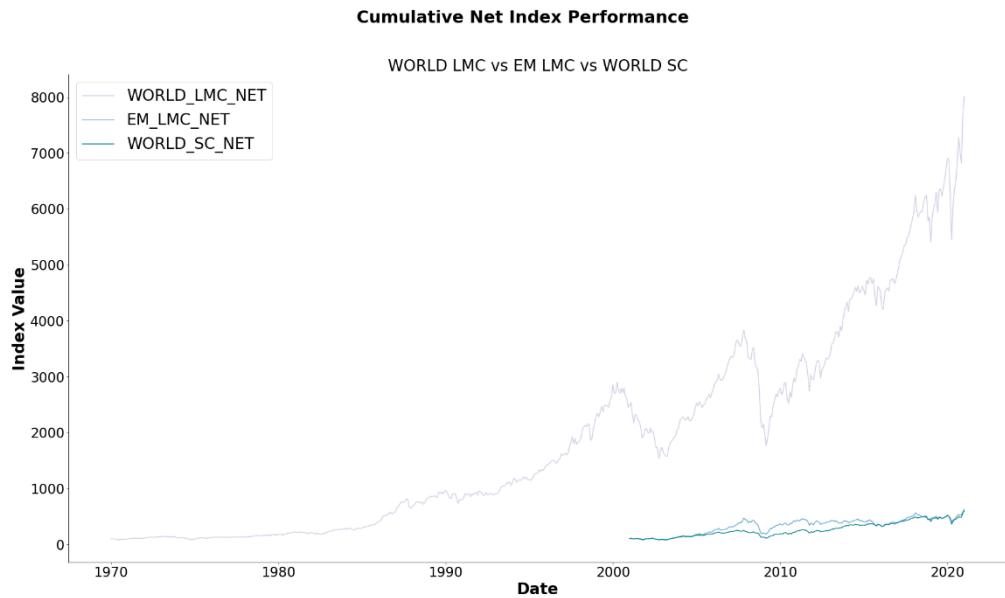
**Net:** in this index the net dividends will be reinvested in the index. This is a commonly realistic investment behavior and mirrors accumulating indices best.

The Net variant is a good choice, but for investors there always will be costs (e.g. administrative fees), if you buy and hold ETF. Therefor the TER will be subtracted from the Net indices. For the LMC indices the annual TER will be estimated with 0.2% p.a., for the SC index it will be estimated with 0.35% p.a. Those estimations are rather conservative and are easily met in Germany.

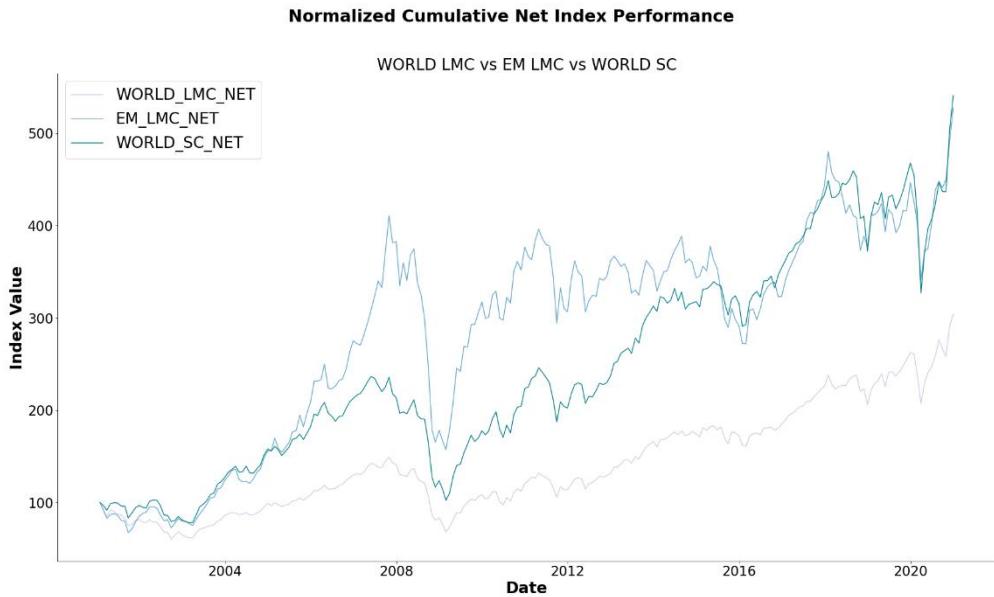


This graph for the WORLD LMC shows the impact of this TER subtraction.

The next graph shows the development of the three indices WORLD LMC, EM LMC and WORLD SC, each time for the NET variant, which will be in the focus from now on.



It is clearly visible, that all indices start with a base score of 100. Due to different starting dates, the performance cannot be compared with each other without normalization. Therefor all data will be normalized to 100 at the last starting date. Also, all data before the starting date gets cut for better visualization.



With this kind of normalization, the performance can be compared much better, but the usable timespan gets rather short. For all further analyses throughout this project in general, this kind of normalization will be used. Whenever suitable the whole available dataset will be used (for MSCI World LMC data from 31.12.1969 until 31.12.2020).

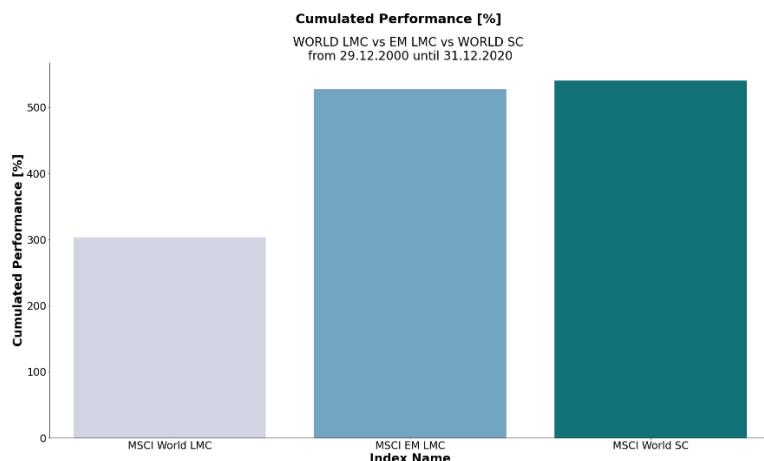
## C. Results

In this section a deeper insight into the data will be given and some easy regression models will be created.

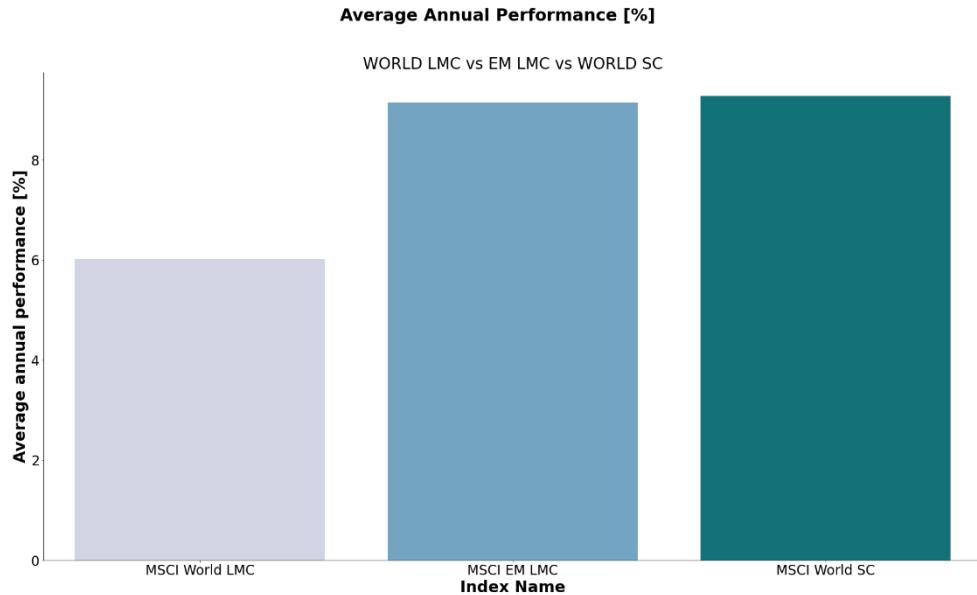
### C.1. Evaluation & Modeling

#### General Evaluation

First, the cumulated performance from 29.12.2000 until 31.12.2020 and the average annual performance will be shown.

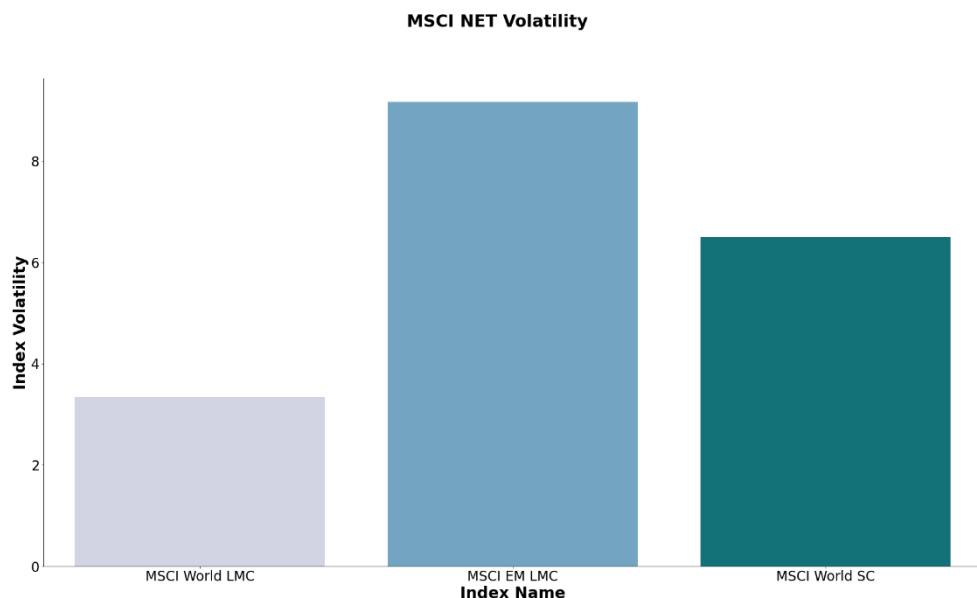


The bar plots clearly show that the cumulated performance is positive. The WORLD LMC performs good and achieves about **300%** of its original value. But the EM LMC and WORLD SC perform even better and both achieve around **500%**.



The average annual performance looks similar, of course. All indices achieve an average annual performance of **6% to 9% p.a.** Keep it in mind, that the theoretical costs for ETF already got subtracted for those data. Inflation is not considered, though.

In finance theory performance is closely related to risk. Typically values like the volatility of your data is chosen to describe this risk. Therefor the volatility will be plotted next.

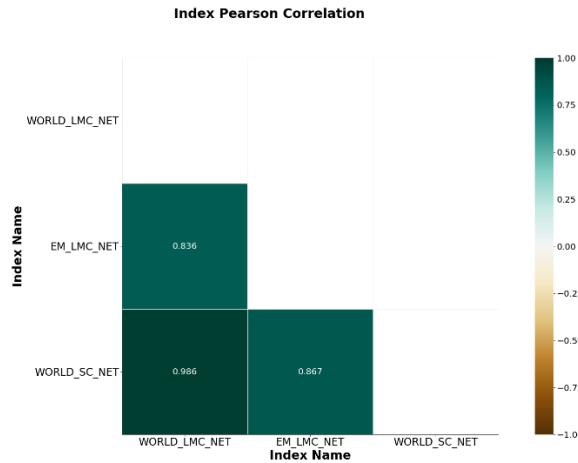


The volatility for the World LMC is lower than the volatility for the other observed indices. This fits to the expectation and the worse performance in comparison to EM LMC and World

SC. It also fits to the expectation that the risk for the industrialized regions with its large-caps can be expected to be lower.

More interesting is the observation, that the volatility for the World SC is lower than for the EM LMC (despite the effect, that the performance was comparable).

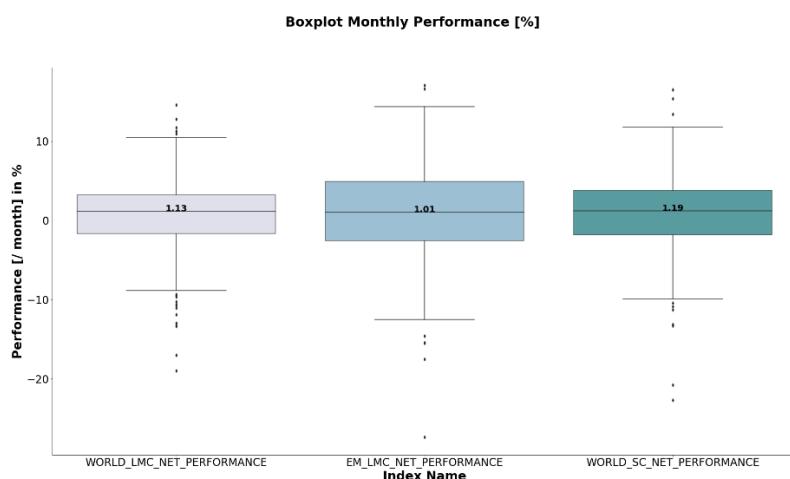
Next the linear Pearson correlation for those indices will be shown.



The World LMC and World SC show a high correlation of **0.986** (possible maximum is 1). The correlation of the EM LMC to the other two indices is lower and about **0.83** to **0.86**. To reduce the risk of a portfolio it is best to have assets included, which have a positive performance, but do not correlate positively.

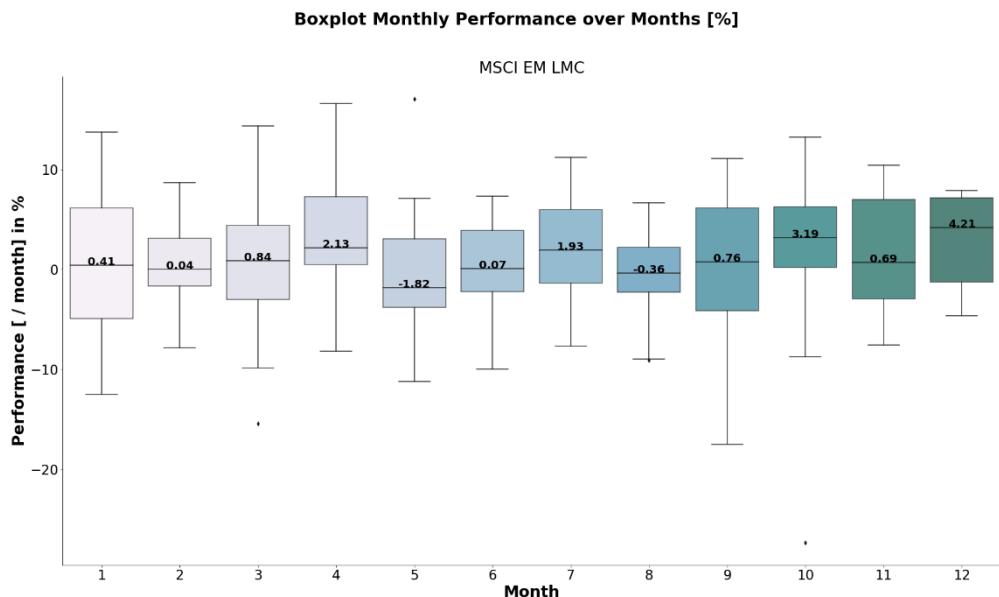
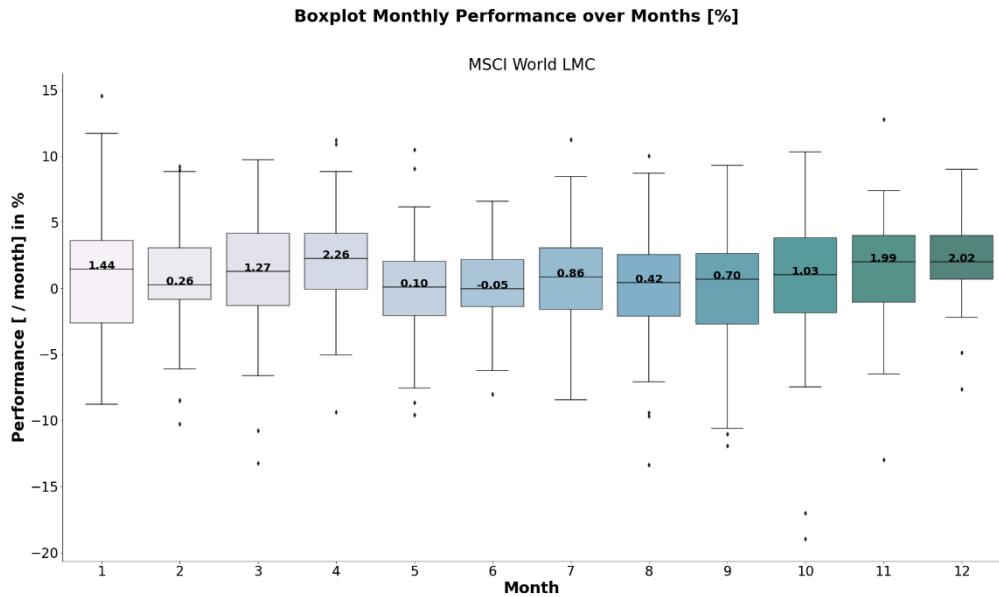
In this shown case, all values are in a similar range, which means, that a crash probably affects all indices. This is not surprising since all indices reflect to the stock market and therefore cover the same asset group. From correlation point of view it would be better to mix EM LMC and one of the two World indices in the portfolio, because the correlation is lower.

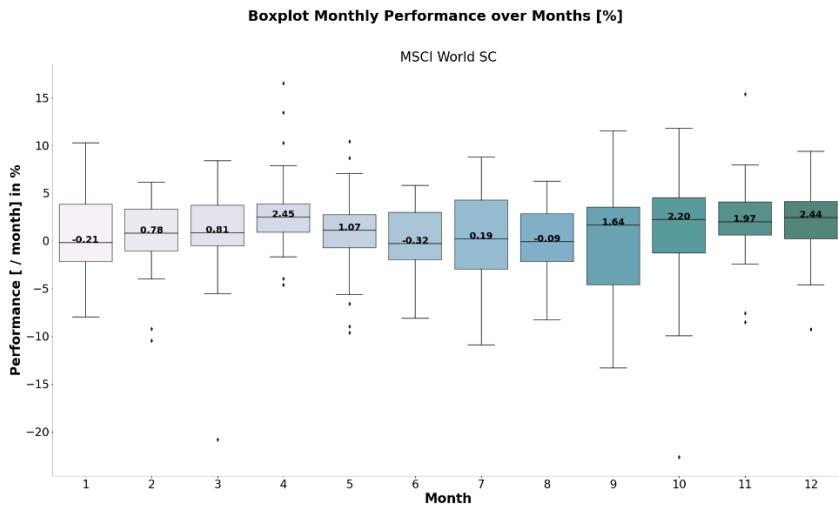
To show the span width of the performance a Boxplot will be shown. It should be considered that this time the performance is shown per month.



As expected, the highest performance for the World LMC is lower and the lowest performance is higher than the performance of the other two indices. This fits well to the observed volatility values. The median values for the three indices are **1.132%**, **1.009%** and **1.191%**. The median shows that value, for which 50% of values are lower and 50% of values are higher and differs from the often-used average value. The median value is more robust against extreme outliers in both directions.

The previous shown information is not incredibly special. Next, the performance will be shown for its corresponding month for all three indices. With this evaluation it can be checked whether there is a season dependency on the data and whether there are "better" or "worse" months to be invested.



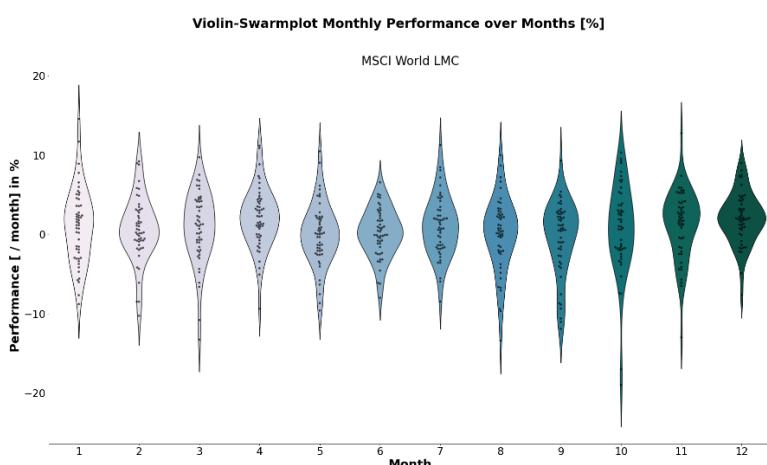


In general, it can be seen, that most of the median values are positive or only very slightly negative (Minimum median values: **World LMC = -0.046% /month**, **EM LMC = -1.823 % /month** and **World SC = -0.321% /month**).

Interestingly all three median values for April are relatively high, at least much higher than the months next by (**World LMC = 2.265% /month**, **EM LMC = 2.131 % /month** and **World SC = 2.447% /month**). A hypothesis for this positive behavior in April could be, that often companies publish their annual reports within the first three months. Perhaps the reports in average do not satisfy the expectations and therefor result in a lower performance during these months.

Additionally, it can be seen in the chart, that especially the last quarter of the year shows a better median performance as well. This could be driven by the winter and Christmas sales. The rather bad summer performance might be the result of the vacation season. Keep it in mind, that all those hypotheses are not proven and the values might just be the result of random effects. To solve those hypotheses additional data could be gathered and further analyses could be started.

To get a better impression of the distribution of the Boxplot and the Number of data points, two violin plots will be shown.

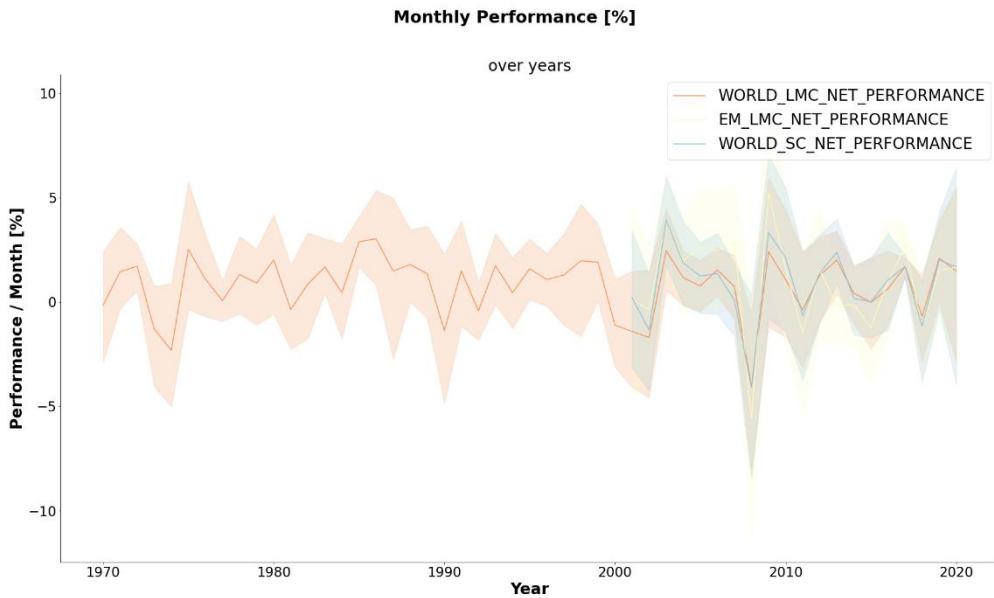




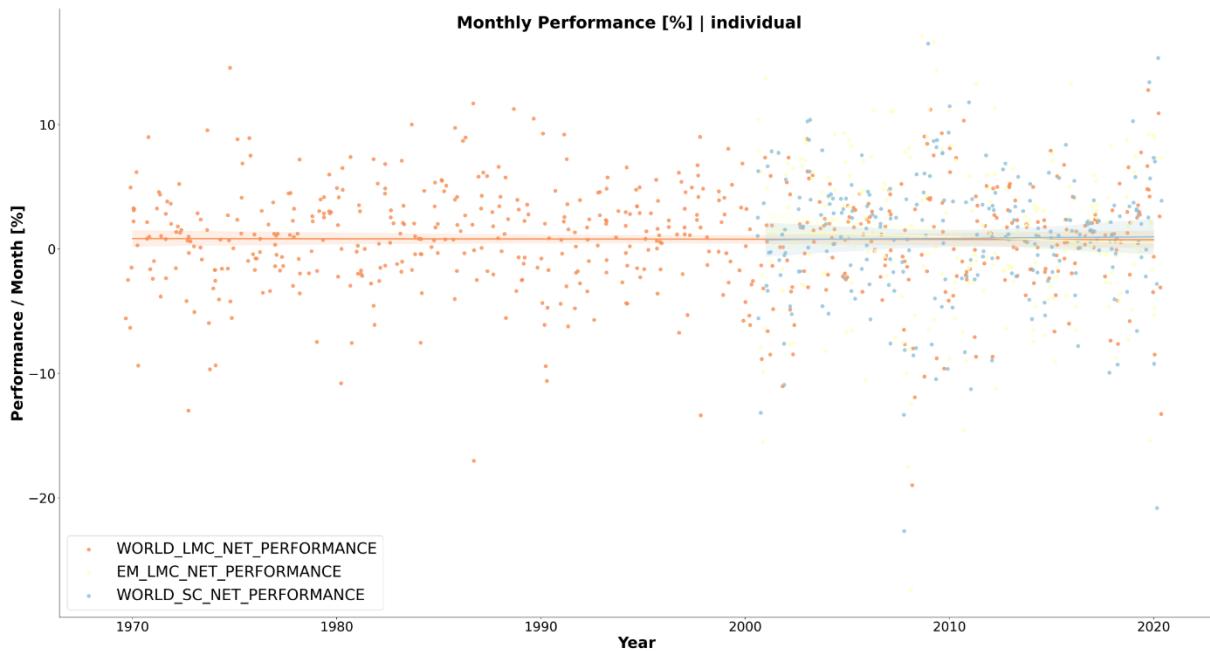
It is more difficult to use the data, but the violin plot helps to visualize the distribution shape and to detect distributions, which differ from normal gaussian distributions.

### Regression & simple Model building

The next plots will use Seaborn's LMplot function to perform a regression over the monthly performance data for all three indices. The World LMC is colored orange, the EM LMC is colored yellow and the World SC is colored blue.

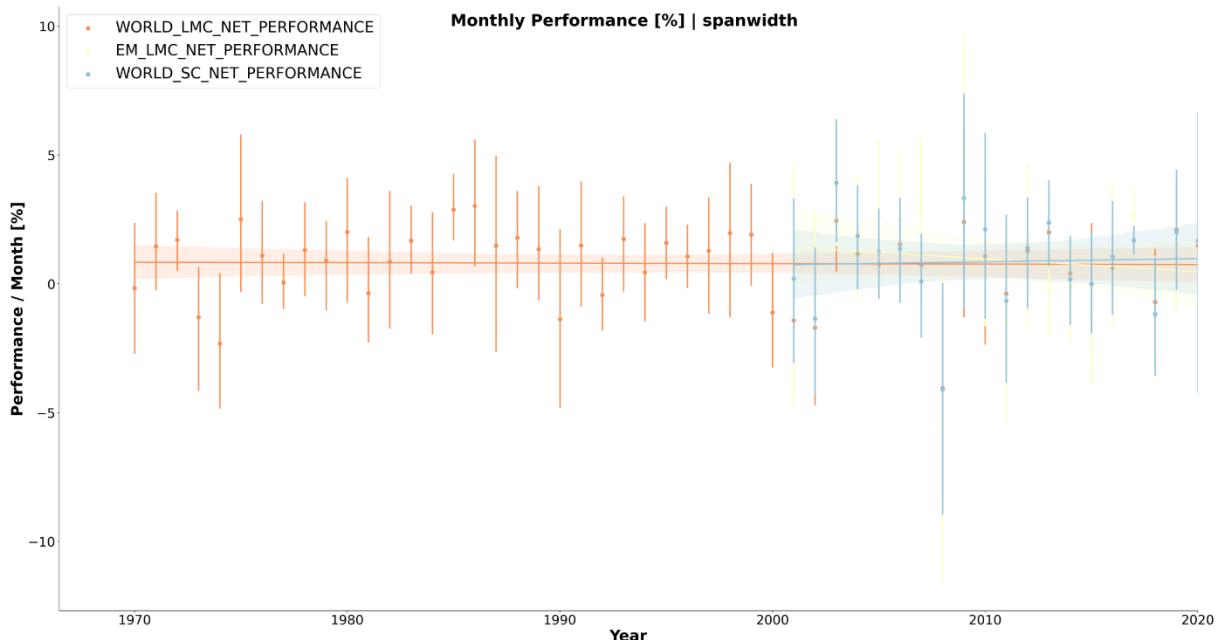


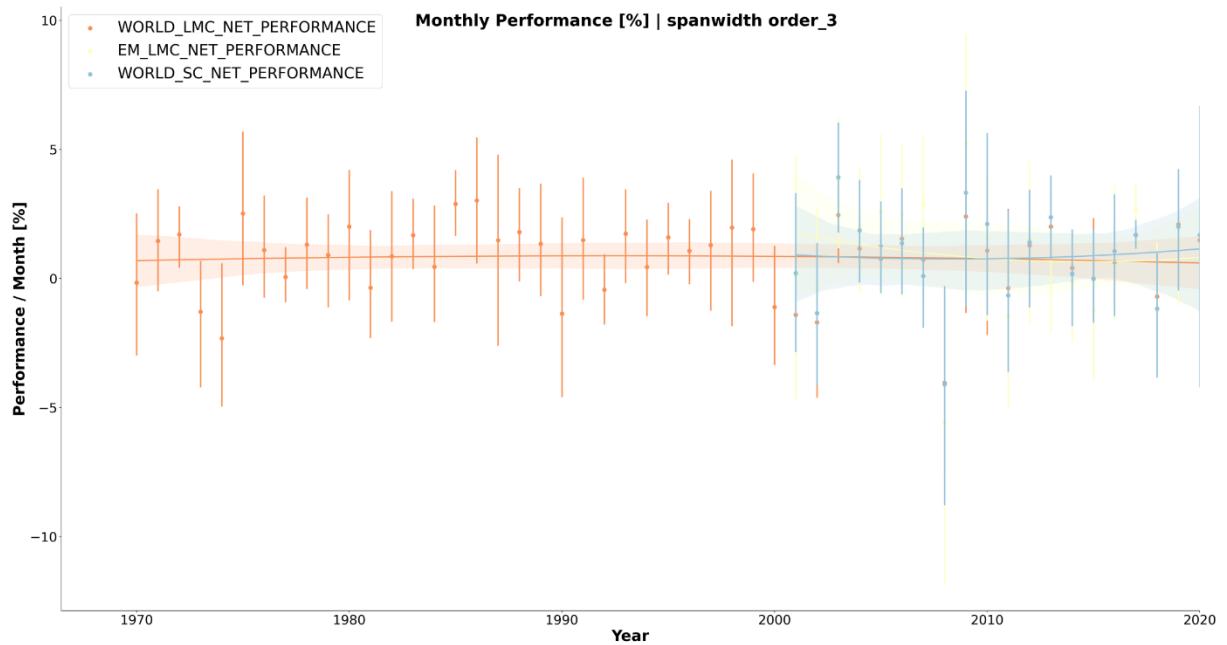
First the underlying data are shown. The dark line shows the average value for its corresponding year, the light-colored area shows the range from the minimum to the maximum value.



In this plot a linear regression can be seen. It is easily detectable, that the single values are distributed in a wide range around it and it seems to be difficult to get useful regressions. Still a more detailed look on the plots might help. Therefor the same regression is plotted, but this time the mean value as well as the span width for the annual data is plotted. As a result, the plot gets better readable.

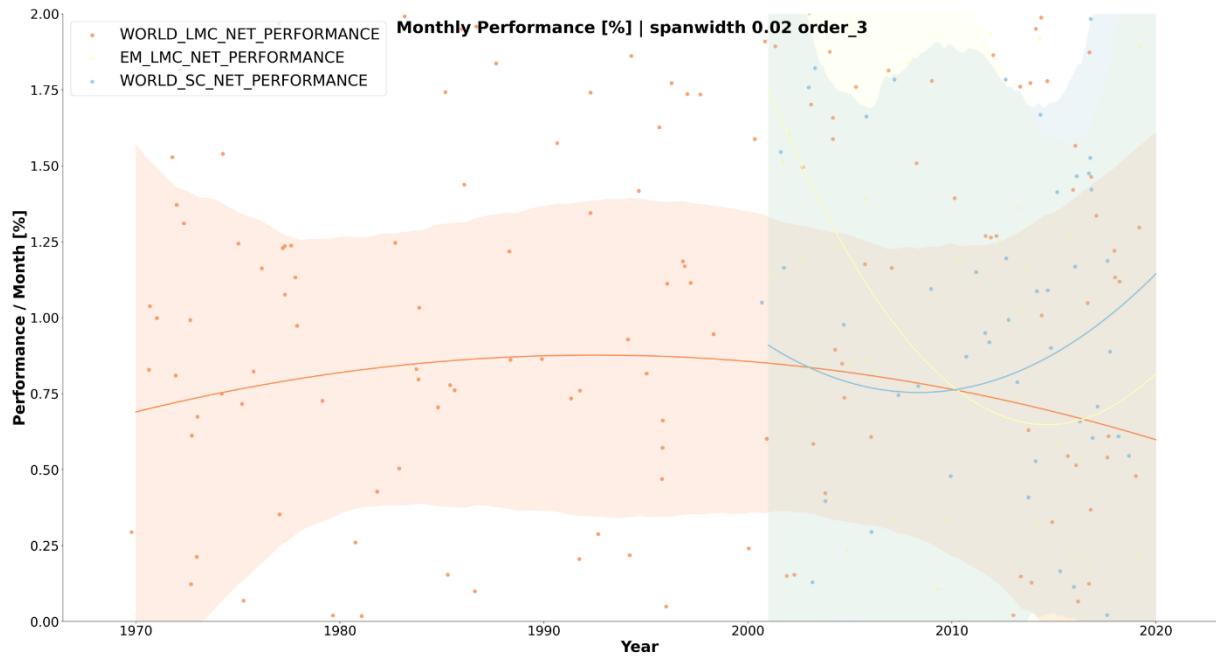
The two plots below show the regression in a linear variant (order 2) and a quadratic variant (order 3). Both variants do not fit that well.





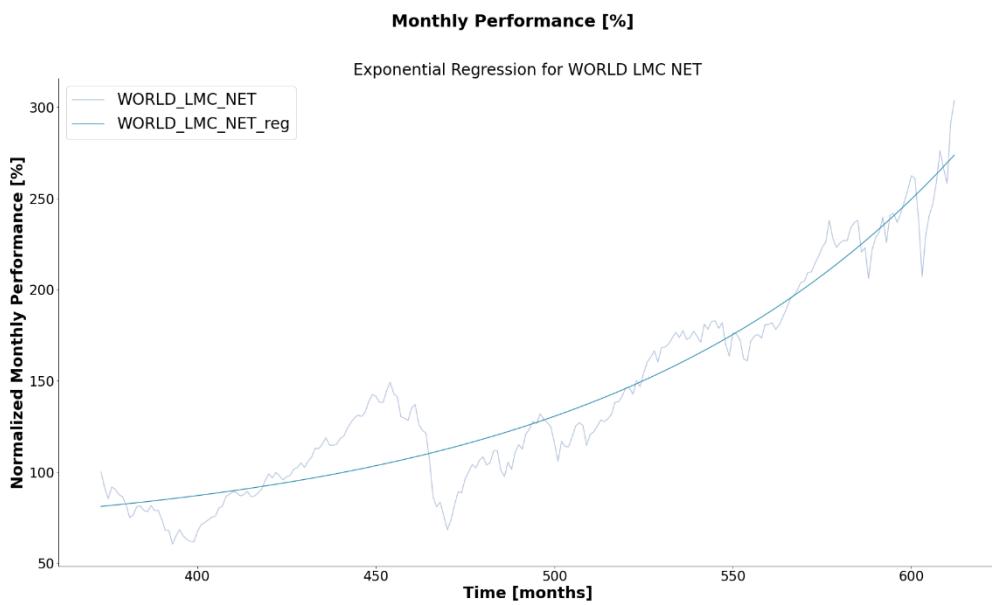
If those plots get zoomed in the trend can be evaluated better. The linear trend for World LMC shows a slightly decreasing effect, the EM LMC shows a sharp decreasing trend and the World SC shows a positive trend. Despite that information it is obvious, that the regressions do not work well.



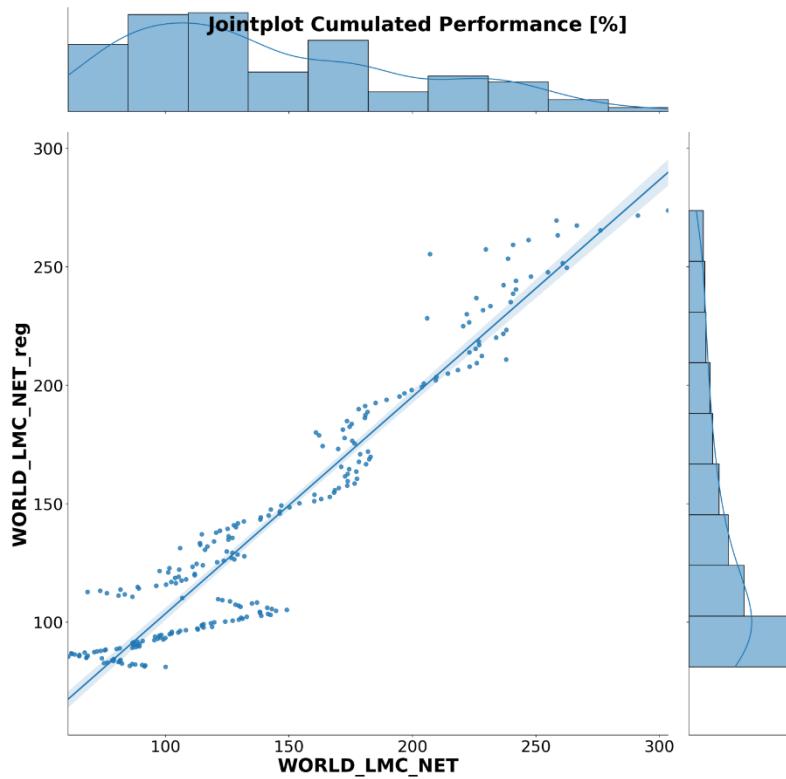


### Curve\_fit regression

To improve the regression an exponential function is defined and the `curve_fit` function is applied (using SciPy Optimize). This time the original index data is used. The data begins with the starting month in Dec 1969. The regression will be done over the absolute number of months.

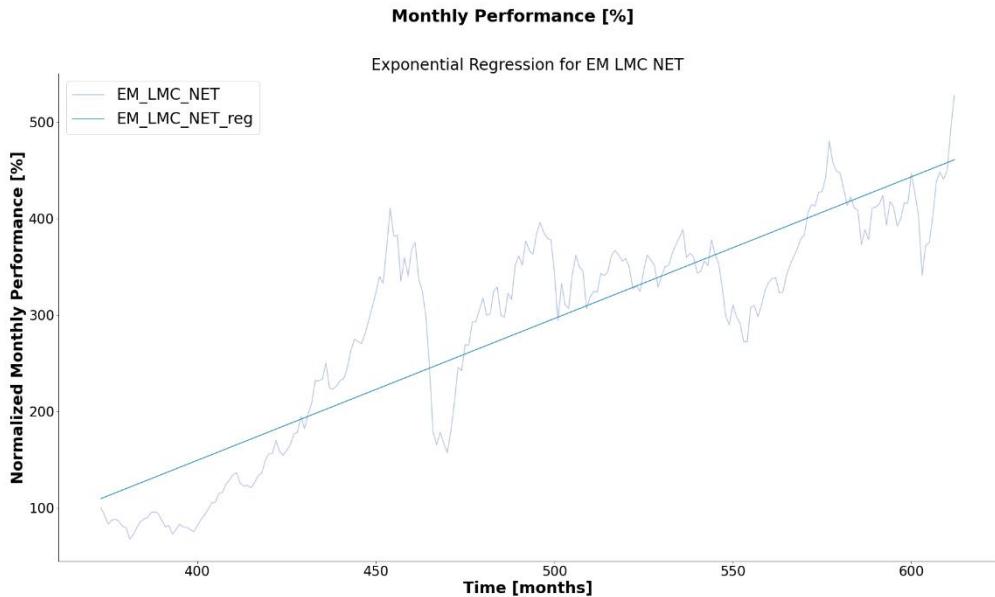


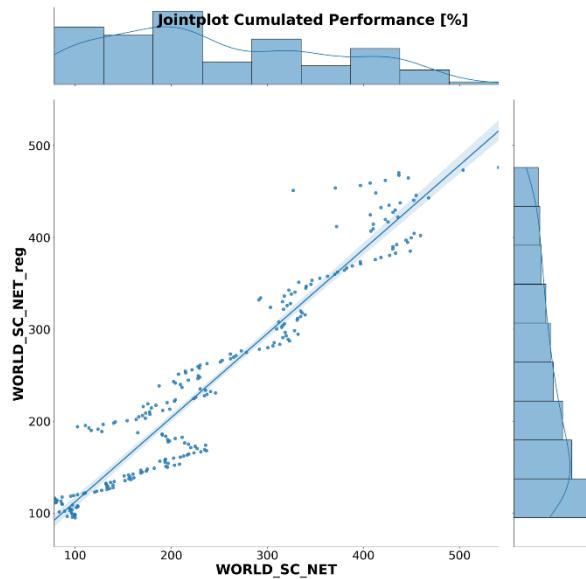
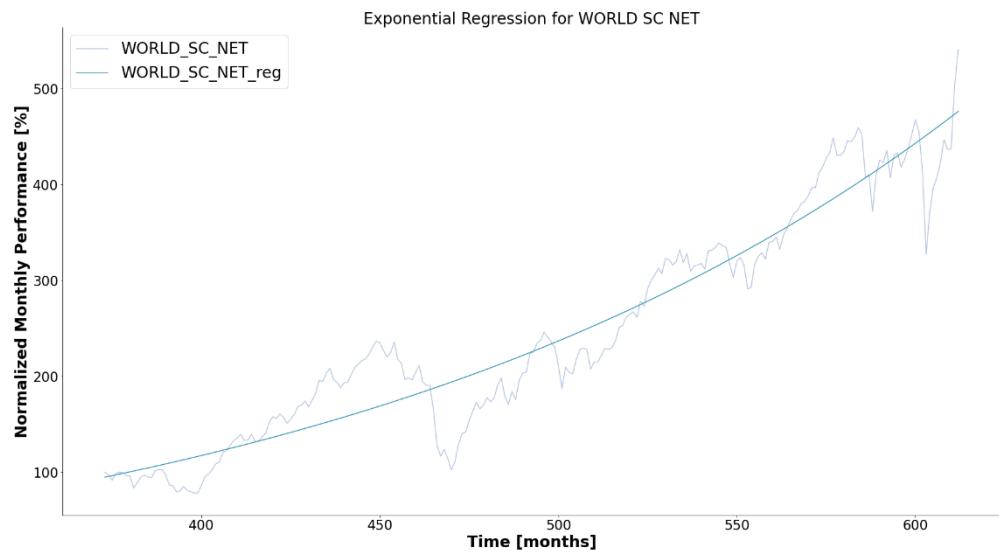
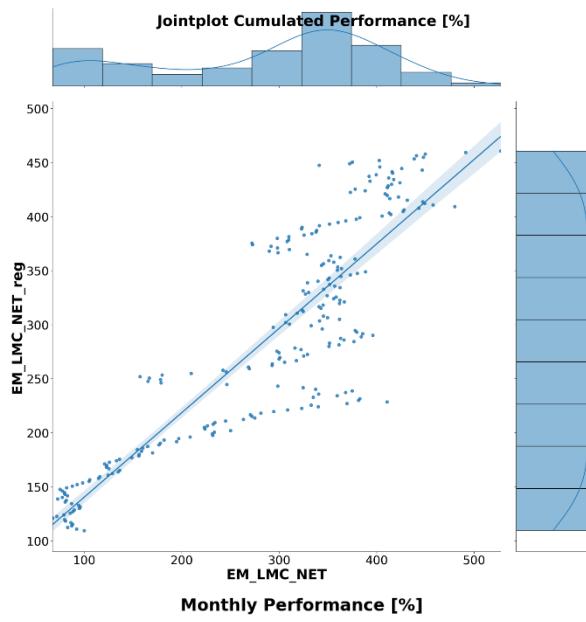
The `curve_fit` seems to work, but obviously crashes cannot be considered in the model. Also, it is not possible to rate the accuracy of the model easily, because  $R^2$  for example cannot be applied. The reason is, that this regression works for a non-linear problem and therefore the weighted linear square of  $R^2$  would deliver useless results, if applied.



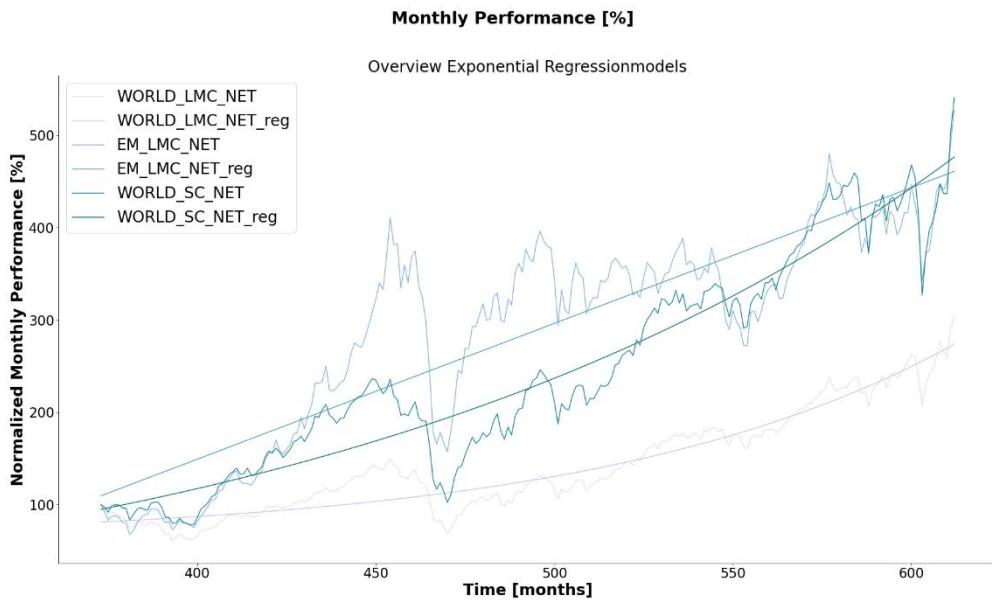
The joint plot of the regression and the original data shows the same behavior well. The general regression seems to work, the volatility itself is missing in the regression.

For EM LMC and World SC the graphs look similarly.





In the last plot, an overview over all three regression models is given.

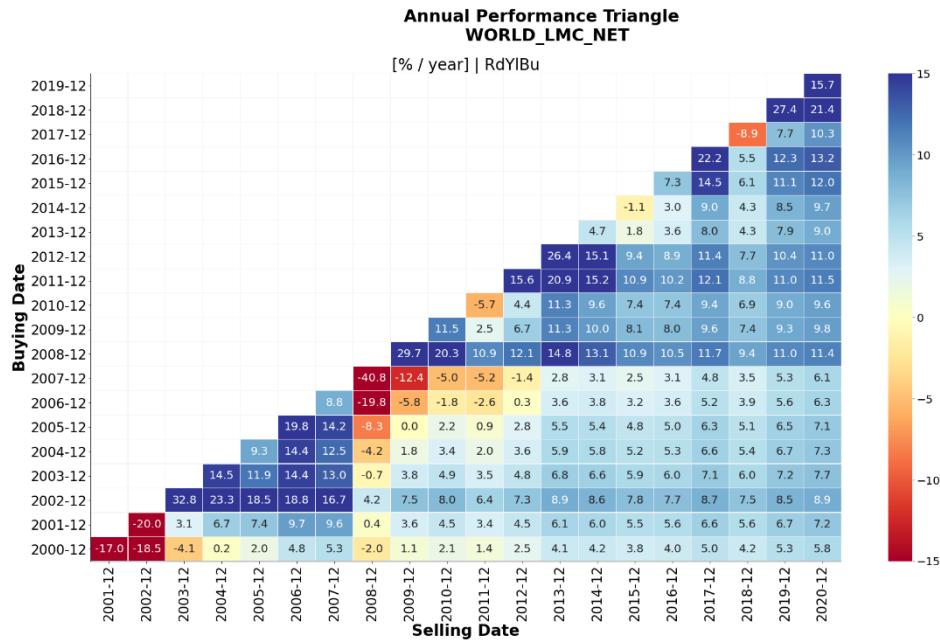


As a summary, it seems to be possible to do a regression for the general performance, but the models are rather inaccurate. Still the data might be used to detect an overpriced or underpriced market situation. Further analyses would be necessary to get more insights here. Also, it should be considered, that no-one knows, whether the data, which was used for the regression, is representative for the future, as well.

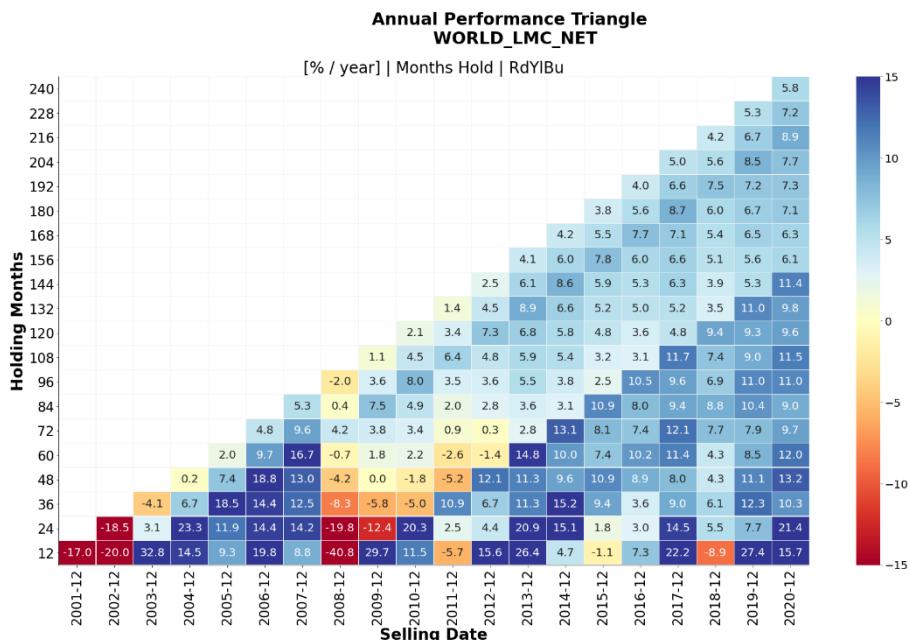
### Annual Performance Triangles

Next performance triangles are created. These triangles show the annual performance in dependency to the defined selling and buying date. Additionally, another set of performance triangles are created. These performance triangles depend on the selling date as well as the number of months, which the stock shares were held. Those triangles are created for all three indices in the same matter. Typically, the creation of heatmaps with seaborn is easy, but for these analyses the data must be prepared with iterative calculations in a first step. The calculation can be seen in the python code.

## World LMC | 12months step | start from 12/2000

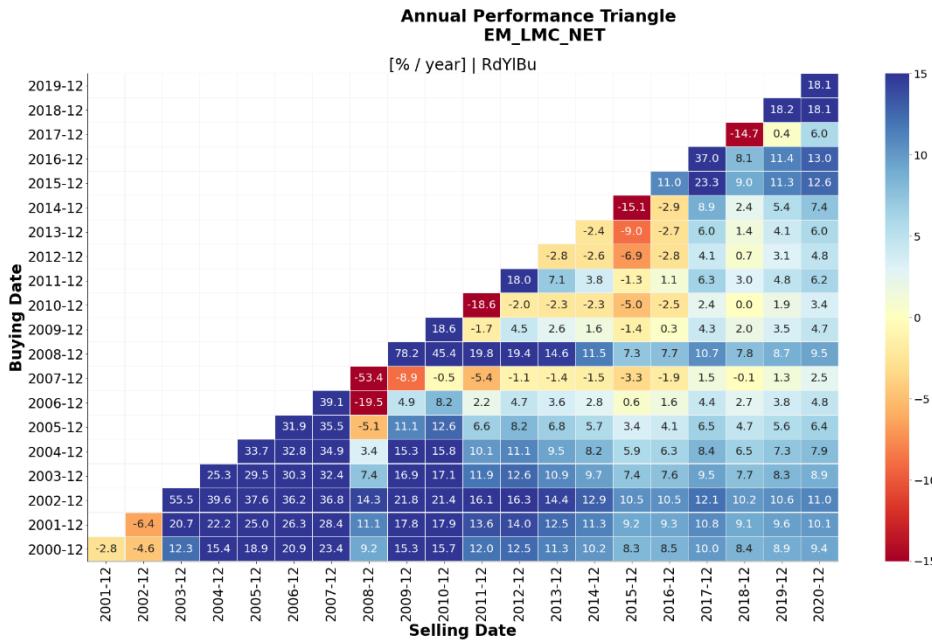


You can see, that the further on the left and top you look, the more extreme the annual performance is. The reason for this behavior is, that for short investment times the volatility has a major impact on the performance of the investment (in both, positive and negative directions). The longer the shares will be hold, the more the annual performance will narrow to values between 4 and 7 percent.

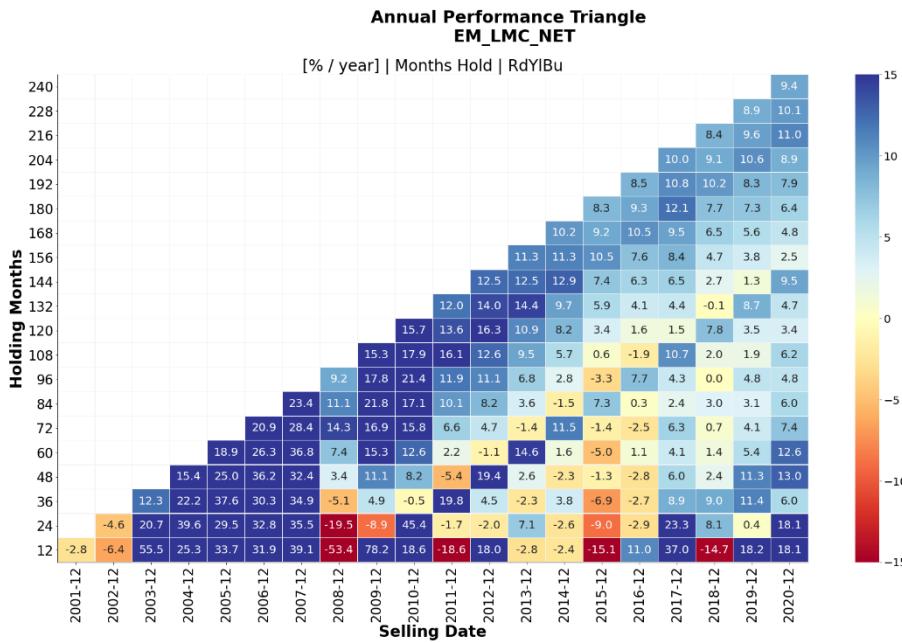


There was no loss at all during the last 20 years if you hold your shares for at least 9 years. This also represents the worst case within this timespan. In this situation you bought the ETFs on nearly the highest level before the dotcom crisis in 2001 and 2002 and sold the ETFs immediately after the economic crisis in 2008.

## EM LMC NET | 12months step | start from 12/2000

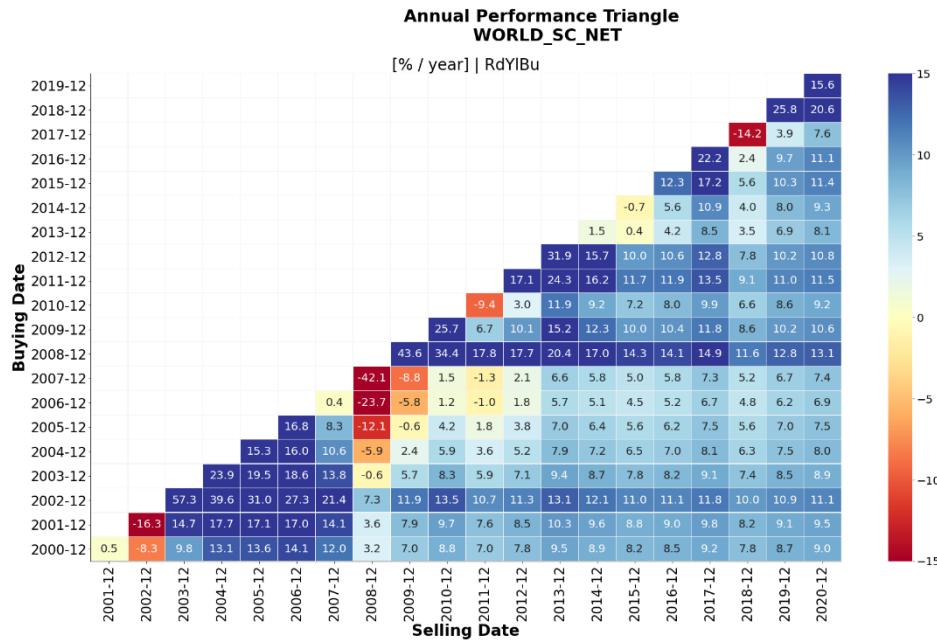


The EM LMC shows a similar behavior, but the maximums of the annual performances are higher than for the World LMC. But the potential losses are higher too.

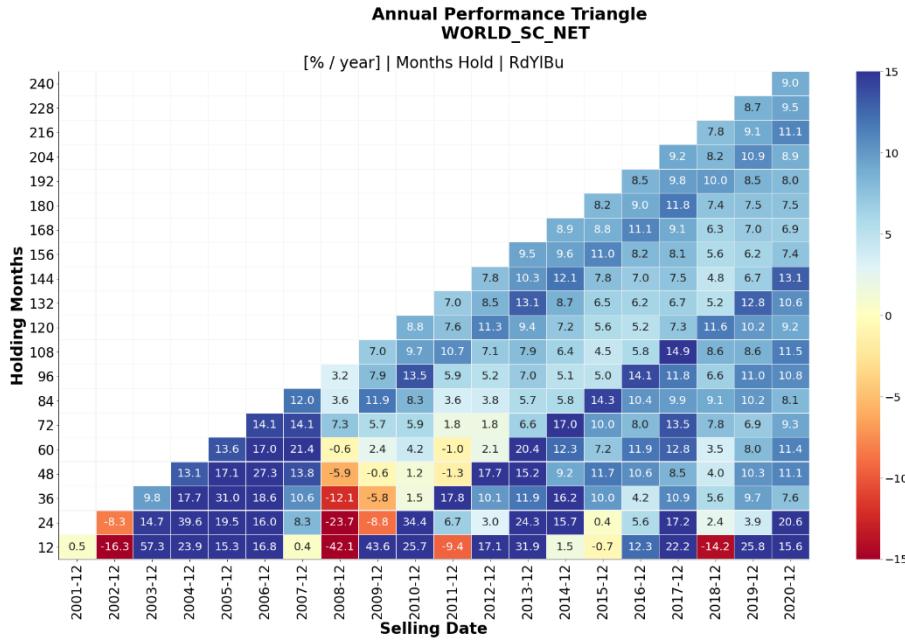


Also, the values narrow down for long investment times, but due to the in general higher risk on a slightly higher level between 6 and 10 percent.

## WORLD SC NET | 12months step | start from 12/2000

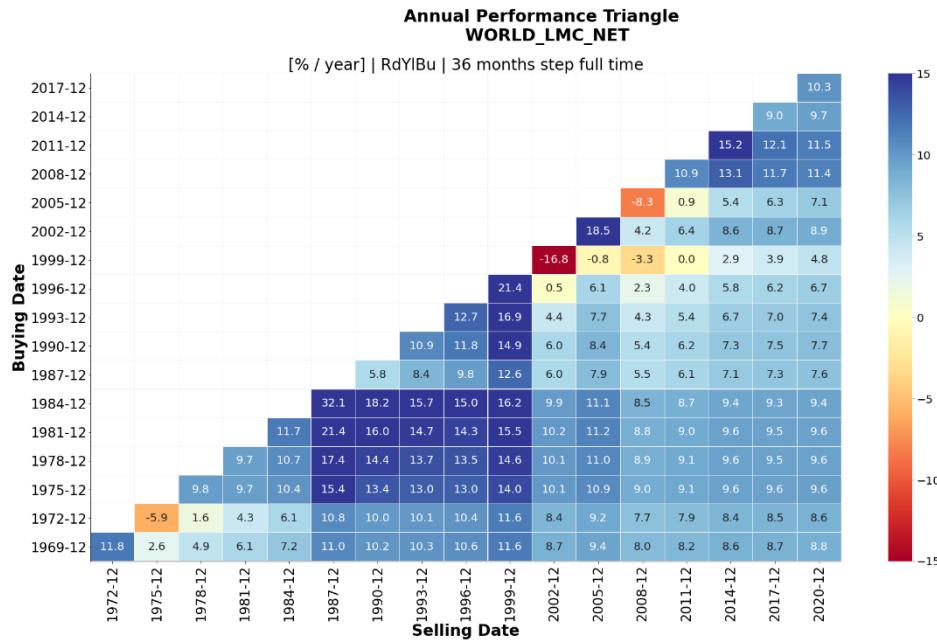


The World SC performs like the EM LMC. Still, the potential losses and short-time gains are on a lower level. This also corresponds well with the lower volatility.

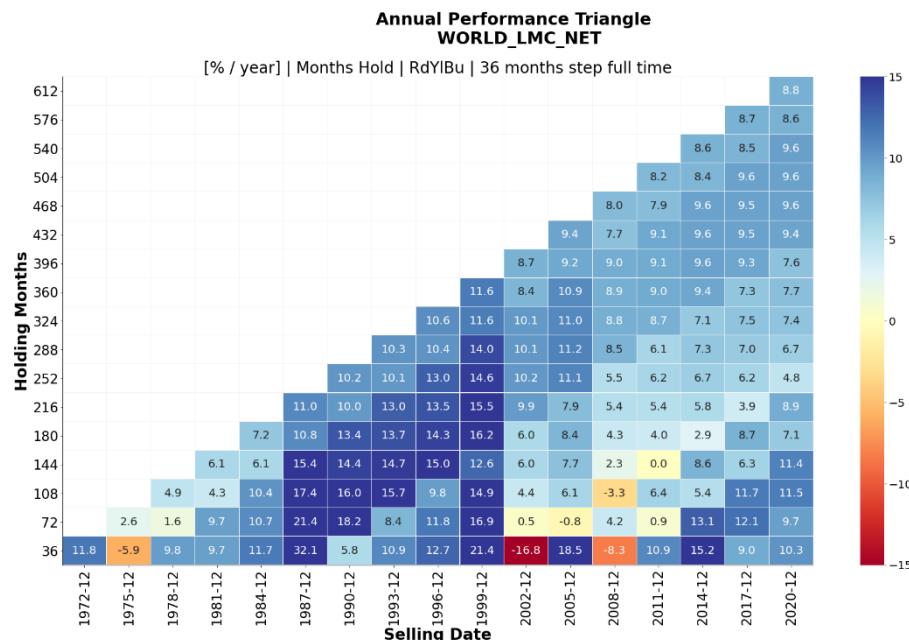


For long investments the annual performance is between 7 to 10 percent.

## World LMC | 36months step | start from 12/1969



In this graph a much longer timespan is shown. Keep it in mind, that each box represents a time step of 3 years now instead of a time step of 1 year as in the graphs before. Due to this change, the number of situations with loses decrease a lot (due to a longer minimum investment time). It can also be seen that the annual performances between 1984 and 1999 were on a higher level than for other time periods.



For exceedingly long investment periods (longer than 30 years) the average annual performance is between 7.3 and 11.6 percent. Again, the only loss occurs for a maximum investment timespan of 9 years (108 months).

## C.2. Investment Strategies - Market Timing

Next, possible investment strategies get evaluated. First the topic market timing is handled and secondly the topic rebalancing gets covered.

For the market timing strategy six basic scenarios are defined:

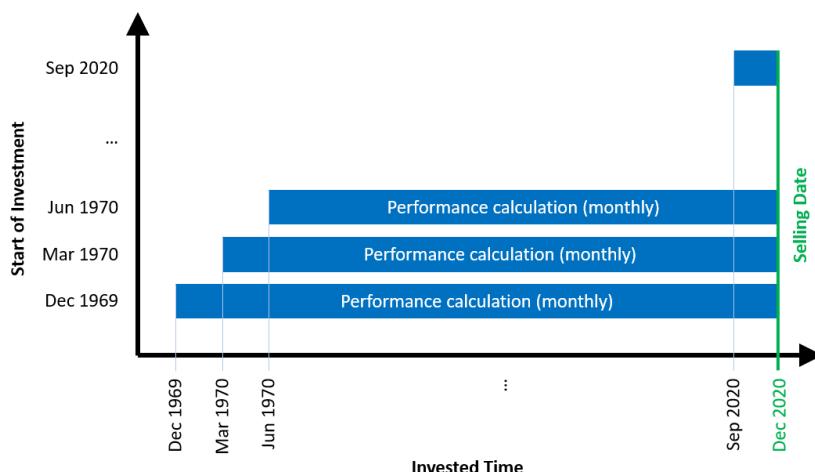
1. The capital is invested completely at the start date. Variant name: **InvStart**
2. The capital is invested equally splitted over the full covered timespan. Variant name: **InvEqual**
3. The capital is invested completely after the first occurring crash of at least 2% (compared to the previous month). Variant name: **InvCrash2**
4. The capital is invested completely after the first occurring crash of at least 5% (compared to the previous month). Variant name: **InvCrash5**
5. The capital is invested completely after the first occurring crash of at least 10% (compared to the previous month). Variant name: **InvCrash10**
6. The capital is invested completely after the first occurring crash of at least 15% (compared to the previous month). Variant name: **InvCrash15**

All investments will be done for all available time periods, always ending with the last datapoint (December 2020).

An example for InvStart:

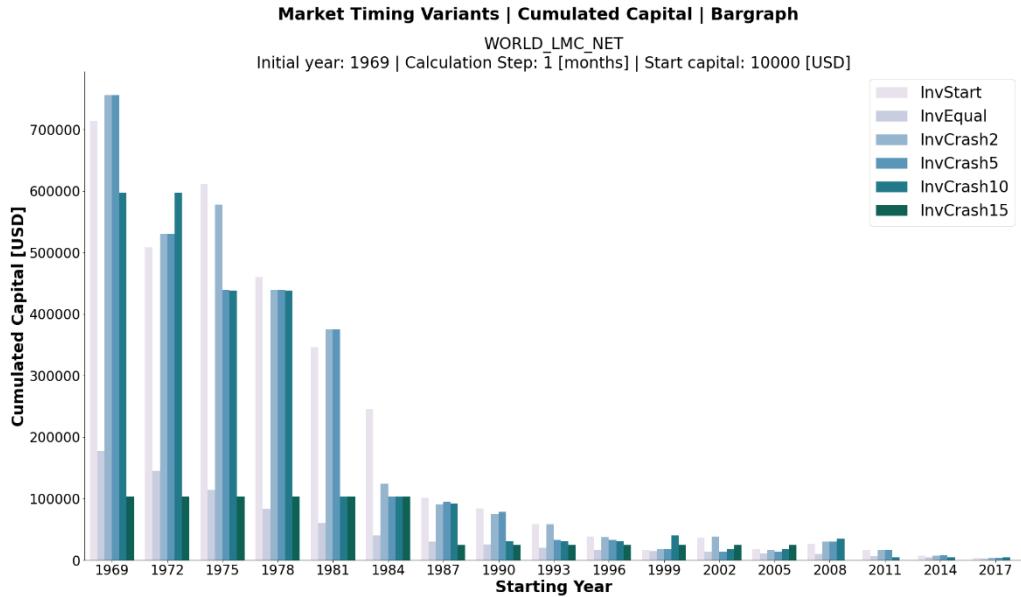
The first investment for LMC World is done in December 1969. Then the cumulated performance is calculated until December 2020. Hereby each month is considered for the performance calculation.

Next the function jumps to the next given start date, in this example 3 months later. This time the Investment capital gets invested completely in March 1970. The performance calculation is then done for each individual month until December 2020. The procedure for all the following time periods is identical. Logically the periods get shorter by the defined number of months for each step. For all investments, the cumulated performance is calculated for the last available data point in December 2020 (fictional selling date). Taxes, which are caused by the selling action are not considered, though.

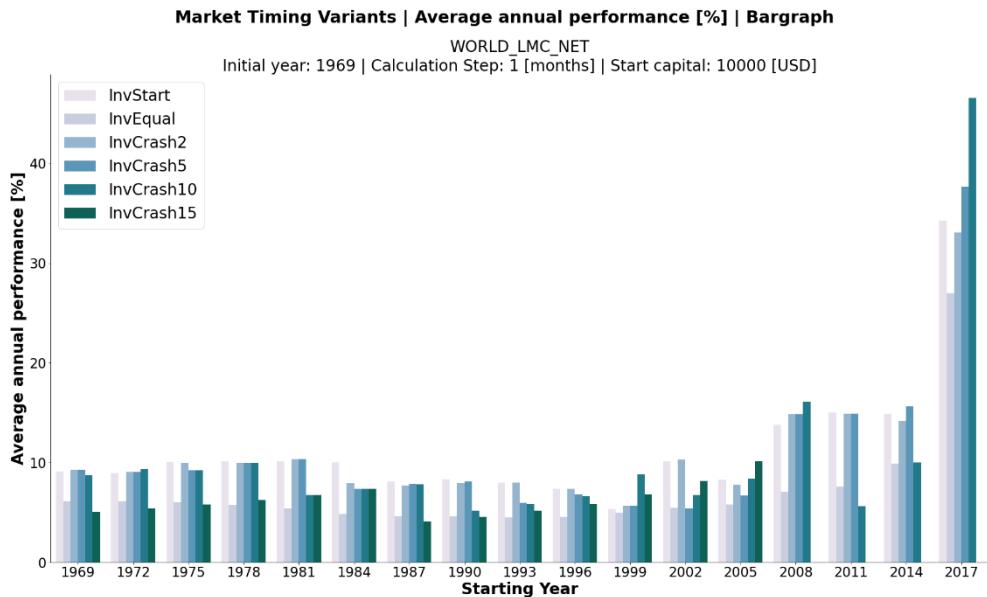


The following plot shows the cumulated capital based on the investment year and the investment variant. The invested start capital is 10,000 USD.

### WORLD LMC NET | 36months observation step | start from 12/1969



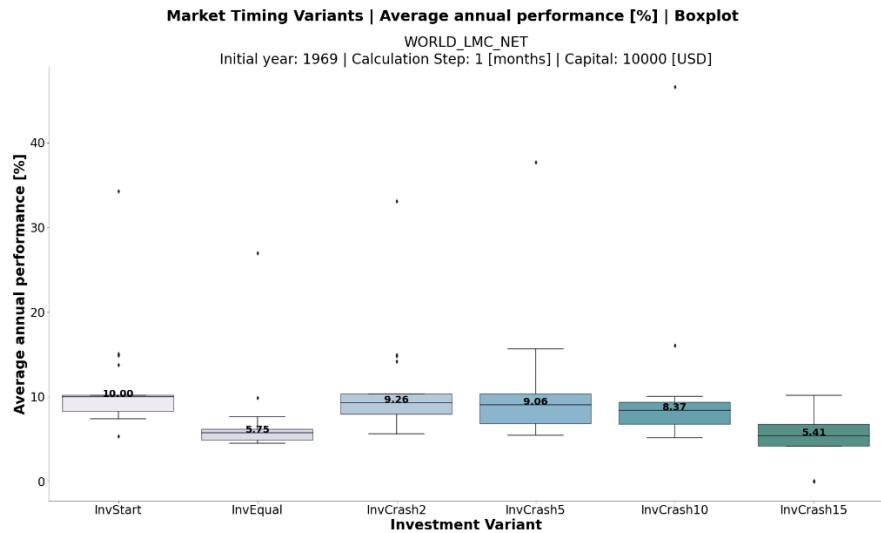
The exponential behavior of the investment over time can be seen clearly. For basically all starting years the equally distributed investment over the full timespan performs worse than the initial investment at the start. The longer the observed period, the larger the disadvantage gets.



Next, the average annual performance is plotted. It is obvious, that the longer the observed periods for the investment get, the lower the potential average annual performances get. Due to the fact, that in December 2020 the stocks were on an all-time high, the incredibly

good short-term performances are not surprising. The reader should keep in mind, that high potential short-term losses also are possible for other selling dates.

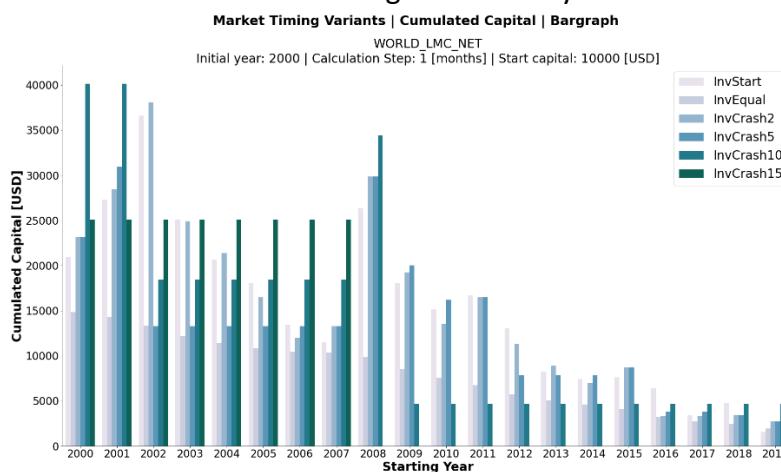
More difficult as the decision whether to invest equally distributed or immediately is the question, whether the investment after a crash is benefitable over the initial investment at the observation start. Therefor the boxplot of the average annual performance can be checked:



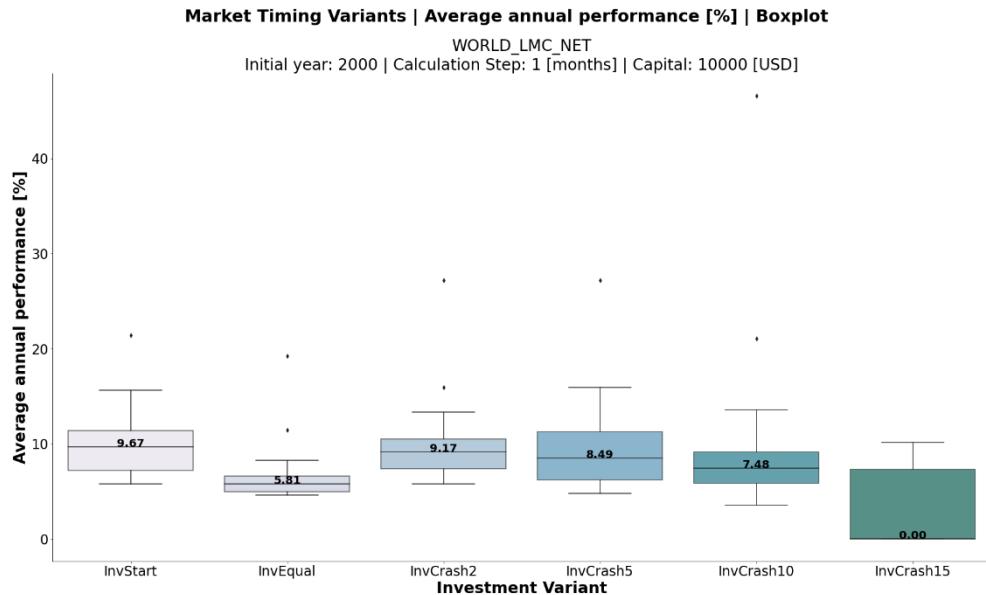
The median value for all observed average annual performances for all analyzed starting days is 10.0% for the investment directly at the start. For the crash variants these values are lower (9.3%, 9.1%, 8.4% and 5.4%). This means, that you statistically lose potential performances, if you wait with the investment. The chance to get high performances raises with the crash level. In contrast, the risk to get no or a low performance (because the capital is not invested at all or gets invested at higher index levels than in the initial investment year) raises as well.

### WORLD LMC NET | 12months observation step | start from 12/2000

Next, again the World LMC is observed. This time the data for a 12months step is observed and the initial investment start begins with the year 2000.

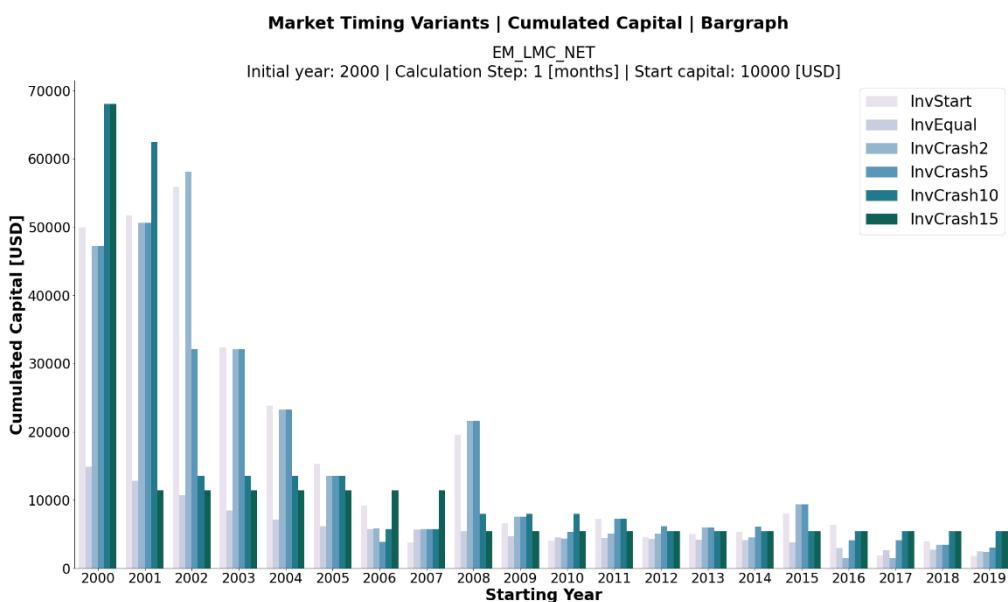


In general, Crash investment seems to perform better, because the starting date is set quite before the dotcom crash. But the following boxplot unveils, that still the statistically better choice for the World LMC is, to invest right at the start. In the years after 2009 basically no performance was achieved for the high crash levels.

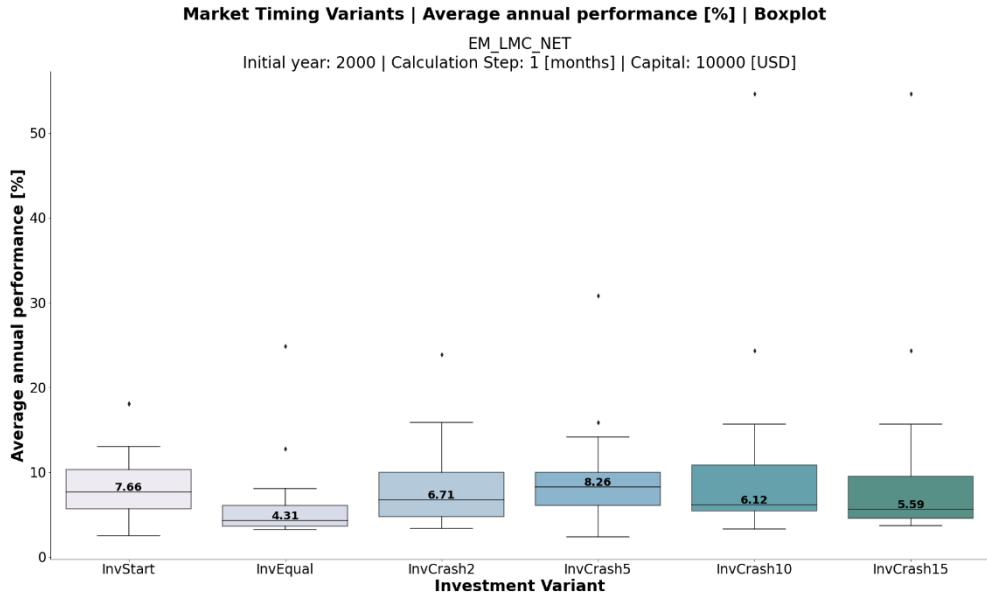


### EM LMC NET | 12months observation step | start from 12/2000

The EM LMC shows a bigger shift to the crash investing potential. For nearly all starting years the cumulated capital is on a similar level. Still there are starting years, in which the lost potential profit can be enormous (e.g. 2002), if the crash level is chosen to high.

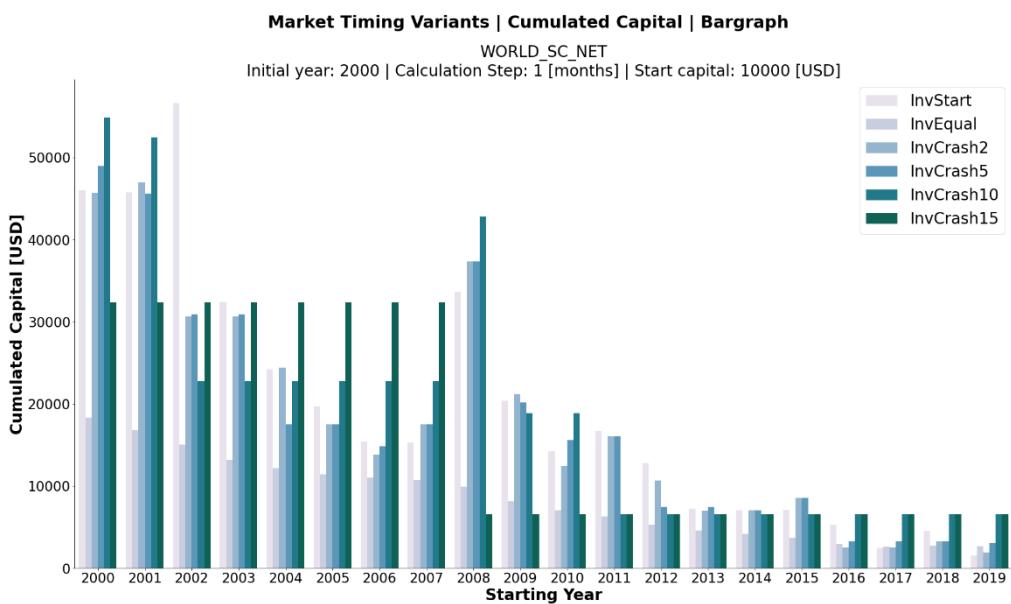


The median value for a crash level of 5% / month has the highest average annual performance of 8.3%, followed by the investment at the start with 7.7%. Again, the volatility for the crash level investments is higher. Also, the higher volatility of the EM LMC might be the reason, why crash level investing for this index might be suitable.

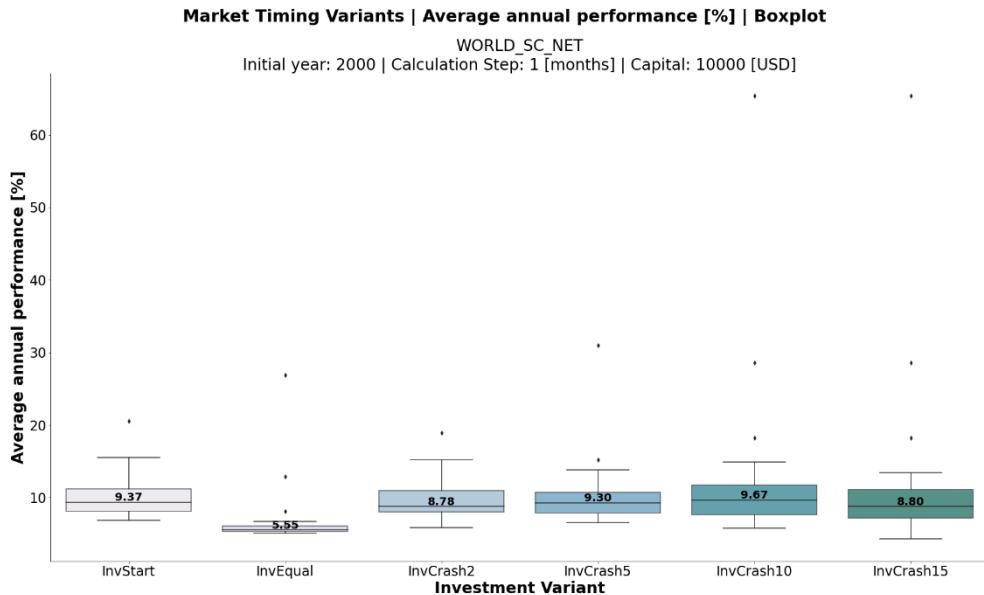


### WORLD SC NET | 12months observation step | start from 12/2000

The World SC also shows a good crash level investing performance. Again, many starting years show similar or higher cumulated capitals in comparison to the investment right at the start. But also, the lost potential profit can be significant (e.g. 2002).



The World SC also has a higher volatility than the World LMC. Therefor again the crash level investing seems to work well. But there is not too much potential for additional annual performance. With a Crash Level of 10% the median average annual performance is 9.7%, for the investment right at the start it is 9.4%.



### C.3. Investment Strategies - Rebalancing Effects

In this chapter rebalancing strategies for those three indices get analyzed. The method works like this:

First the portfolio mix gets defined. The analyzed variants are:

1. World LMC 50%, EM LMC 30%, World SC 20%
2. World LMC 80%, EM LMC 15%, World SC 5%
3. World LMC 30%, EM LMC 30%, World SC 40%
4. World LMC 15%, EM LMC 80%, World SC 5%
5. World LMC 15%, EM LMC 5%, World SC 80%

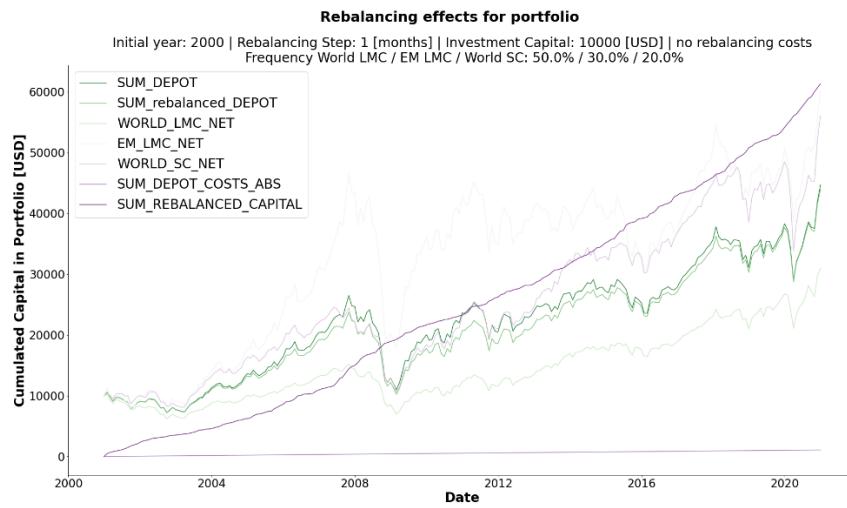
These five distributions are chosen randomly.

Secondly the rebalancing step gets defined. There will be a variant for the portfolio rebalancing for each 1, 3, 6, 12, 24, 36, 48, 60 and 120 months.

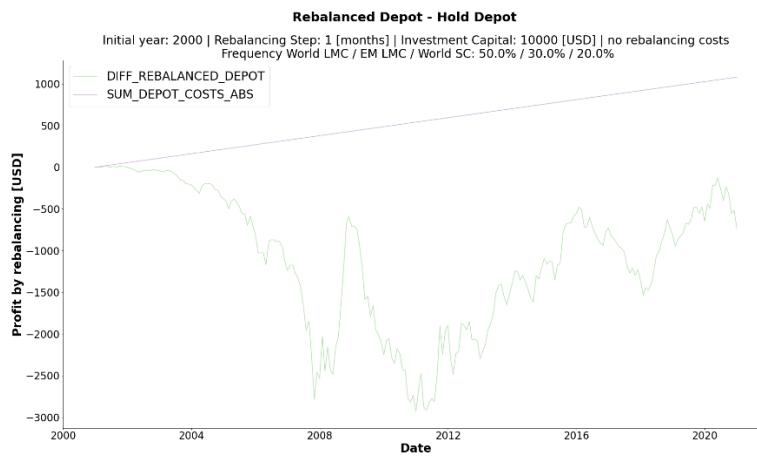
Also, there will be three calculated variants for different cost approaches for the rebalancing step. Variant 1 will not apply any costs for the rebalancing step, variant 2 will apply constantly 1.5 USD per transaction and variant 3 will apply 1.5% costs relatively to the sum of the individual transaction. If all three indices get changed by the rebalancing, three transactions are necessary. If there is any rebalancing at all, at least two transactions are necessary. If you lower one investment there will be an increase of at least another investment.

The algorithm does the performance calculation like before monthly. All three indices are calculated individually. The sum of the depot without rebalancing will be the sum of all individual start capital performances. Additionally, for each rebalancing step the algorithm evaluates whether the portfolio fits to the initial target mix, which basically never happens due to individual index performances. Therefor the algorithm will then increase or lower the capital of each index in such a manner, so that it fits to the initial target mix. Each time the relative or absolute costs will be added separately.

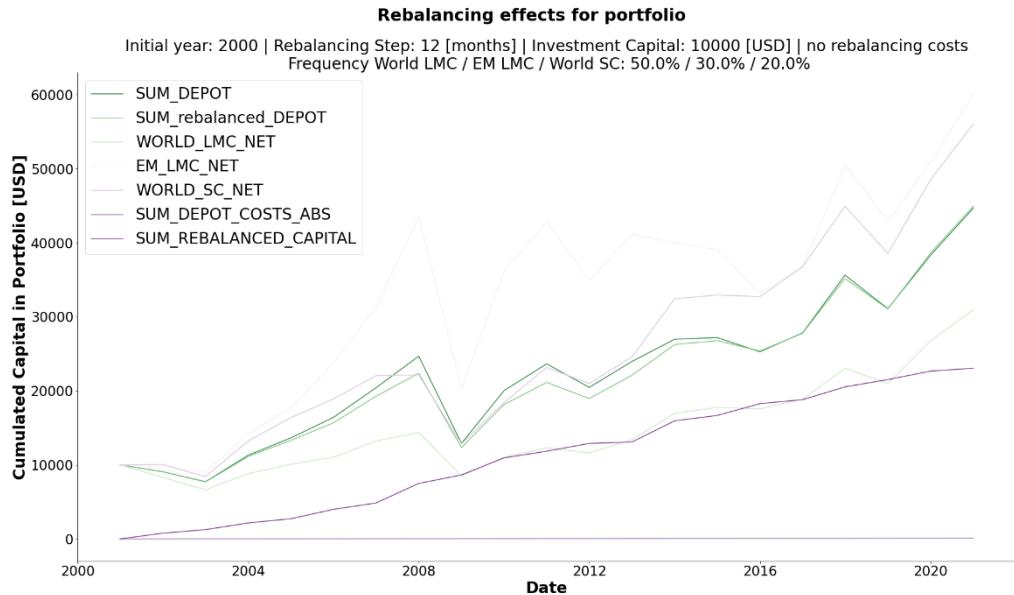
In the next graph you can see results, which explain the performance of those rebalancing strategies.



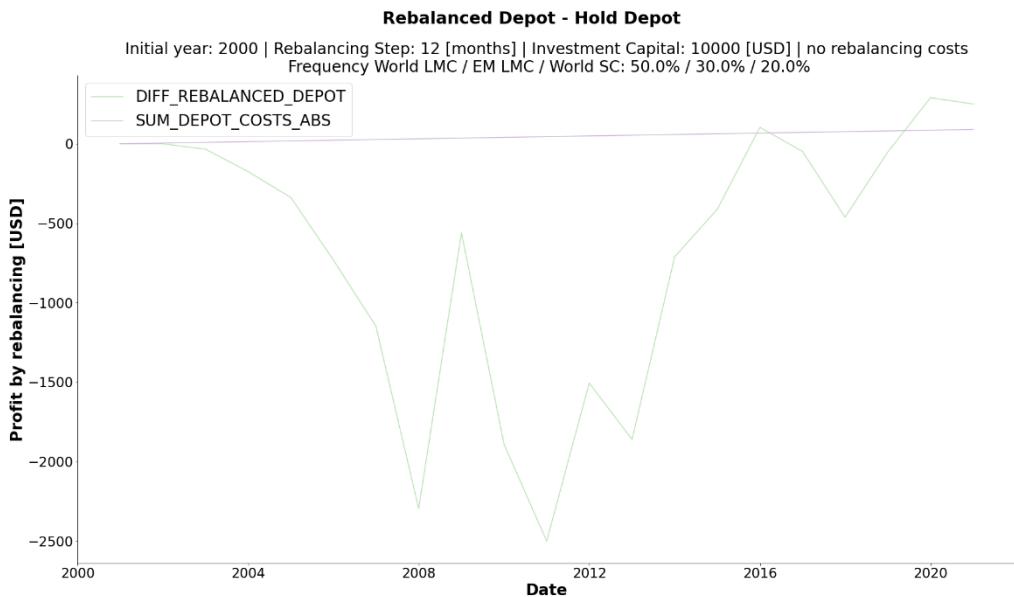
This graph shows the original accumulated capital for the World LMC, EM LMC and World SC in light colors for a starting capital of 10,000 USD. More important are the two dark green lines. SUM\_DEPOT shows the accumulated capital of the portfolio with the initial portfolio mix and no rebalancing at all. Obviously, the accumulated capital in the end is between the three original indices. The line SUM\_rebalanced\_DEPOT shows the accumulated capital of the rebalanced Depot. In the subtitle the rebalancing step is mentioned, which shows, that this depot was rebalanced once per month. The purple lines show the cost and totally "moved" capital by transactions. The costs obviously stay at 0 USD for this variant, but the money, which was rebalanced gets to significant values after a while.



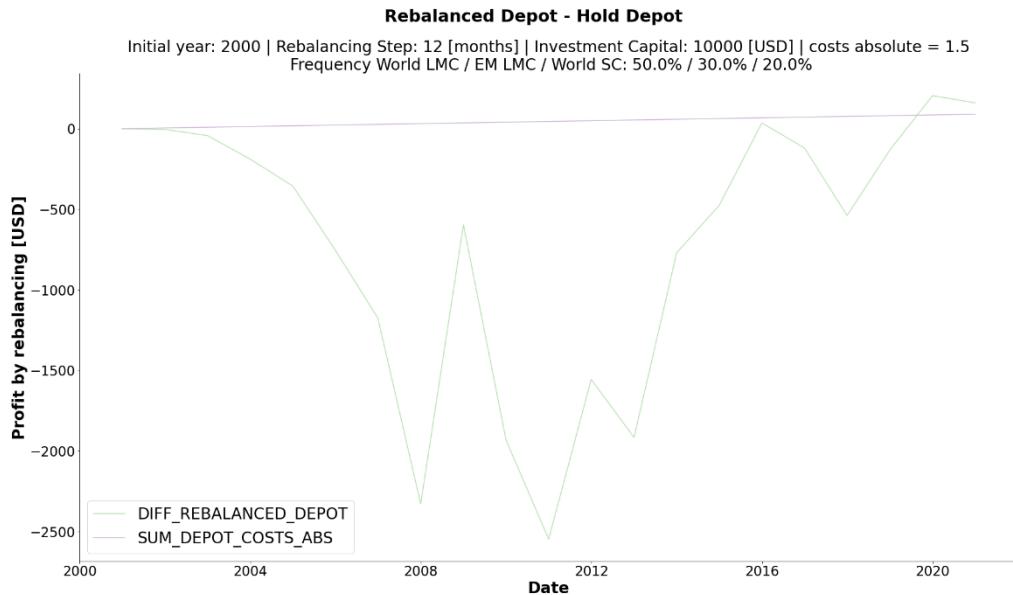
This graph shows the difference of the depot with the original mix (no rebalancing steps) and the depot, which was actively rebalanced. Obviously, the green line starts with 0, but it falls to negative values soon and never reaches positive values. This means, that for this variant the rebalancing causes loses in comparison to no rebalancing. The investor should have avoided this investment strategy.



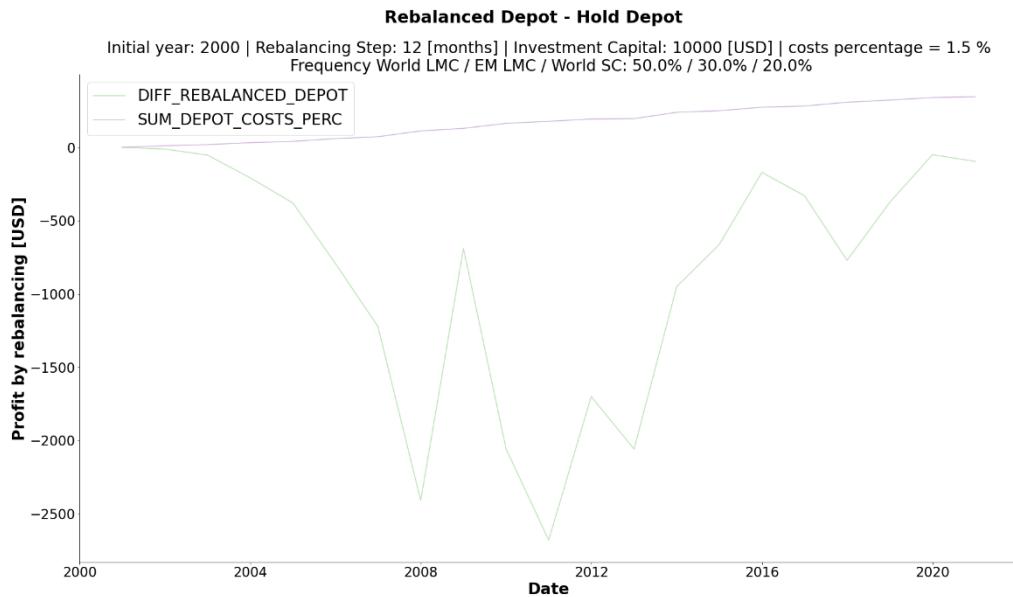
If the rebalancing step gets changed to 12 months, the situation improves. The handled capital gets lower and the end capital is higher than the benchmark. It should be considered that no costs are applied for this variant. Also, there were nearly 15 years, in which the rebalanced portfolio does worse in comparison to the benchmark.



If absolute costs are considered as well, the situation changes slightly.



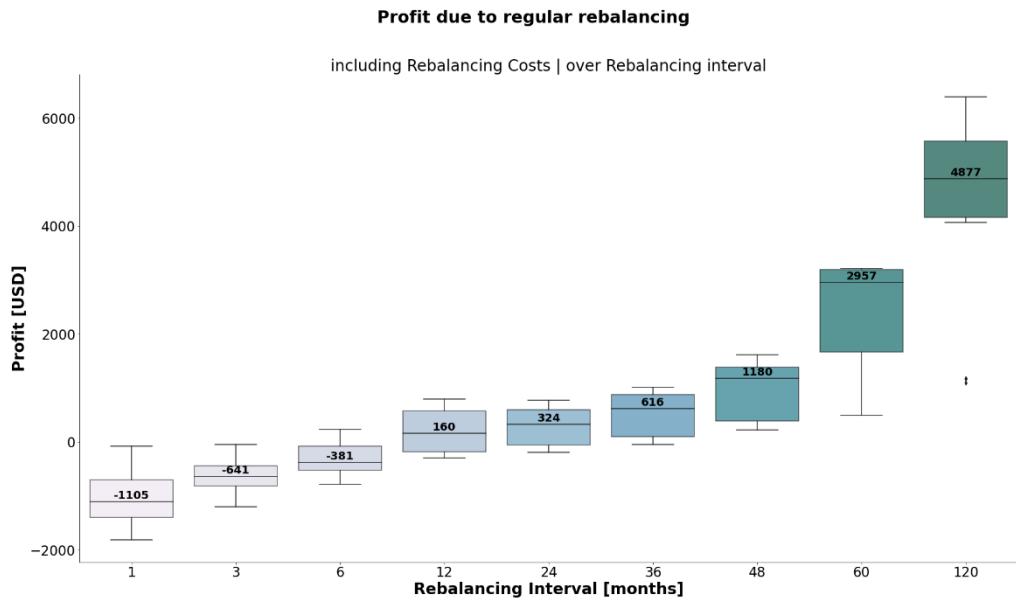
The rebalanced portfolio is on a slightly lower level (check the year 2016), but in general the influence of the absolute costs is low for this amount of invested capital. The costs are more relevant if percentage costs are applied.



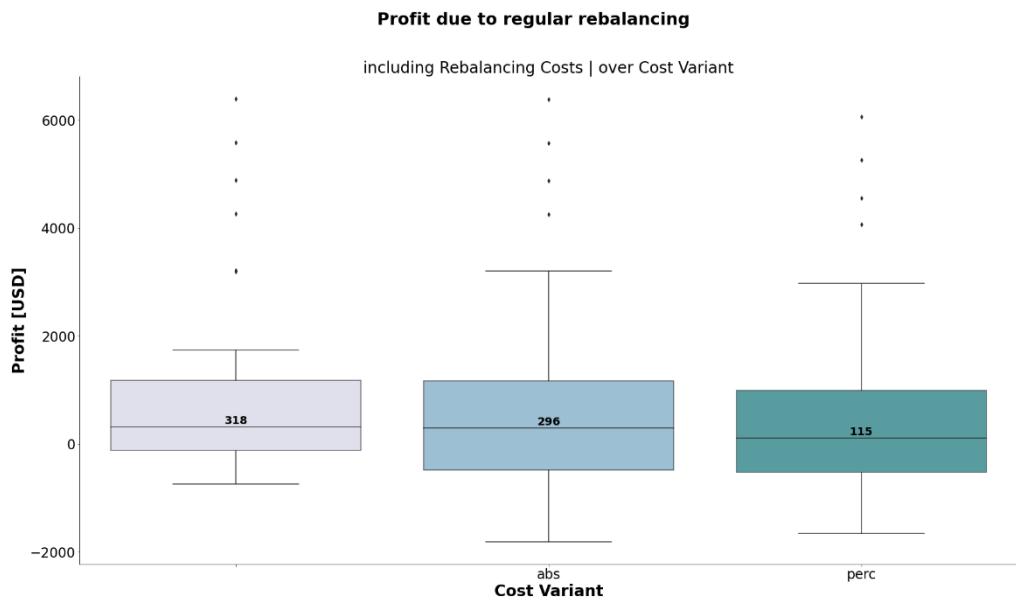
For this variant, the losses are even higher and cause a loss in the end.

Such kind of plots exist for every considered variant. To easier the evaluation, a boxplot is created out of those variants. The boxplot shows the Profit in USD caused by the rebalancing actions. Keep it in mind, that these boxplots include all mixed portfolios and all cost variants at this point. Despite that, one trend can be seen fairly good: Rebalancing causes losses, if it is done too often. The longer the rebalancing steps are, the higher is the chance to increase

the portfolio profit. Still the effect is rather low, if the overall high profit of about 35,000 USD for this example is considered.

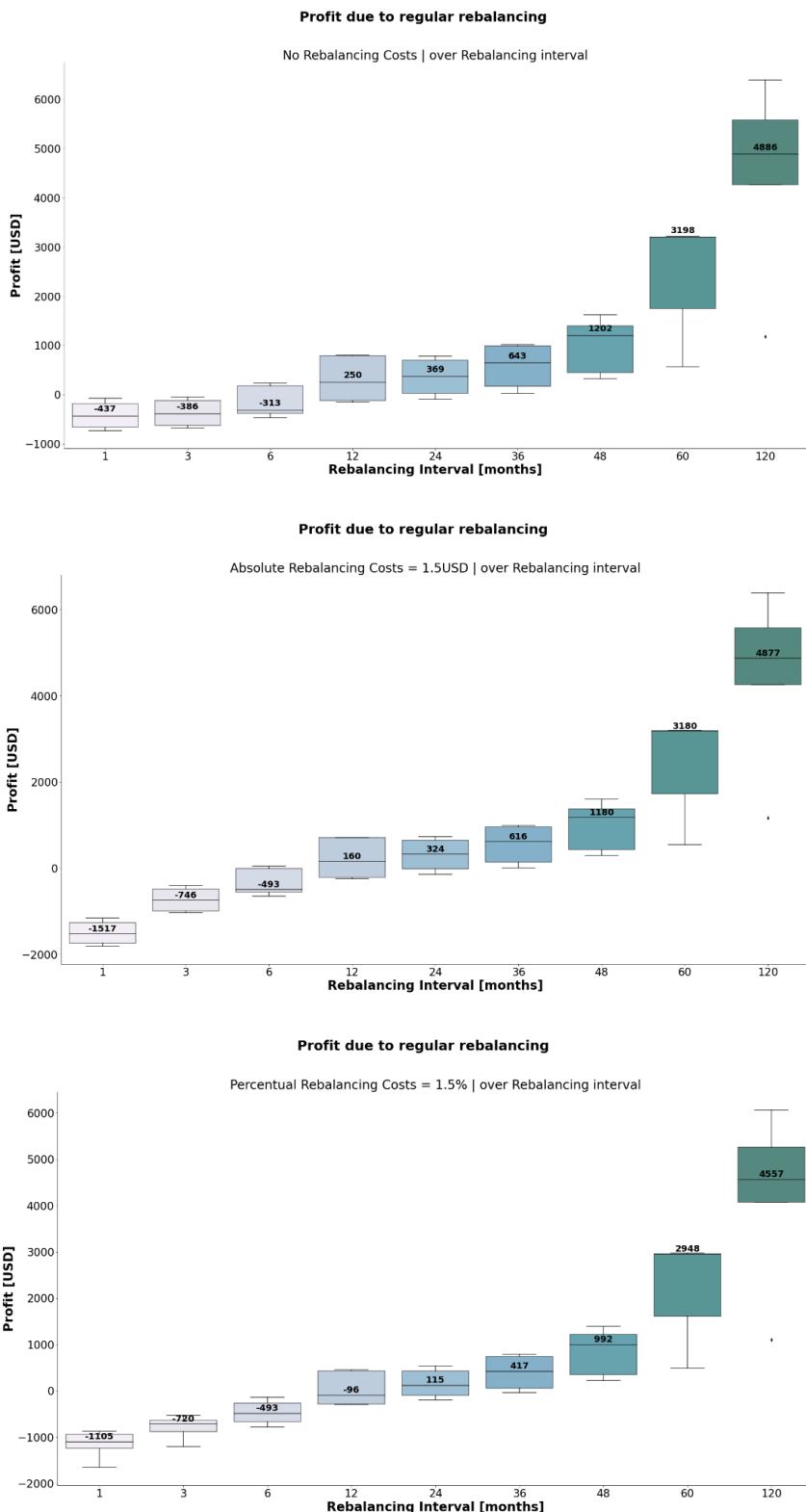


The next boxplot also includes all variants but shows the cost variants over the x-axis.



The median profit of all evaluated variants is similar for the absolute and no transaction cost variants. The median of the percentage variant is much lower. This situation can be completely different if different investment capitals are chosen in the beginning.

Next, there will be a look on the rebalancing intervals for all portfolio variants. But this time the data will be filtered for the three available cost variants of No Rebalancing Costs, Absolute Rebalancing Costs of 1.50USD and Percentage Rebalancing Costs of 1.5%.



In general, the previous behavior gets confirmed. The variants of No costs and absolute costs are similar. The variant with 1.5% costs shows the same trend, but on a lower level. Therefore the rebalancing should be done even more rarely for this variant.

## D. Discussion

A lot of information is given in the previous results chapter. Within this chapter the methods and results get discussed in a general manner.

### Regression & simple Model building

Unfortunately, there are only few parameters available and the behavior obviously is not linear. This causes a more difficult model building. For this case, the exponential function was used and the `curve_fit` function applied. Due to the nonlinear behavior, it is difficult to rate, how good the model works for the data. Subjectively the regression covers the general behavior, but due to the short time periods, missing accuracy measurements etc. it is not recommended to use this regression for extrapolation purposes at this point. Also, it does not cover crashes in a good way, which clearly can be seen in the created Jointplots.

To do proper regressions in a first step more parameters about the data should be collected (e.g. about companies, countries, market capitalization of the indices, etc.).

### Annual Performance Triangle

The annual performance triangles give good insights into the available data. Also, it can be used to derive rather general action strategies for the investor. It gives good indications on how long the shares should be hold and what general annual performance might be expected.

### Investment Strategies - Market Timing

The created boxplots unveil, that statistically the investor should not wait with the investment if there is no additional information about the market situation. But the more volatile the index, the higher the chances are, that crash-investment (after defined crash-levels) might work.

### Investment Strategies - Rebalancing

This report shows that rebalancing can have a positive influence, even for assets of the same type (stock shares). But the rebalancing should be done rarely (3 - 5 years and longer) and only if the transaction costs are low.

## E. Conclusion

The conclusion for investors is simple. Buy the target portfolio and prepare yourself to hold it for at least 10 years. The target portfolio benefits from the higher volatility of the EM LMC and World SC and therefor the overall portfolio performance can benefit. Waiting with the investment to do proper market timing statistically causes disadvantages. Investing in equally spaced amounts over long investment timespans results by far in the worst performance. Rebalancing can help, if it is done rarely (rather some years than months) and with low transaction costs.

Within the observed timespans index investing performed well for nearly all investigated variants. Keep in mind, that the EM LMC and World SC timespan only covers rather short timespan since 2000.

Predicting crashes (at least without further data) is not possible. Doing a regression works, but it is difficult to evaluate the accuracy for this non-linear model. Therefor it is hard to decide, whether the prediction was too conservative or too optimistic.

In general, the results of this evaluation support the known investment strategy "Buy and Hold".

Or in the words of André Kostolany:

*If you want to make money with shares, you don't do it with your head or your stomach, but with your bottom, by sitting on it for as long as possible.*

## F. References

- [1] [MSCI World - Wikipedia article](#)
- [2] [MSCI Index Solutions – msci.com](#)
- [3] [Tracking Error - Wikipedia article](#)
- [4] [Exchange-traded fund – Wikipedia article](#)
- [5] [Total expense ratio – Wikipedia article](#)
- [6] [Rebalancing Investments- Wikipedia article](#)
- [7] [Spyder IDE Webpage - spyder-ide.org](#)
- [8] [Development platform - github.com](#)
- [9] [Package for scientific computing with Python - numpy.org](#)
- [10] [Fast, powerful, flexible and easy to use open source data analysis and manipulation tool - pandas.pydata.org](#)
- [11] [Python data visualization library based on matplotlib - seaborn.pydata.org](#)
- [12] [Comprehensive library for creating static, animated, and interactive visualizations in Python - matplotlib.org](#)
- [13] [SciPy is a Python-based ecosystem of open-source software for mathematics, science and engineering - scipy.org](#)