Out-of-Sample Performance of Modern Portfolio Strategies

Master’s Thesis
in the Department of Economics of the University of Konstanz

By: Marlon Haudek
Litzelstetter Straße 31C
78467 Konstanz

Period of completion: 24th April 2023 until 24th August 2023

1st assessor: Prof. Marcel Fischer
2nd assessor: Prof. Axel Kind

Konstanz, 11th August 2023
Abstract

This paper evaluates the out-of-sample performance of four modern portfolio strategies, which are the minimum-variance portfolio, the Jorion’s Bayes-Stein minimum-variance portfolio, the 1/N portfolio, and the Equity Market Neutral portfolio implemented by a Convolutional Neural Network. The out-of-sample performance is tested on two time horizons (2010-2019 and 2010-2022) on the German stock market using Return, Volatility, Sharpe Ratio, and Drawdown as performance measurements. The empirical results show the minimum-variance portfolio has on average the lowest annual volatility and max drawdown for both time horizons. Whereas the Equity Market Neutral portfolio has the highest average annual return and Sharpe Ratio in both time horizons. However, these results need to be verified in further investigations, for example adding transaction costs to each portfolio strategy, which can result in drastically different performance.
# Table of Content

List of Figures .............................................................................................................ii  
List of abbreviations ....................................................................................................iii  
1. Introduction ..............................................................................................................1  
2. Theoretical Framework ............................................................................................4  
   2.1 Minimum-variance portfolio (Plug-in approach) ..............................................4  
   2.2 Jorion’s Bayes-Stein estimator approach for the minimum-variance portfolio ...11  
   2.3 1/N portfolio strategy with rebalancing .............................................................14  
   2.3 Equity Market Neutral portfolio implemented by a CNN ..................................16  
3. Methodology and Data ............................................................................................30  
   3.1 Methodology .......................................................................................................30  
   3.2 Data ....................................................................................................................34  
4. Empirical findings ..................................................................................................39  
   4.1 Empirical findings for the time horizon 2010 – 2019 .........................................39  
   4.2 Empirical findings for the time horizon 2010 – 2022 .........................................43  
   4.3 Comparison of results with existing literature ....................................................47  
5. Conclusion ..............................................................................................................49  
Reference List ..............................................................................................................54
List of Figures

Figure 1: Efficient Frontier .................................................................6
Figure 2: Position of minimum-variance portfolio on the efficient frontier ........8
Figure 3: Components of a single node .................................................19
Figure 4: Linear activation function.......................................................19
Figure 5: Sigmoid activation function....................................................20
Figure 6: Hyperbolic tangent activation function ....................................21
Figure 7: Rectified Linear Unit activation function ...................................22
Figure 8: Architecture of a neural network .............................................22
Figure 9: Advantages of having hidden layers within a neural network .......23
Figure 10: Functional concept of how a convolution layer works .............25
Figure 11: Flowchart of the EMN portfolio strategy ..................................28
Figure 12: Rolling window approach ....................................................31
Figure 13: Cumulative return chart of the DAX for the time horizon 2010 - 2019 ........35
Figure 14: GDP development of the German economy ..........................36
Figure 15: Development of unemployment rate in Germany ....................36
Figure 16: Development of ECB’s funds rate ........................................37
Figure 17: Cumulative return chart of the DAX 2010 – 2020 ....................38
Figure 18: Cumulative returns of portfolio strategies & DAX 2010 - 2019 ..........40
Figure 19: Cumulative returns of portfolio strategies & DAX 2010 – 2019 .........45
List of abbreviations

EMH = Efficient Market Hypothesis

EMN = Equity Market Neutral

CNN = Convolutional Neural Network

DAX = German stock market index (*benchmark*)

1/N = 1/N portfolio strategy with rebalancing

min = minimum-variance portfolio strategy using a “Plug-in approach”

bs-min = Jorion’s Bayes-Stein estimator approach for minimum-variance portfolio

emn-cnn = Equity Market Neutral portfolio strategy implemented by a convolutional neural network
1. Introduction

Nowadays, due to the low-interest rate income received on bank accounts, the investment topic becomes more and more relevant to everybody who wants to earn yields on her/his capital. For most (retail) investors the starting point of their investment journey is the equity market. As enthusiastic as they are when deciding to make more out of their money, as quickly they are overwhelmed by the abundance of portfolio strategies the economic literature offers. Thereby not only the invention of new portfolio strategies is important, but also the performance testing in terms of out-of-sample performance on different markets and different economic environments. For practitioners, it is of no use if a portfolio strategy only delivers good results theoretically and on paper. At the end of the day, a portfolio strategy must optimally bring positive money into investors' coffers in every market and in every market situation. This is the exact point where the goal of this paper is.

The goal of this paper is to evaluate the out-of-sample performance of conceptionally different portfolio strategies, such that investors know (i) which different approaches exist in the current economic literature and (ii) how each portfolio optimization concept performs according to different performance measurements in direct comparison to each other and to a benchmark. Let me briefly review which conceptionally different portfolio strategies are chosen in this paper and why.

Today a huge variety of portfolio optimization strategies exist in the literature. However, this was not always the case, a few decades ago only a handful of portfolio optimization strategies existed, and even fewer were scientifically investigated. In the 1950’s an economist named Harry Markowitz invented a groundbreaking new theory called Modern Portfolio Theory (MPT) on which many of today’s portfolio optimization strategies are based on. The quint essence of MPT is, that an investor can maximize her expected return for a given risk by investing in stocks that are as less correlated as possible. Because of the great impact on the research for portfolio optimization in this paper, one basic portfolio optimization strategy which has its roots in the MPT is implemented. Over time much research has been done around the MPT and many variations are invented. One very significant paper that investigates the out-of-sample performance of different portfolio strategies, including MPT portfolio strategies, is the paper by DeMiguel, Garlappi, & Uppal (2009). They found that the minimum-variance portfolio exhibits great performance compared to other portfolio strategies in terms of
Sharpe Ratio, Turnover, and Certainty Equivalent over multiple stock market portfolios in a time horizon between 1963 and 2004. This result prompts the inclusion of the minimum-variance portfolio as the first portfolio strategy for out-of-sample performance investigation in this paper. However, one major drawback of the minimum-variance portfolio is that it suffers from an estimation error, which can lead to sub-optimal wealth allocation for the portfolio resulting in bad out-of-sample performance. To cope with the estimation error in the minimum-variance portfolio a Bayesian approach can be used. In the paper by DeMiguel et al. (2009) the Bayes-Stein approach introduced by Jorion (1986) achieves on average also good out-of-sample performance in terms of Sharpe Ratio, Turnover, and Certainty Equivalent. Therefore, it is also worthwhile to implement the Bayes-Stein approach by Jorion (1986) as the second portfolio strategy in this paper. Since the minimum-variance and Jorion’s Bayes-Stein approach are settled in the MPT, this paper is enriched by a third portfolio optimization concept, that is more naively oriented. The paper by DeMiguel et al. (2009) proposes a 1/N portfolio strategy which, according to the results of the paper, outperforms on average all other investigated portfolio strategies in their paper over all performance measurements and different datasets. Such a result gives the reason to conduct further research around the out-of-sample performance of the 1/N strategy as the third portfolio optimization strategy in this paper. Since the introduction of the minimum-variance portfolio and the 1/N portfolio has been a while, this paper aims to close the gap between the past and today. In times of computer hardware faster than ever, the unstoppable rise of neural networks is foreseeable. Therefore, this paper investigates into a fourth portfolio optimization strategy, which is a Convolutional Neural Network (CNN) that implements the Equity Market Neutral (EMN) strategy presented by Wu, Syu, Lin, & Ho (2021). The reason why exactly this strategy is chosen is based on the findings from the paper by Gunjan & Bhattacharyya (2023), which compares several neural networks with respect to their out-of-sample performance. The EMN implemented by a CNN performs best on average in terms of Sharpe Ratio.

In summation, this paper offers the reader four portfolio optimization strategies, of which three of them are conceptionally different. Another contribution to the existing literature is that the portfolio strategies are applied to the German stock market index DAX on a weekly timeframe. To test the out-of-sample performance of the portfolio strategies two different time horizons are chosen. The first one is from 2010 until 2019 and the second one is from 2010 until 2022. The shorter time horizon is chosen because
it corresponds to a period of stable economic growth of the German economy and therefore a steady stock market growth without serious corrections or bear markets. The goal of the shorter time horizon is to evaluate the performance of the four portfolio strategies under good economic conditions. The longer time horizon, 2010 - 2022, inherent a change in the market environment happening in the years 2020 until 2022 because of COVID-19. The stock market responds to it with a higher stock market volatility with serious drops. The goal of the longer time horizon is to evaluate how the portfolio strategies perform under the condition of a changing market environment.

Thereby the paper aims to answer the following research questions for the corresponding two time horizons:

(i) Do any of the portfolio strategies presented in this paper achieve a higher annual return than the DAX as a benchmark?

(ii) Do any of the portfolio strategies presented in this paper achieve a lower annual volatility than the DAX as a benchmark?

(iii) Do any of the portfolio strategies presented in this paper achieve a higher annual Sharpe Ratio than the DAX as a benchmark?

(iv) Do any of the portfolio strategies presented in this paper achieve a lower Drawdown than the DAX as a benchmark?

To evaluate the performance measurements a rolling window approach is used in order to calculate the optimal portfolio weights for each portfolio strategy. As an in-sample period length used by the rolling window approach 60 periods are considered. The consideration is based on the paper by DeMiguel et al. (2009) which justifies that it is a reasonable in-sample length for this methodology framework.

Turning now toward the empirical findings. This paper delivers promising results namely, two portfolio strategies are able to statistically outperform the DAX as a benchmark in several performance measurements. The min portfolio is able to reach a statistically lower average annual volatility of 0.1068 and 0.1242 for the respective time horizons 2010 – 2019 and 2010 – 2022. Whereas the DAX has for, the same time horizons, a volatility of 0.1449 and 0.1517, respectively. Also, the min portfolio achieves the lowest max drawdown of all portfolio strategies with 0.2049 for the 2010 – 2019 time horizon and 0.2810 for the 2010 – 2022 time horizon. The second portfolio strategy,
which achieves statistical outperformance over the DAX and all other portfolio strategies, is the emn-cnn portfolio in terms of average annual return and Sharpe Ratio. The emn-cnn portfolio exhibits an average annual return of 0.2132 and 0.2213, while the DAX only achieves an average annual return of 0.0969 and 0.0860 for the shorter and longer time horizon, respectively. In terms of Sharpe Ratio, the emn-cnn portfolio stands out with an average annual Sharpe Ratio of 1.7929 compared to the DAX with 0.5611 in the time horizon of 2010 – 2019. For the time horizon of 2010 – 2022 the emn-cnn portfolio outperforms with a Sharpe Ratio of 1.6563 while the DAX exhibits a Sharpe Ratio of 0.5276. Furthermore, the emn-cnn portfolio outperforms all other portfolio strategies in terms of return and Sharpe Ratio in this paper.

The rest of the paper is structured as follows. Section 2 provides all necessary background information to the reader regarding the minimum-variance portfolio optimization, the Jorion’s Bayes-Stein minimum-variance portfolio optimization, the I/N portfolio optimization, and the EMN portfolio optimization implemented by a CNN. Section 3 describes in the first part the methodology, i.e. the rolling window approach used by this paper. Also, it states the corresponding formulas for the performance measurements used. In the second part of section 3 descriptive statistics regarding the data used in this paper enriched by macroeconomic information about the German economy for the two time horizons are provided. The empirical findings with reasonable explanations and comparison to existing literature can be found in section 4. The last section concludes the paper and its findings. In addition, critical points and further investigation possibilities are mentioned.

2. Theoretical Framework

2.1 Minimum-variance portfolio (Plug-in approach)

The concept of the minimum-variance portfolio strategy can be assigned to the Modern Portfolio Theory, which was invented by Markowitz in 1952 and published in the paper “Portfolio Selection”. Since then, the MPT developed itself into a well-researched concept in the literature over the past decades. Not only its widespread use by the literature in different markets and market conditions but also the decent number of variations and extensions that economists came up with for improving the MPT and its
portfolios. Because of the sheer impact of the MPT has had back in the days and still has nowadays on economic research, it is almost self-explanatory that this paper, which investigates the out-of-sample performance of modern portfolio strategies includes a portfolio optimization strategy from this concept.

However, before we dive into the mathematics and the reason why exactly the minimum-variance portfolio is investigated with respect to its out-of-sample performance in this paper, necessary background information about the MPT is delivered to the reader. Actually, Markowitz proposed in 1952 not directly the minimum-variance portfolio under the umbrella of the MPT, the first approach he presented, is the mean-variance portfolio optimization concept. As the name might suggest, it creates a portfolio from a combination of the first two moments, expected mean and variance. Thereby, it is important to understand, that the portfolio with the maximum expected return is not necessarily the portfolio with the lowest variance (Markowitz, 1952). Also, a risk-averse investor, who wants to minimize her risk by minimizing the portfolio’s variance is not done by investing in many equities. In the context of the MPT Markowitz states, that an investor must invest in many different equities with low correlation. Here is an example for clarification. The variance of a portfolio consisting only of tech companies has a higher variance than a portfolio with the same number of equities but consisting of equities from different sectors like oil, technology, pharma, and consumer goods. This is obvious, since the “tech only” portfolio directly depends on the ups and downs of the tech sector, while the “mixed” portfolio is more diversified. By combining maximum expected return and minimum variance through an optimized wealth allocation the mean-variance portfolio offers an advantage over other portfolio optimization concepts. It offers a portfolio with the maximum expected return while simultaneously not exceeding a certain variance. Also, the mean-variance portfolio can be interpreted as offering the minimum variance for a given expected return. Mathematically, the mean-variance portfolio is expressed by the following optimization problem:

\[
\max_{w_t} \ w_t^T \mu_t - \frac{1}{2} \lambda w_t^T \Sigma_t w_t \\
\text{s.t. } w_t^T \mathbf{1}_N = 1
\]

Equation (1) states thereby the tradeoff between the portfolio’s expected return and the portfolio’s variance for period \( t \). \( \mu_t \) is an \( N \)-dimensional mean return vector of each asset, where \( N \) is the number of assets in the portfolio. \( w_t \) is the weight vector at period \( t \) that
is solved by the optimization problem with the dimensions $1 \times N$. $\Sigma_t$ is the $N \times N$ variance-covariance matrix at period $t$. $\lambda$ represents the risk-aversion parameter, the higher this parameter the more preference an investor has toward a low variance of her portfolio. The constraint in equation (2) assures that all weights must sum up to one, meaning all wealth must be invested in the assets of the portfolio and no cash positions are allowed. Thereby, $1_N$ is a $N$-dimensional vector of ones. Also, the mean-variance portfolio allows for long and short positions. The relative weight vector for the mean-variance portfolio received by solving the optimization problem above is presented by equation (3).

$$\hat{w}_{\text{mean-var}} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{1_N\hat{\Sigma}^{-1}\hat{\mu}_t} \quad (3)$$

Note, the expected mean $\hat{\mu}$ and variance $\hat{\Sigma}$ are estimators estimated from the data. One can construct a line of portfolios with different expected returns for given possible minimum variances (Markovitz, 1952). This line is called the efficient frontier, exemplarily visualized in Fig. 1.

![Efficient Frontier](image)

*Figure 1: Efficient Frontier which presents optimal portfolios that offer the maximum expected return for given possible minimum variances.*

All portfolios below the efficient frontier are non-optimal and can be replaced by an optimal portfolio that has a higher expected return for a given variance or a lower variance for a given expected return (Markowitz, 1952). As already mentioned, to calculate those portfolios it needs two parameters: The expected future returns in order to calculate the expected means and the variances of the assets in order to calculate the variance-covariance matrix. Since future expected means and the covariance matrix are estimated
it is obvious, that as long as an estimation is done, there is also an estimation error. However, according to Merton (1980) it is more difficult to estimate future expected means than covariances. Therefore the resulting estimation error of the estimated expected means is way larger and has a higher impact than the estimation error emerging from the estimation of the variance-covariance matrix (Merton, 1980). Merton supports his argument by comparing the variance of the estimated mean $\hat{\mu}$ needed for the calculation of the expected mean and estimated variances $\hat{\sigma}^2$ needed for the calculation of the variance-covariance matrix. Note, the lower the variance of an estimator the more accurate the estimator is, resulting in smaller estimation errors. The variance of the expected mean is defined in equation (4).

$$\text{Var}(\hat{\mu}) = \frac{\sigma^2}{h}$$  \hspace{1cm} (4)

Where $h$ is the length of the time horizon of returns used for the calculation of the expected mean and $\sigma^2$ is the variance of the returns within that time series of returns. As the reader might notice, the variance of the expected return only decreases if a longer time series of returns are used for its calculations. On the other hand, the variance of the returns, as a time series, might significantly change over time, resulting in heteroscedasticity of the expected return estimator (Merton, 1980). This is a clear disadvantage regarding the accuracy of the expected mean estimator. Further Merton shows that the accuracy of the variance estimator can be expressed by the following equation:

$$\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n} + \frac{4\mu^2 h}{n^2}$$  \hspace{1cm} (5)

Where besides, the already known parameters $h$ and $\sigma^2$, $n$ and $\mu$ determine the accuracy of the variance estimator. $\mu$ is the mean of the variances within the time series of variances that are used for the estimation of $\hat{\sigma}^2$. If $\mu$ increases so does $\text{Var}(\hat{\sigma}^2)$. As this is exogenously given by the data nothing can be done at this point to decrease $\text{Var}(\hat{\sigma}^2)$. More interestingly though is $n$, which is the frequency of the time series of variances used for estimation. The variance of the variance estimator, i.e. its accuracy, can be significantly improved by using a higher frequency of data for the estimation. This is a practical advantage of the variance estimator’s accuracy because the frequency $n$ influences $\text{Var}(\hat{\sigma}^2)$ to the second power. Thus $\text{Var}(\hat{\sigma}^2)$ rather depends on the frequency than on the length of the time series used for the estimation. For example, using weekly
or even daily data, instead of monthly or yearly, leads to significantly more accurate estimates. While the estimate of the expected return estimated on the same estimation window has such a high estimation error, that is almost useless, due to the heteroscedasticity of returns and a too small estimation window (Merton, 1980).

Hence, from Merton one can suggest that using the expected mean in portfolio optimization can lead to serious misallocations of portfolio weights. Therefore, Jagannathan & Ma (2003) state that the estimation error of the expected mean is so large that it is not very different when ignoring it. This is the reason why most of today’s literature only focuses on a modification of the mean-variance portfolio which only uses the variance for its optimal wealth allocation. This portfolio is called the minimum-variance portfolio. Figure 2 shows its position on the efficient frontier.

![Minimum-variance Portfolio](image)

**Figure 2:** Position of minimum-variance portfolio on the efficient frontier as the portfolio with the lowest variance with the lowest expected return.

The minimum-variance portfolio is the far most left portfolio of all optimal portfolios on the efficient frontier. This means it offers on one hand the lowest optimal expected return, but on the other hand, also the lowest optimal variance, which is especially for risk-averse investors very suitable. Fortunately, the optimization problem of the mean-variance portfolio only needs a small adjustment to turn into the minimum-variance optimization problem.

\[
\begin{align*}
\min_{\mathbf{w}_t} & \quad \mathbf{w}_t^T \Sigma \mathbf{w}_t \\
\text{s.t.} & \quad \mathbf{w}_t^T \mathbf{1}_N = 1
\end{align*}
\]

Equations (6) and (7) present the minimum-variance optimization problem. Whereby, only equation (6) exhibits changes. Equation (7) is the same constraint as it is for the
mean-variance optimization problem. The difference in equation (6) to (1) is that the expected return part is canceled and only the variance part is taken in the optimization. The relative weight vector of the minimum-variance portfolio is expressed by the following:

\[
\hat{w}_{\text{min-var}} = \frac{\sum^{-1}1_N}{1_N^T \Sigma^{-1} 1_N}
\]

By comparing the relative weight vector of the minimum-variance portfolio with the one of the mean-variance portfolio the reader sees, that the expected return estimates are canceled for the minimum-variance weight vector. In theory, this should result in superior out-of-sample performance of the minimum-variance portfolio compared to the mean-variance portfolio, because of smaller estimation errors.

A paper that puts that theory to the test is the paper “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?” by DeMiguel et al. (2009). They compare different portfolio strategies, two of which are the mean-variance portfolio and the minimum-variance portfolio. The paper’s framework uses different market portfolios as “assets” for portfolio optimization. These market portfolios inherent 10 – 20 different assets from US equity market sectors. The optimal weight allocations are calculated based on monthly returns. Further, out-of-sample performance is evaluated by using the Sharpe Ratio, Certainty Equivalent, and Turnover as performance measurements. In all three performance measurements, the minimum-variance portfolio achieves an outperformance over the mean-variance portfolio (DeMiguel et al., 2009). These results seem to support the arguments of superiority taken by Merton (1980) and Jagannathan & Ma (2003) of the minimum-variance portfolio compared to the mean-variance portfolio. Therefore, it delivers the reason why the minimum-variance portfolio is chosen for out-of-sample performance investigation in this paper.

As in every theory, model, or in this case portfolio optimization concept it is necessary to ask the question of what assumptions are needed to be taken. Therefore, this paragraph lists the necessary assumptions for the minimum-variance portfolio. First, since the minimum-variance portfolio is settled in the MPT all necessary assumptions of the MPT must hold. These assumptions are: Investors are in general risk-averse, further market returns are given and normally distributed such that expected returns, variances, and covariances can be calculated. Another assumption, which must hold in the context of the
MPT is that markets need to be efficient, i.e. available information is incorporated into the stock prices (Markowitz, 1952). Turning now toward assumptions that must hold but are out of the context of the MPT. One additional crucial assumption is assets can be held at any amount, meaning partial share purchase is possible. A sophisticated reader might have noticed another more indirect assumption that specifically in the framework of the minimum-variance portfolio is made: The true variance-covariance matrix $\Sigma$ in the optimization problem is simply replaced by an estimated variance-covariance matrix $\hat{\Sigma}$ that is estimated from the past data without dynamic adjustments (predictions) about possible future changes. DeMiguel et al. (2009) called this a “plug-in approach”. Lastly, since past performance is used to predict future performance the assumption is made that good-performing stocks are performing well in the future and vice versa. However, this further requires the assumption that the stock’s attribute (good or bad performance) is stationary. Unfortunately, the stationary assumption cannot be proven and is usually not fulfilled in financial markets, since market dynamics change over time (DeMiguel et al., 2009; Merton, 1980). Nevertheless, DeMiguel et al. (2009) apply the portfolio strategy.

To close the section in the following advantages and disadvantages of the minimum-variance portfolio are provided to the reader. Starting with the first advantage is that in order to calculate the optimum portfolio weights the minimum-variance strategy only needs stock returns, which are nowadays easily accessible on the internet. Moreover, since the mathematical concepts behind the minimum-variance portfolio are not complex, it is easy to implement and apply. This advantage goes hand in hand with the small computational burden the portfolio strategy has and therefore it can be run on any computer hardware. Lastly, the MPT is well investigated resulting in a rich pool of assistance available on the internet, providing a low barrier to do further investigations. Changing now perspectives and examining the disadvantages the minimum-variance portfolio strategy comes up with. As the reader remembers, the portfolio strategies assume normally distributed returns. It is questionable if a normal distribution of returns might always be the case (DeMiguel et al., 2009). Another disadvantage is that the minimum-variance portfolio tends to assign extreme positive and negative weight allocations far from optimal, because of the results from the estimated variance-covariance matrix (DeMiguel et al., 2009). A third attributable disadvantage is the missing connection between past and today/future since the variance-covariance matrix is only estimated on past historical data, without a prediction approach for possible future developments. Finally, one critical disadvantage pointed out in the paper by DeMiguel et
al. (2009) is the severe estimation error the variance-covariance matrix suffers from depending on the chosen time series length for the estimation. The authors show that the estimation window necessary for a tolerable low estimation error would be 3000 periods for a portfolio consisting of 25 assets and astonishing 6000 periods for 50 assets. These estimation window lengths are far from what is applied in practice. Practitioners commonly apply an estimation window length of 60 or 120 periods. Therefore, it is obvious that the variance-covariance matrix suffers from significant estimation errors.

A portfolio strategy that tries to cope with the disadvantage regarding the estimation error is a Bayesian approach invented by Jorion (1986) called the “Bayes-Stein estimator”. The next section is dedicated to explaining the necessary background information to the reader.

2.2 Jorion’s Bayes-Stein estimator approach for the minimum-variance portfolio

Since the common recognition of minimum-variance portfolio suffering from an estimation error by the literature, there are several techniques invented for tackling it. One that delivers very good out-of-sample performance in comparison to other approaches and portfolio strategies is Jorion’s Bayes-Stein estimator applied to the minimum-variance portfolio strategy that uses a shrinkage technique (DeMiguel et al., 2009). The underlying concept of Jorion’s Bayes-Stein estimator is elaborated in the next subsections.

In general, a shrinkage estimator, which the Bayes-Stein estimator belongs to, applies some sort of distance reduction (shrinkage) from the estimated value to a predetermined overall expected mean. The overall expected mean is calculated on the volatility of means an asset has had in the past (Jobson & Korkie, 1980). This dynamic allows to connect past historical data with the ability to predict possible future changes in expected means. From a performance perspective, the new shrunken estimator is sometimes better, but never worse than that of the original one, according to Jobson & Korkie (1980). Because of the reduction in estimation error from which the minimum-variance portfolio strategy suffers, the fairly intuitive mathematical and statistical theory behind it, and the direct implementation capability into the minimum-variance portfolio strategy the Jorion’s Bayes-Stein estimator approach is included in this paper for investigation of its out-of-sample performance compared to other modern portfolio strategies.
Let us get into the mathematics regarding the application of Jorion’s Bayes-Stein estimator on the minimum-variance portfolio strategy. Recall the relative weights vector, equation (8), receiving by solving the minimum-variance optimization problem. In order to implement the Bayes-Stein approach from the optimization problem’s perspective, the calculation of the relative weights vector in the minimum-variance problem stays the same, but the variance-covariance estimate \( \hat{\Sigma} \) remains no longer a classical “plug-in” approach, instead it is replaced by a Bayes-Stein estimated variance-covariance \( \hat{\Sigma}_{\text{Bayes-Stein}} \). The new relative weights vector for the Bayes-Stein minimum-variance portfolio strategy is presented by equation (9).

\[
\hat{w}_{\text{Bayes-Stein}} = \frac{\hat{\Sigma}_{\text{Bayes-Stein}}^{-1} \mathbf{1}_N}{\mathbf{1}_N^T \hat{\Sigma}_{\text{Bayes-Stein}}^{-1} \mathbf{1}_N} \tag{9}
\]

So far so good regarding the implementation, now we dive into the theory behind Jorion’s Bayes-Stein estimator and how \( \hat{\Sigma}_{\text{Bayes-Stein}} \) is calculated.

The work by Jorion builds upon previous studies by Barry (1974), Brown (1976), and Klein & Bawa (1976), which all contribute and lead to the objective to minimize estimation risk in context of optimal portfolio choice. Jorion considers estimation risk as a source of uncertainty about a present event whose probability of occurrence in the past is not observable. Since it is unobservable this turns out to be a problem. The solution to that problem Zellner & Chetty (1965) deliver by expressing the uncertainty with a predictive density function. A predictive density function is a distribution of possible unobserved values (events) conditional on the observed values (events) (Zellner & Chetty, 1965). Jorion shows the predictive density function can be described as \( p(\mathbf{r}|\mathbf{y}, \Sigma, \lambda) \), consisting of a future returns vector \( \mathbf{r} \), that is conditional on today’s return observation \( \mathbf{y} \), the real variance-covariance matrix \( \Sigma \), and a precision parameter \( \lambda \). With a mean of the predictive density function stated in equation (10).

\[
E[\mathbf{r}] = (1 - w)\mathbf{Y} + w\mathbf{1}_N\mathbf{Y}_0 \tag{10}
\]

\[
w = \frac{\lambda}{T + \lambda} \tag{11}
\]

\[
\mathbf{Y}_0 = \frac{\mathbf{1}_N^T \Sigma^{-1} \mathbf{1}_N}{\mathbf{1}_N^T \Sigma^{-1} \mathbf{1}_N} \mathbf{Y} \tag{12}
\]
Where \( w \) is the shrinkage coefficient, \( Y \) is the vector of observed averages with dimensions \( N \times 1 \), \( 1_N \) is a \( N \)-dimensional vector of ones, and \( Y_0 \) is the overall expected mean. The calculation of the shrinkage coefficient \( w \) and the overall expected mean \( Y_0 \) are presented in equations (11) and (12), respectively. Thereby is \( T \) the number of time periods the estimation window contains. However, the mean of the predictive density function is for the minimum-variance portfolio not the point of interest. Therefore, equations (10), (11), and (12) are stated for completeness of Jorion’s derivations regarding the predictive density function. The point of interest to use for the minimum-variance portfolio is stated by equation (13), which is the variance-covariance matrix of the predictive density function.

\[
\text{Cov}(r) = \Sigma(1 + \frac{1}{T + \lambda}) + \frac{\lambda}{T(T + 1 + \lambda)} \frac{1_N^T 1_N}{\Sigma^{-1} 1_N^T 1_N} \tag{13}
\]

In order to transform equation (13) into the estimated \( \hat{\Sigma}_{\text{Bayes-Stein}} \), that is needed for the calculation of the relative weight vector in equation (9), estimates for the parameters \( \Sigma \) and \( \lambda \) are needed. Luckily, both parameter can be estimated directly from the data. For the variance-covariance matrix \( \Sigma \) Jorion uses an estimation suggestion from the paper by Zellner & Chetty (1965), which uses the usual unbiased sample variance-covariance matrix \( S \) and adjusts it by a factor that is a combination of the number of assets in the portfolio \( N \) and the number of time periods \( T \). The respective formula is stated by equation (14).

\[
\hat{\Sigma}_{\text{Zellner-Chetty}} = \frac{T - 1}{T - N - 2} S \tag{14}
\]

Regarding the precision parameter \( \lambda \) Jorion presents a PDF of \( \lambda \) that is gamma distributed \( p(\lambda | \mu, \eta, \Sigma) \) with a mean of

\[
E[\lambda] = \frac{N + 2}{(Y - 1_N Y_0)^T \Sigma^{-1} (Y - 1_N Y_0)} \tag{15}
\]

Since now every parameter from equation (13) is either known or estimated by the data (16) presents the calculation of Jorion’s Bayes-Stein estimator for the variance-covariance matrix used in (9) to calculate the relative weight vector for the minimum-variance portfolio strategy.
\[ \hat{\Sigma}_{\text{Bayes–Stein}} = \hat{\Sigma}_{\text{Zellner–Chetty}} (1 + \frac{1}{T + \lambda}) + \frac{\lambda}{T(T + 1 + \lambda)} \frac{1^T_N}{1^N} \hat{\Sigma}_{\text{Zellner–Chetty}}^{-1} 1^N \]

Regarding the assumptions of the Bayes-Stein minimum-variance portfolio, it is quite obvious that on one hand all assumptions made on the base case of the minimum-variance portfolio in section 2.1 must hold, too. On the other hand, one additional further assumption has the be made. The Bayes-Stein estimator can only be used if the observed random variables are greater than one and there is a reasonable correlation between them. Both conditions are given in a portfolio optimization context (Jorion, 1986).

Because of its straightforward implementation, the Bayes-Stein estimator approach by Jorion keeps all advantages from the basic minimum-variance portfolio strategy. Further, it gains the advantage by reducing the estimation error of the basic minimum-variance strategy. Also, all necessary parameters can be estimated directly from the data. However, the Bayes-Stein minimum-variance strategy also relies on normally distributed returns. Moreover, a critical disadvantage of the Bayes-Stein approach is if the means have unconventional values there is no guarantee that the Bayes-Stein estimator leads to a significant reduction in estimation error. Quite the contrary it can lead to serious deterioration of the estimator resulting in false portfolio weight allocations (Efron & Morris, 1977).

Therefore, it is even more important to provide high-quality data to the portfolio optimization models such that they can optimally calculate portfolio weight allocations and do not suffer from poor data problems or false return distribution assumptions. In the next two sections, the paper presents two further portfolio strategies, that either completely ignore the data or do not rely on distributional assumptions.

2.3 I/N portfolio strategy with rebalancing

Despite sophisticated investigations and advancements in portfolio optimization approaches over the last decades, authors like Benartzi & Thaler (2001) and Huberman & Jiang (2006) could document, that investors tend to allocate their wealth evenly across their assets. This behavior did not go unnoticed by economic researchers and thus in 2009 DeMiguel et al. raised the question “How inefficient is the I/N portfolio strategy?” In their paper they compare optimal portfolio strategies, like mean-variance, minimum-variance, market-value weighted portfolios, and Bayesian approaches to the out-of-
sample performance of the 1/N approach. The performance is evaluated in terms of Sharpe Ratio, Certainty Equivalent, and Turnover for each portfolio strategy on six different datasets. Thereby each of the six datasets consists of 10 – 20 different sector portfolios. Each sector portfolio is constructed of stocks belonging to certain sectors, market values, or fundamental factors of companies. The results of the paper by DeMiguel et al. are surprising. Regarding the Sharpe Ratio the 1/N portfolio has across all six datasets an average rank of four. Only the minimum-variance portfolio and two further extensions of it are able to outperform the 1/N portfolio. For the Certainty Equivalent, the 1/N portfolio performs best on average across all six datasets, where second and third ranks are again the minimum-variance portfolio and one of its extensions. Second best performance on average achieves the 1/N portfolio with respect to Turnover. Here, only the value-weighted portfolio outperforms the 1/N portfolio. Overall, the surprising result of the work of DeMiguel et al. is that the 1/N portfolio performs best when the average across all performance measures is taken. This remarkable finding and the conceptional difference of the 1/N portfolio compared to the minimum-variance approach provides reason to include the 1/N portfolio in this paper.

Mathematically speaking is the calculation of the relative portfolio weight vector very simple, as equation (17) shows.

\[ w_t^{1/N} = \frac{1}{N_t} \]  

(17)

Where \( N_t \) is the number of assets the portfolio consists of at period \( t \). The reader can interpret equation (17) such that the wealth distribution of an investor is evenly distributed among the number of assets in the portfolio. For example, if the portfolio consists of 10 stocks, every individual stock gets 1/10th of the total investor’s wealth. Thus the 1/N portfolio does not estimate or optimize portfolio weights, it naively distributes wealth and completely ignores the stock market data.

Nevertheless, two assumptions need to be made. First, all wealth must be distributed into the assets/stocks, such that the sum over all weights must be equal to one. i.e. no cash positions can be taken. Second, partial share purchases should be possible, meaning assets can be held at any amount.

Because of its mathematical simplicity and the few assumptions the reader may ask what advantages and disadvantages the 1/N portfolio strategy might entail. Starting with
the first and most obvious advantage, the 1/N portfolio strategy is very easy and intuitive to implement. Also, since no estimation is done in order to calculate the portfolio weights, no estimation error can occur. Moreover, the data quality does not matter because it completely ignores the data. Likewise, it does not rely on any distributional assumption. Regarding the 1/N portfolio’s disadvantages, while estimation approaches, like minimum-variance, usually perform better if the number of stocks is small, the 1/N strategy performs worse (DeMiguel et al., 2009). The reason for this is, the lower the number of stocks in a portfolio the lower degree of diversification a naive approach achieves, resulting in a higher unsystematic risk. In addition, even if the number of stocks is large enough to diversify away unsystematic risk and only systematic risk remains, the 1/N portfolio just yields the market return. An investor can never expect to outperform the market significantly, because it completely ignores the data (DeMiguel et al., 2009).

At this point of the paper, three out of four portfolio strategies are presented to the reader. In the next section the most modern portfolio optimization concept is introduced to the reader, a Convolutional Neural Network which uses a hedging strategy named “Equity Market Neutral” for portfolio optimization.

2.3 Equity Market Neutral portfolio implemented by a CNN

Due to technological advances in computer hardware, desktop computers have nowadays more computational power than huge IBM servers in the 1980s. These changes also affect the research in the field of portfolio optimization. Portfolio optimization models, that require high computational power can now be investigated with most of the hardware at our homes. One of them are neural networks.

But why are neural networks so interesting for portfolio optimization? Neural networks have several advantages over other portfolio optimization approaches. One advantage is their numeric nature. Nominal techniques require a transformation from numeric data into nominal data, which comes with a loss in information. Furthermore, different time intervals and transformation techniques lead to different results (Bahrammirzaee, 2010). Neural networks do not have such an issue, because they are capable of processing “raw” numeric data without conversion to nominal values. Another edge over other portfolio optimization concepts is that neural networks do not need any distributional assumptions regarding the input, i.e. returns (Bahrammirzaee, 2010). Also, they are more flexible in terms of they can use new data for already trained models to
update them and generate new results. On the contrary, other portfolio optimization approaches or in general statistical methods can only use old and new data to retrain their models separately. Another big advantage is neural networks can generate estimates from data without any application of a model. Therefore, they capture relations/interaction effects among variables and formulate their own models with respect to the given data (Bahrammirzaee, 2010).

At the writing of this paper, a very recent study is conducted by Gunjan & Bhattacharyya (2023) comparing classical portfolio optimization approaches (mean-variance, etc.) to modern, intelligent approaches (different neural network strategies, quantum computing strategies, etc.). They find neural networks perform very well against other approaches on average. Among different neural network strategies one particular portfolio optimization strategy, the Equity Market Neutral (EMN) portfolio implemented by a Convolutional Neural Network (CNN), taken from the paper by Wu et al. (2021) performs on average best. Because of these findings and the advantages of neural networks over other portfolio optimization approaches the presented strategy by Wu et al. (2021) is used as the fourth and last portfolio strategy to investigate its out-of-sample performance compared to other modern portfolio strategies in this paper.

Let us begin by diving into the concepts behind the EMN strategy. Wu et al. (2021) take the basic concept of the EMN from the paper by Patton (2009), which is a hedging strategy in terms of it exploits differences in stock trends. It sells (shorts) stocks that are relatively weak and holds (buys) stocks that are relatively strong, such that it earns the spread between the two positions (Patton, 2009). Therefore, two independent neural networks are trained, one only allowed to open long positions and the other one only short positions. A further second restriction is imposed, that is for both independent neural networks the weights of their long and short positions must sum up to one respectively, as equations (18) and (19) show.

\[ W_L = (W_{L1}, W_{L2}, \ldots, W_{LM}), \quad W_{Li} \in [0,1], \sum_{i=1}^{M} W_{Li} = 1 \]  

(18)

\[ W_S = (W_{S1}, W_{S2}, \ldots, W_{SM}), \quad W_{Si} \in [0,1], \sum_{i=1}^{M} W_{Si} = 1 \]  

(19)

Where \( M \) is the total number of positions, long or short, are taken respectively. To prevent a situation in which the same stock is bought by a long position and sold by a short
position simultaneously the final portfolio weights are calculated by the summation of the corresponding weights for each stock, shown by equation (20).

\[
\sum_{i=1}^{M} W_{Ci} = \sum_{i=1}^{M} W_{Li} - W_{Si} = \sum_{i=1}^{M} W_{Li} - \sum_{i=1}^{M} W_{Si} = 1 - 1 = 0
\]  

Where \( W_{Li} \) and \( W_{Si} \) are the long (\( L \)) and short (\( S \)) weights assigned to a specific stock \( i \). \( W_{Ci} \) is the combined final weight for stock \( i \). It is worth noticing, the investment amount needed for the longs is gained from the short positions Wu et al. (2021). From equation (20) one can see that the size of longs is equal to the size of short positions which is indicated by the zero at the end of equation (20). This means in theory the EMN not imposes any investment capital and net investments are zero (Wu et al., 2021). In practice however, shorting a stock requires a certain margin on the broker’s bank account to maintain the short position. As the reader notices, the EMN strategy is not difficult to understand, but what is more complex is the implementation of the portfolio strategy through a Convolutional Neural Network. Therefore, the following question is raised to provide the necessary background information.

How do neural networks in general and specifically the CNN work? And how is the EMN implemented by CNN?

Since the usage of neural networks in finance is relatively new compared to other portfolio optimization techniques it is not assumed that the reader is familiar with the basic concepts of neural networks, therefore this paper introduces them to the reader. Important, only basic concepts of neural networks are explained by the following, since it would be beyond the scope of this paper to provide detailed explanations of the functioning and concepts of neural networks. Sophisticated readers are referred to study the paper “Neural networks and their applications” by Bishop (1994), which provides more detailed and general explanations of how neural networks work.

The reader can image a neural network of interconnected nodes, as a human brain has. Each node can receive several input signals, called connections, and provides an output signal. For a node in a neural network to each connection a weight is assigned. Figure 3 visualizes these weight assignments. For each given node, input signals are multiplied by the assigned weights and are summed together in the next step. This summed value is referred to as summed activation potential for that specific node (Coakley & Brown, 2000).
The summed activation is then transformed via an activation function, also called transfer function, see Fig. 3. In the end the activation function decides the output of the node. In other words, whether the node is activated or not (Coakley & Brown, 2000). The activation of nodes within a neural network leads to a learning of the network, similar to a human brain. Different activation functions lead to different learning mechanisms of the network. In general, there are two different types of activation functions, linear and non-linear activation functions. Linear activation functions do not apply any transformation to the summed activation at all (Goodfellow, Courville, & Bengio, 2016). Figure 4 displays the activation potential a node outputs with a linear activation function. Where the x-axis is the summed activation potential and the y-axis scales the output generated by the activation function. The resulting blue line is the by the activation function transformed output of the node.

**Figure 3: Components of a single node**

**Figure 4: Linear activation function result in a linear activation of the node according to the summed activation input.**

Source: https://mlnotebook.github.io/post/transfer-functions/
The reader can interpret Fig. 4 such that the summed activation potential is received by the activation function and is one-to-one transformed as an output signal of the node (same scaling of the x- and y-axis). In other words, no filter is applied to the weighted summed input signals the node receives (Goodfellow et al., 2016). Such networks are used for simple regressions, because they are very easy to train, but cannot learn complex relations. Non-linear activation functions, on the other hand, are more difficult to train, but in exchange, they can learn more complex structures. Traditionally, two widely used non-linear activation functions are the sigmoid activation function and the hyperbolic tangent (tanh) activation function. The sigmoid activation function is also called logistic activation function, because of how it works and looks on a graph, see Fig. 5.

![Sigmoid Function](https://mlnotebook.github.io/post/transfer-functions/)

*Figure 5: Sigmoid activation function transforms summed activation potential into values between 0.0 and 1.0 and then not activates or activates the nodes at a threshold of 0.5, respectively.*  
*Source: https://mlnotebook.github.io/post/transfer-functions/*

The sigmoid activation function takes the summed activation potential and transforms it into a value between 0.0 and 1.0, visualized by the blue line in Fig. 5. For summed activation potentials bigger than 1.0 it transforms them to 1.0, and for inputs smaller than 0.0 it transforms them to 0.0. Thereby, the sigmoid activation function decides to activate the node if the outcome of the transformation is bigger than 0.5. In this case, the output of the node is 1 in binary terms (Goodfellow et al., 2016). The same applies vice versa. For a long time sigmoid activation functions were default activation functions because of these characteristics. Similarly shaped is hyperbolic tangent (tanh) activation function, shown by Figure 6.
Figure 6: Hyperbolic tangent activation function transforms summed activation potential into values between -1.0 and -0.0 and then not activates or activates the node at a threshold of 0.5, respectively.  
Source: https://mlnotebook.github.io/post/transfer-functions/  

The difference is it transforms received summed activation potential values to values between -1.0 and 1.0. Whereby, the threshold for activation lays here at 0.0. So if the transformed summed activation potential is above 0.0 the node outputs 1 in binary terms (Goodfellow et al., 2016). This allows a slightly different learning of the neural network compared to a network that uses sigmoid activation functions, but the differences are not explained further and readers are referred to the paper by Goodfellow et al. (2016) since this is out of the scope of this paper.  

However, sigmoid and hyperbolic tangent activation functions do have a disadvantage that should be not underestimated. They saturate, meaning larger values than 1.0 and smaller values than 0.0 or -1.0 for sigmoid and tanh respectively are transformed such that the values equal 1.0, 0.0, or -1.0 with respect to the given summed activation potentials and depending on the activation function. This results in the characteristic of being only sensitive to changes around their thresholds, 0.5 for sigmoid and 0.0 for tanh. Once saturated, it becomes challenging for the network’s learning mechanism resulting in a stagnation of performance improvement of the neural network’s model (Goodfellow et al., 2016). To tackle the problem researchers invented the so-called “Rectified linear activation function” (ReLu), which is more sensitive to every given summed activation potential and avoids saturation. Therefore, the ReLu is a better choice when it comes to learning complex structures and continuing weight adaptation for performance improvement (Goodfellow et al., 2016). As Fig. 7 shows, this activation function returns based on the summed activation potential provided 0.0 or the same values as the summed activation potential has had.
Figure 7: Rectified Linear Unit activation function transforms summed activation potentials into 0.0 for values smaller than 0.0 and applies no transformation to values bigger than 0.0. Source: https://mlnotebook.github.io/post/transfer-functions/

As the reader sees from Fig. 7 the ReLu is linear for values greater than 0.0. This brings a lot of desirable properties of linear activation function when training a network, but also combines it with non-linearity for values smaller than 0.0 (Goodfellow et al., 2016). Because of these properties, the ReLu activation function is commonly used nowadays. Also, in the CNN from the paper by Wu et al. (2021) for the EMN.

Now, that the reader knows how a node functions and how input signals are processed by a node, the focus is shifted one level higher to the architecture of a neural network. Therefore it is important to note, that nodes can be grouped into several layers which then represent the architecture of the network. A neural network consists of an input layer, one or more hidden layers, and an output layer, Fig. 8 visualizes that.

Figure 8: Architecture of a neural network in general with one input layer, one or more hidden layers and an output layer.

In the input layer, each node corresponds to a different explanatory variable. In the hidden layers the learning, fitting, and generalization of the neural network to the given data
happens. By the way, a neural network with more than one hidden layers is called Deep Learning Neural Network. The output layer is usually one node that represents the output of the neural network (Sermpinis, Karathanasopoulos, Rosillo, & La Fuente, 2021). In case of portfolio optimization, the output would be the relative portfolio weight vector. The architecture of a neural network describes how the nodes are interconnected and how many layers are used. Thereby, the most significant difference in the functioning and learning of a neural network depends on how many hidden layers the network has. For understanding, three cases are presented to the reader over the next lines. Case one no hidden layers, case two one hidden layer, and case three two hidden layers. Most of the economic literature does not apply more than three hidden layers therefore the explanations up to the case with two hidden layers should be sufficient since the inclusion of further hidden layers does not change the basic concepts necessary for understanding. Since this part could be hard to understand for the reader one the first attempt Fig. 9 visualizes the differences in learning a neural network has with no, one, and two hidden layers. By looking at all three panels of Fig. 9 the reader can see squares and circles which represent two different kinds of data points within a sample. A neural network is able to divide the sample by hyperplanes such that it minimizes data points from being on the “wrong side” indicated by the white and blue background color in Fig. 9. Note that the hyperplane can be highly nonlinear. The left panel of Fig. 9 refers to the case when there is no hidden layer within the neural network. This case resembles the most to the mean squared error approach (Coakley & Brown, 2000).

![Figure 9: Advantages of having hidden layers within a neural network](image)

By adding one hidden layer the network is now able to create new hyperplanes for each node that is within that hidden layer (Coakley & Brown, 2000). In the case of the middle panel from Fig. 9, there are two nodes within one hidden layer. The neural network can
now divide the sample by using two hyperplanes, which increases its accuracy of having the corresponding data points on the “right side”. Thereby, the selection/combination of the hyperplanes such that most data points are on the “correct side” is done by the output layer (Coakley & Brown, 2000). The right panel of Fig. 9 shows a network with two hidden layers, each with two nodes. Based on the example for one hidden layer the second hidden layer functions as an “output layer” which selects/combines both hyperplanes from the first hidden layer such that the most data points are on the “right side”. From this in the next step, the actual output layer selects/combines the two hyperplanes from the second hidden layer such that by combining all hyperplanes most data points are on the “right side” (Coakley & Brown, 2000). The result is an even higher increase in the accuracy of the network. The combination of the four hyperplanes is visualized in the right panel of Fig. 9 by the blue background.

Now that the reader is familiar with the general basic concepts of neural networks and how they function, it is important to mention that there are several different classes of neural networks. For example, Radial Basis Function Networks (RBF), Higher Order Neural Networks (HONN), Recurrent Neural Networks (RNN), or Convolutional Neural Networks (CNN). All function in different ways. As already mentioned, the network used by the paper of Wu et al. (2021) is a CNN. This is because a CNN is very successful in detecting patterns in complex data. Originally, CNNs are used in processing multidimensional data like images and video to detect patterns. Wu et al. (2021) have found these capabilities also be useful from a finance perspective. Over the next lines, the paper explains the necessary knowledge to the reader for understanding how a CNN functions. However, the explanations are limited to the essentials. Readers who want to gain more knowledge about how a CNN works are referred to the paper “Introduction into Convolutional Neural Networks” by Jianxin Wu (2017). As all other neural networks, a CNN has an input layer, several hidden layers, and an output layer. Thereby, the input is fed sequentially (period after period) through the layers in which it is processed such that the desired output (portfolio weights) is generated (Jianxin Wu, 2017). It is important to note, that in a CNN an additional layer before the output layer is included in the network. This layer is no hidden layer instead it is used for error backpropagation (Jianxin Wu, 2017). The concept of error backpropagation is to compute output values based on input values, then compare the resulting output values to desired target values. In the next step, an error between the output value of the network and the desired target value is computed, which is transformed into an error signal. This error signal is propagated back
through the network to the input layer, which leads to weight adjustments on the input signal of each node. The described process is repeated until the error is acceptably small (Coakley & Brown, 2000). The advantage of an error backpropagation layer is that the network can update itself more efficiently improving the training process and gaining a higher learning rate (Jianxin Wu, 2017). The learning of a CNN happens in the equally named convolution layers added as hidden layers. In the following the functioning of a convolution layer is explained, but without too much detail and not digging into the deep mathematics. Again, further and deep explanations for sophisticated readers can be found in the paper by Jianxin Wu (2017). A convolution layer gets an input called tensor. Over this input tensor a convolution kernel is sequentially overlaid to perform mathematical tasks. The resulting output of these mathematical tasks is smaller than the input into the convolution layer. That is the point where the network makes generalizations and learns from the given input data (Jianxin Wu, 2017). Figure 10 provides a simplified example of how a convolution layer works. The input tensor for the convolution layer is shown as a matrix in panel (b) of Fig. 10. The convolutional kernel, in this example 2 x 2, in panel (a) of Fig.10 is overlayed over the convolution input tensor beginning in the top left corner. The position of the first overlay is represented by the blue dotted line in panel (b). Mathematically, the product between the convolution input tensor and the convolution kernel is taken for the corresponding same positions. At the end, the sum over the products is taken (Jianxin Wu, 2017).

Figure 10: Functional concept how a convolution layer works with a 2 x 2 kernel. 
Source: adapted from (Jianxin Wu, 2017)

For the top left overlaying between the input tensor and the kernel the mathematics are shown by equation (21).
After the calculation is done, the kernel shifts one row down for which the same mathematics are applied. In Figure 10 this is indicated by the red dotted line. This procedure is repeated until the bottom of the input tensor is reached. Then the kernel shifts one column to the right and repeats the mathematics until the bottom of the input tensor is reached again. After that the kernel shifts again one column to the right, and so on until the kernel reaches the bottom right of the input tensor. The resulting generalized output of the convolution layer, in Fig. 10 right matrix in panel (b), is then finished and can be taken as input for the next convolution layer to perform further generalizations (Jianxin Wu, 2017).

Finally, enough background information is provided to the reader to start explaining how the CNN is implemented into the EMN portfolio strategy in the paper by Wu et al. (2021). The design of the network used by Wu et al. (2021) is stated below in table 1, which is copied from their paper. By examining the reader recognizes, the network consists of an input layer, which receives an input tensor with the dimensions $M \times 4 \times N$.

Table 1  Design of the Convolutional Neural Network

<table>
<thead>
<tr>
<th>Layer</th>
<th># Filter</th>
<th>Filter size</th>
<th>Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>2</td>
<td>$(M \times 4 \times N)$ Vector</td>
<td></td>
</tr>
<tr>
<td>Conv.2D</td>
<td>48</td>
<td>$(1, 3^{rd}$-dim previous output)</td>
<td>relu</td>
</tr>
<tr>
<td>Conv.2D</td>
<td>1</td>
<td>$(1, 3^{rd}$-dim previous output)</td>
<td>relu</td>
</tr>
<tr>
<td>Flatten</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td></td>
<td></td>
<td>softmax</td>
</tr>
</tbody>
</table>

| M-dimensional Vector |

Source: (Wu et al., 2021)

Where $M$ represents the number of stocks in the portfolio, 4 is the number of features here Open, High, Low, and Close prices for each stock, and $N$ is the length of the time series for each feature. In the paper of Wu et al. (2021) the time series length is set to 20, however in this paper a different time series length is chosen. Reasons for that are described in section 3.1. The following formulas explain how the input tensor $T_t$ period $t$ is constructed:

$$T_t = [p_t^1, p_t^2, \ldots, p_t^M]$$  \hfill (22)

$$p_t^i = [\text{Open}_t^i, \text{High}_t^i, \text{Low}_t^i, \text{Close}_t^i]$$  \hfill (23)
One can see the input tensor $T_t$ includes a $4 \times N$-dimensional price vector $P^i_t$ for each stock $i$. Each price vector $P^i_t$ for each stock $i$ consists of four features, which are the normalized $Open^i_t, High^i_t, Low^i_t, Close^i_t$ vectors with the dimension $1 \times N$. The normalization of the $Open^i_t, High^i_t, Low^i_t, Close^i_t$ is done by a division of the latest closing price $close^i_t$. This changes the absolute prices into relative prices. Note, that the lower notation ($open, high, low, close$) represents absolute price data from the stock exchange. The reason for the normalization of the input tensor is that it now contains the changes and fluctuations of the stock prices, which is important for the training of the network (Wu et al., 2021). Returning to table 1 of the network design, the CNN contains three hidden convolution layers, whereby the column “filter size” can be interpreted as the kernel size for the convolution layers. Also, one can see the ReLu activation function is used. The layers ‘flatten’ and ‘dense’ are included to process and transform the output from the convolution layers such that the data gets into an $M$-dimensional vector for the output layer (Wu et al., 2021). Finally, the output layer returns the relative weights for the corresponding stocks in an $M$-dimensional vector. In order to set up and run the CNN properly two more things are needed. That is one the network’s environment and two the policy function. Let us start with the simpler one, the policy function. The policy function, also called reward function, provides rewards to the neural network after each training iteration (Wu et al., 2021). Examples of policy functions can be Volatility, Return, Sharpe Ratio or Maximum Drawdown. The goal of the network is always to optimize its output such that it gets the highest reward from the policy function (Wu et al., 2021). Wu et al. (2021) decide to use the Sharpe Ratio as a policy function because it provides a tradeoff between profitability and risk. Since this makes sense, the Sharpe Ratio is also adopted as a policy function here in this paper. The last part, that is needed to set up and run the CNN is the network’s environment, which can be seen metaphorically as a manager. The
environment receives the input tensor, slices it into training batches, and provides them in iterations to the network. In each iteration it receives the output (portfolio weight vector) from the network, lets the policy function calculates the corresponding reward, and transfers the reward back to the network, such that the network can update itself accordingly. Also, it decides when to stop the training for each training batch and starts a new iteration in the training process (Wu et al., 2021). Now that all separate parts of the EMN portfolio strategy and its implementation by the CNN are explained the following flowchart (Fig. 11), copied and slightly modified from the paper by Wu et al. (2021), provides a visualization of the interrelations. In the flowchart one can see that from the input tensor there is a solid arrow labeled “Training Data” and a dashed arrow labeled “Testing Data”. The reasoning is the CNN gets trained by a bunch of data and then its out-of-sample performance is tested on data, that the CNN has never seen before.

![Flowchart of the EMN portfolio strategy applied by a CNN with all its components and main executions.](image)

*Figure 11: Flowchart of the EMN portfolio strategy applied by a CNN with all its components and main executions.*
Besides the fundamental information about how a CNN is used to implement the EMN portfolio strategy, there are parameters that need to be calibrated such that the CNN functions properly. Unfortunately, Wu et al. (2021) do not provide in their paper a detailed list of the calibrated parameters. The paper with the title “A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem” by Jiang, Xu, & Liang (2017) resembles the most the framework by Wu et al. (2021). Further, there are projects available on GitHub around the paper by Jiang et al. (2017). Therefore, the calibration of the parameters of the CNN is adapted from their work.¹

Due to the strong advantages when applying a neural network, i.e. no distributional assumptions, model-free estimator, etc., only a few more general assumptions are needed in order to apply the EMN portfolio strategy. The first assumption is all wealth must be distributed into assets/stocks, hence no cash positions can be taken. Second, assets can be held by any amount, which means partial share purchases are possible. Lastly, markets are efficient, meaning available information is incorporated into the stock’s prices.

By looking at the advantages of neural networks, the on average remarkable performance, and the relatively weak assumptions one can get the impression that neural networks are the holly grail in portfolio optimization, but not all that glitters is gold. There are some downsides. One is the complexity of neural networks. Oftentimes they are difficult to understand, especially for beginners, since many problems result in a problem of 3rd or 4th-dimensional order. Further, neural networks can be complex in their structure and construction, i.e. how many layers should one uses, which layers, which activation function, etc. Complexity is also present in the calibration of neural networks, how to choose the split size between training and testing data, how large should the training batch be, etc. Besides the complexity, there is also the disadvantage of overtraining a neural network on the training data, leading to false correlation detection by the network, which finally results in a bad out-of-sample performance (Sermpinis et al., 2021). Also, depending on the setup and despite the advancement in hardware power, neural networks can bring a high computational burden (Sermpinis et al., 2021). In the end, the last disadvantage is a neural network is a Blackbox, meaning no one knows what correlations the network finds in the data, since no one can see into a neural network.

3. Methodology and Data

3.1 Methodology

The focus of this section is to provide the reader an insight into the methodology used to answer the research questions. With the methodology presented in the following out-of-sample performance measurement distributions, i.e. Return, Volatility, Sharpe Ratio, and Drawdown distributions, for each portfolio strategy are generated, which enables to calculate means and test for statistical differences in order to draw empirical findings. Also, the formula for each performance measure is provided. It is important to note, the paper by DeMiguel et al. (2009) uses monthly stock returns for the calculation of portfolio weights and the resulting out-of-sample performance. Here in this paper, weekly returns are used due to the following reasons: Compared to the monthly timeframe, the weekly timeframe reflects faster changes in market behavior. Hence the portfolio strategies can reallocate faster to newly given market conditions therefore it should boost their performance, in theory. Moreover, the weekly timeframe is not too small and still contains the essential information on economic events. This point is in terms of the statement by Fama (1976) about how an appropriate timeframe should look like. Also, by deviating from the monthly timeframe, my paper aims to enrich the existing literature, because to my best knowledge, no paper has investigated in an out-of-sample performance comparison of the incorporated portfolio strategies on a weekly timeframe.

Now let me start by explaining how the in-sample dataset is defined, which is later used to calculate the portfolio weights, and then how the portfolio weights are applied. In order to create out-of-sample distributions for Return, Volatility, Sharpe Ratio, and Drawdown a rolling window approach is used. The following figure gives the reader an overview of how the rolling window approach is applied.
Figure 12: Rolling window approach used to create out-of-sample performance distributions for each performance measure in order to draw empirical findings with statistical power.

Step 1: Create In-Sample Dataset

In the first step, an in-sample dataset (estimation window) is selected. It incorporates all stocks from the benchmark index at time period $t$. Whereby $t$ is one period before the first/next out-of-sample performance is evaluated. Since this paper is inspired by the methodology from the paper of DeMiguel et al. (2009) the same in-sample period length of 60 periods is chosen. Hence, the in-sample dataset starts at period $t - 60$ and ends at period $t$.

Step 2: Estimate covariances/train CNN

After the estimation window is selected, the estimation of the variance-covariance matrix is done via plug-in approach for the minimum-variance portfolio and for the bayes-stein minimum-variance portfolio according to equation (16). For the $1/N$ portfolio there is no estimation needed. Therefore, the $1/N$ portfolio skips Steps 1 and 2 from Fig. 12, since the optimal portfolio weights stay at $1/N$. Also, the CNN starts training on the given estimation window.

Step 3: Calculation of portfolio weights

Step 3 implies the calculation of the relative portfolio weight vector according to equations (8), (9), and (17) for the minimum-variance portfolio, bayes-stein minimum-variance portfolio, and $1/N$ portfolio, respectively. Regarding the EMN portfolio, after the long and short optimal portfolio weights are generated separately, see equations (18) and (19), the portfolio weights are combined as described in equation (20).
Step 4: Calculation of out-of-sample performance

Since relative portfolio weights are calculated, in step 4 they are applied to the out-of-sample period in order to evaluate each portfolio strategy’s performance. For the minimum-variance, bayes-stein, and 1/N portfolios the out-of-sample period is the next period \((t + 1)\) after the last period \(t\) of the estimation window. Since a trained CNN is commonly tested on a time series rather than on a single period a 50/50 split between training and testing data is applied in this paper. This is inline with the application in the paper by Wu et al. (2021). To clarify, all portfolio strategies in this paper, also the EMN portfolio get a 60-period long estimation window in order to estimate and optimize. While the minimum-variance, bayes-stein, and 1/N portfolios apply their portfolio weights in Step 4 on one period, the EMN portfolio because of the implementation by a CNN tests itself on 60 out-of-sample periods, resulting in 60 out-of-sample performance periods.

As already mentioned, to calculate the portfolio’s out-of-sample performances Return, Volatility, Sharpe Ratio, and Drawdown are used as performance measures. Those are calculated in Step 4. The corresponding formulas and some comments are listed below. Equation (28) shows the calculation of the portfolio return applied out-of-sample.

\[
Portfolio \text{ Return}_t = \mu_t^{OS} \times w_t
\]  
(28)

Where \(\mu_t^{OS}\) corresponds to the out-of-sample return at period \(t\) of each stock in the portfolio. \(w_t\) is the relative portfolio vector of the corresponding portfolio strategy applied at period \(t\). Since the resulting portfolio returns are weekly portfolio returns the portfolio returns from (28) are multiplied by 52, because a year has 52 weeks, to obtain the annual portfolio return. The next performance measure used is the portfolio’s volatility, shown by equation (29).

\[
Portfolio \text{ Volatility}_t = \sqrt{w_t^T \times \Sigma_t^{OS} \times w_t}
\]  
(29)

Where \(\Sigma_t^{OS}\) corresponds to the out-of-sample portfolio’s variance-covariance at period \(t\). Also here it is necessary to transform the portfolio volatility from weekly to annually. This is done by a multiplication of \(\sqrt{52}\). A portfolio’s volatility is referred to as the risk of a portfolio. The less volatility the less risk is associated with the portfolio and is more preferred by risk-averse investors. As a third performance measure, the Sharpe Ratio is used calculated by equation (30).
Portfolio Sharpe Ratio\(_t\) = \(\frac{\mu_{t}^{OS} \cdot w_{t} - \bar{rf}}{\sqrt{w_{t}^{T} \cdot \Sigma_{t}^{OS} \cdot w_{t}}}\)  \hspace{1cm} (30)

\(\bar{rf}\) refers here to the average risk-free rate taken over the whole time horizon used for the out-of-sample performance evaluation. Same as for the portfolio’s return and volatility, the portfolio’s Sharpe Ratio is given in weekly terms in the first instance and needs to be multiplied by \(\sqrt{52}\) to receive the annual Sharpe Ratio. The Sharpe Ratio is used as a performance measure because it offers a ratio that displays a tradeoff between the portfolio’s excess return (nominator) and its volatility aka risk (denominator). The higher a Sharpe Ratio the lesser riskier the portfolio is for a given return or in other words the more return an investor gets for a given risk. Usually, the Sharpe Ratio should be at least positive, meaning the portfolio is able to earn more than the risk-free rate. The last performance measurement chosen in this paper is the so-called Drawdown. Volatility incorporates negative and positive portfolio fluctuations. However, no investor dislikes positive fluctuations in her portfolio. The Drawdown only incorporates negative portfolio fluctuations, thus it is a measurement for the downside risk of a portfolio. A portfolio’s Drawdown is the percentage difference between the portfolio’s last peak and the subsequent drop. Mathematically it can be expressed by the following equation:

\[
\text{Drawdown}_{t} = \frac{\text{Through Value}_{t} - \text{Peak Value}_{t-n}}{\text{Peak Value}_{t-n}}
\]  \hspace{1cm} (31)

Where the \(\text{Peak Value}_{t-n}\) is the latest highest value of the portfolio. The \(\text{Through Value}_{t}\) is the current portfolio’s value after the highest value of the portfolio (\(\text{Peak Value}_{t-n}\)) was reached. Further, this paper differentiates between average Drawdown and Maximum Drawdown. The Maximum Drawdown is the maximum drop that occurs for a portfolio over a certain predetermined period of time. In this paper, the Max Drawdown is calculated for the respective time horizons.

In Step 4 of Fig. 12 the just explained formulas for performance measurements are applied to calculate the out-of-sample performance for each portfolio strategy. In order to obtain more out-of-sample performance datapoints the in-sample window is “rolled” forward by the length of the out-of-sample. Meaning for the minimum-variance, bayes-stein, and 1/N portfolios the in-sample window is shifted by one period and for the EMN portfolio by 60 periods. Then Steps 1-4 are repeated. This concept is called “rolling-window” approach. One important note regarding Step 1. As the estimation window shifts
it is checked, whether the stock composition of the DAX has changed. If so, the corresponding changes in stock composition are also applied to the in-sample dataset.

**Step 5: Test statistical significance of results**

After the rolling-window is applied for the corresponding time horizon, a distribution of out-of-sample datapoints for each performance measurement is generated. Those distributions are used to calculate means. Also, T-tests are applied to compare whether the distributions of the resulting performances of the applied portfolio strategies statistically differ compared to the DAX performance distributions. T-tests are done on common p-values, i.e. 0.01, 0.05, 0.1.

Through the presented methodology the paper is enabled to answer the research questions in the introduction.

3.2 Data

Let me start this section by answering the question: Why is the DAX chosen as a database? To my best knowledge, there is no academic research done on the minimum-variance, bayes-stein, 1/N, and EMN portfolio which uses the DAX as a benchmark index. Most papers lay their focus on the US equity market. For this reason, this paper aims to fill that gap in academic research and provide insights for investors to out-of-sample performance of modern portfolio strategies with the DAX as an underlying benchmark. As data source Yahoo Finance is chosen. The advantage hereby is the easily available data for all stocks that are used by the portfolio strategies to calculate the performance measures. Also, there is a long time history of stock prices available such that there is no limitation from this perspective. Because of Yahoo Finance as a data source, the data quality of the stocks is high. All stocks have clean available data for their corresponding time periods in which they are included in the DAX. One exception is the METRO stock. Unfortunately, no other reliable data source with good data quality could be found to replace the bad data from Yahoo Finance. Therefore, the stock is excluded from the dataset.

This paper works with two time horizons for out-of-sample performance evaluation of the four portfolio strategies and the DAX as a benchmark, one from 2010 until 2019 and the second from 2010 – 2020. Over the next paragraphs descriptive statistics, comments on the macroeconomic environment, and reasons why the time horizons are
chosen are provided. The first time horizon, 2010 – 2019, is displayed in Fig. 13 as a chart of the DAX. The average annual return of the DAX for the time horizon 2010 – 2019 is 0.0969, with an average annual volatility of 0.1449. An average annual risk-free rate 0.0111 can be stated. To calculate the average annual risk-free rate the average of the German 10-Year Bond Yield for the corresponding same time horizon is taken. The data is downloaded from the website investing.com. As an average annual Sharpe Ratio of the DAX 1.4573 can be calculated. Whereas a max drawdown of 0.2813 is reached. From a performance measurement perspective, the time horizon from 2010 – 2019 can be determined as a stable, profitable, not highly volatile time horizon.

![Figure 13: Cumulative return chart of the DAX for the time horizon 2010 - 2019](image)

This is also reflected from a macroeconomic perspective. Starting with a GDP expansion and a declining unemployment rate over the whole time horizon. Figures 14 and 15 show the GDP and unemployment rate development respectively.
Simultaneously there is a decline in European Central Bank (ECB) funds rate, meaning it gets easier for companies to borrow money from banks, which results in an easier
environment companies can grow. Such an environment is called an “easy monetary policy environment”. Figure 16 shows the development of the ECB’s funds rate.

![Figure 16: Development of ECB’s funds rate](image)


Also the decline in ECB’s funds rate mean a decline in return of the risk-free rate because the ECB’s funds rate and the 10-Year German Government Bond move synchronously. The result is a higher investment inflow into the stock market which pushes the stock market higher.

Those three factors are the most important macro economical factors to describe an economy and the influence they have on the stock market. All three factors have positive developments over the time horizon from 2010 until 2019, contributing to a steady incline of the DAX without major corrections or bear markets. Because of the good economic environment and a steady stock market growth this time horizon is chosen to test the performance of the portfolio strategies in order to compare their results to the DAX’s performance.

The second time horizon, 2010 – 2020, displayed in Fig. 17 as a chart of the DAX, exhibits a lower annual average return of 0.086, and a higher average annual volatility of 0.1517. Also, the average annual risk-free rate decreases to 0.0089.
On the other hand, the average annual Sharpe Ratio increases significantly to 2.0811, whereas also the max drawdown increases to 0.3272 compared to the shorter time horizon. By just analyzing the performance measurements, one can see an overall deterioration of the stock market’s performance in general. The only performance measure that increases in a positive way is the Sharpe Ratio. An explanation therefore can be attributed to the relatively larger decline in the risk-free rate compared to the average annual return.

From a macroeconomic perspective COVID-19, a major exogenous negative productivity shock, hits the German economy at the beginning of 2020. Leading to a major drop in the stock market followed by a rally, see Fig. 17. Such a stock market behavior drastically increases volatility, making it more difficult to earn money. Furthermore, due to the COVID-19 shock the German GDP contracts in 2020 and expands again the year afterward, the same holds for the unemployment rate. Both developments can be traced in Fig. 14 and 15 by the reader. In the thread of a recession, the ECB announces an asset purchase program (AAP) which is better known by the name “quantitative easing” to boost the economy. Thereby, the ECB buys predetermined amounts of government bonds or other assets to stimulate the economy. A side effect of quantitative easing is an increase in ECB’s funds rate, shown by Fig. 18, making it harder for companies to borrow money from banks. Overall, the years from 2020 until 2022 add uncertainty in terms of volatility to the whole time horizon from 2010 until 2022. The goal of this time horizon is to investigate how the portfolio strategies react to a change in the macroeconomic environment and an exogenous negative productivity shock.
For completeness, it is worth mentioning, that in 2021 the number of companies in the DAX is increased from 30 to 40. However, there is no trackable effect and the portfolio strategies adapt dynamically to this change due to the construct of the rolling-window approach.

4. Empirical findings

This section presents and analyzes the empirical findings. In the first subsection the findings from the short time horizon (2010 – 2019) are presented and analyzed. In the second subsection findings from the longer time horizon (2010 – 2022) are presented, analyzed, and compared with the findings from the shorter time horizon. The empirical results, i.e. out-of-sample performances are summarized in Table 2 and 3 for the short and long time horizon respectively. In each table, the benchmark (DAX) and the portfolio strategies are listed row-wise, whereas the performance measurements are listed column-wise. In both tables, the values in parentheses are the p-values related to the DAX as a benchmark. Finally, section 4.3 compares the empirical findings to the existing literature.

4.1 Empirical findings for the time horizon 2010 – 2019

Table 2
Out-of-sample performance for time horizon 2010 – 2019

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Return</th>
<th>Volatility</th>
<th>Sharpe Ratio</th>
<th>Avg. Drawdown</th>
<th>Max Drawdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAX</td>
<td>0.0969</td>
<td>0.1449</td>
<td>0.5611</td>
<td>0.0641</td>
<td>0.2813</td>
</tr>
<tr>
<td>1/N</td>
<td>0.0943</td>
<td>0.1355</td>
<td>0.5682</td>
<td>0.0648</td>
<td>0.2711</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(0.21)</td>
<td>(0.99)</td>
<td>(0.86)</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.1192</td>
<td>0.1068</td>
<td>0.7441</td>
<td>0.0450</td>
<td>0.2049</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.21)</td>
<td>(0.99)</td>
<td>(0.86)</td>
<td></td>
</tr>
<tr>
<td>bs-min</td>
<td>0.1335</td>
<td>0.1873</td>
<td>0.6267</td>
<td>0.1582</td>
<td>0.4895</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.00)</td>
<td>(0.88)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>emn-cnn</td>
<td>0.2132</td>
<td>0.7194</td>
<td>1.7929</td>
<td>0.0532</td>
<td>0.3072</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 reports the empirical results in terms of average annual return, average annual volatility, average drawdown, and max drawdown for each of the four portfolio strategies and the DAX as a benchmark for the time horizon 2010 - 2019. In parentheses is the p-value of the statistical difference between the performance measurement value of each strategy from that of the DAX benchmark.
Comparison of average annual return

Analyzing the average annual returns on column one of Table 2. The first point to notice is, none of the portfolio strategies achieves a statistically different annual average return from the DAX. Investigating further, the reader can see that the emn-cnn strategy achieves an outperformance compared to the DAX with an average annual return of 0.2132 while the DAX has 0.0969, which is very remarkable. This outperformance in terms of return can be investigated in more detail in Fig. 18, which displays the cumulative return of each portfolio strategy in the time horizon 2010-2019. It can be seen that the majority of outperformance by the emn-cnn portfolio is achieved in the years 2010-2019. In this period the DAX entered a sideways market with no clear direction, neither up nor down. Here one advantage of the emn-cnn strategy can be beautifully seen which is the relatively quick adaptation to new market environments compared to the other portfolio strategies.

By comparison of the min and the bs-min portfolios, the results reflect expectations. Since the bs-min shrinks its moment to the precalculated overall mean the performance in terms of return is better compared to the min portfolio. This fits in line with the statement of Jobson & Korkie (1980) that the performance of a shrunken estimator is sometimes better, but never worse the that of the original one.

![Figure 18: Cumulative returns of portfolio strategies and DAX for the time horizon 2010 - 2019](image)

A surprising result is the underperformance of the 1/N strategy compared to all other strategies. Both DAX and 1/N have very similar cumulative return trends, see Fig. 18. Also, the average annual return is nearly similar. This finding stays in strong contrast to the findings by DeMiguel et al. (2009). One explanation could deliver the DAX’s
composition. The DAX itself is composited by the market value of its inherent companies. The higher the market value of a company the bigger the share it has in the index. The reason why the return of the DAX and the 1/N portfolio is similar can be explained that the market values of the companies in the DAX do not differ that much, resulting in a relatively equal share size of each company in the DAX. This resembles the 1/N portfolio. Another explanation could be the dataset used in the paper of DeMiguel et al. (2009). They use portfolios that contain sub-portfolios as already explained in section 2.3, whereas in my paper the 1/N portfolio invests directly into stocks. This seems to result in a significant difference in performance since the 1/N portfolio investigated in the paper by DeMiguel et al. (2009) achieves a higher degree of diversification which could lead to outperformance.

For completeness, it is worth mentioning that the cumulative returns of the portfolio strategies and the DAX over the time horizon are: DAX 144%, 1/N 138%, min 214%, bs-min 227%, and emn-cnn 584%.

Comparison of average annual volatility, average drawdown, and max drawdown

Contrarily to the return values, the volatility, and drawdown values are highly significant, except for the 1/N portfolio. Regarding its average annual volatility, average drawdown, and max drawdown the 1/N portfolio is again very similar to the DAX. The reasons for these findings are the same as already mentioned above with respect to the return.

Partly expected but also interestingly, the min strategy has the lowest average annual volatility, average drawdown, and max drawdown. Partly expected because the min strategy optimizes the portfolio’s variance, exactly as it is supposed to do. Since the drawdown can be interpreted as negative volatility it is also no surprise that the average drawdown and max drawdown are low in comparison to the other strategies. Though the result is interesting, since we should expect that the bs-min has the lowest average annual volatility, average drawdown, and max drawdown. Moreover, the difference in volatility and average drawdown between the min and bs-min portfolios are highly statistically significant. Further, by comparison of the average drawdown of the min and bs-min it can be seen, most of the volatility is negative, because the average drawdown of the bs-min (0.1582) is higher than that of the min portfolio (0.0450). Those findings stay in harsh contrast to the statement by Jobson & Korkie (1980) regarding a shrunken estimator is sometimes better, but never worse. Jorion himself detects in his paper two explanations
why bayes-stein estimators might underperform. First, if the sample size is too large. Second, if there are significant mean fluctuations of the variance with the time series dataset (Jorion, 1986). It is unclear which of both reasons or a combination of both is responsible for the underperformance. Since the focus of this paper is to investigate in out-of-sample performance of different portfolio optimization strategies no further investigations are done regarding the exact reasons. Providing a list of possible reasons is considered as sufficient.

Shifting our focus back to Table 2 the reader might notice that the emn-cnn portfolio has the highest average annual volatility of all portfolio strategies. However, by comparing its volatility (0.7164) with the average drawdown (0.0532) one finds that most of the volatility is positive, because the average drawdown value is relatively low. Also, the emn-cnn has the second lowest average drawdown (0.0532), which is significantly different at a 5% level from the min portfolio which has the lowest average drawdown (0.0450). On the other hand, the emn-cnn portfolio has the second highest max drawdown with a value of 0.3072 after the bs-min portfolio with a value of 0.4895.

Overall for the performance measurements, average annual volatility, average drawdown, and max drawdown the min portfolio clearly outperforms.

Comparison of average annual Sharpe Ratio

Same as in the rest of the performance measurements, the 1/N portfolio is not statistically different in its average annual Sharpe Ratio than the DAX’s average annual Sharpe Ratio. Also in absolute values, the 1/N portfolio with an average annual Sharpe Ratio of 0.5682 is nearly the same as the average annual Sharpe Ratio of the DAX with 0.5611.

Ranks 3 and 2 are occupied by the bs-min portfolio with an average annual Sharpe Ratio of 0.6267 and the min portfolio with an average annual Sharpe Ratio of 0.7441, which are statistically indifferent from each other. This is a fascinating result from a statistical standpoint because the Sharpe Ratio is, as the reader knows, a division of the excess return and the volatility. In the findings, the average annual returns of both strategies are statistically indifferent (p-value 0.82), whereas the average annual volatilities of both strategies are statistically different (p-value 0.00).

Regarding the statistical difference between the DAX and the portfolio strategies, only the emn-cnn portfolio is statistically different. Also, only the emn-cnn’s Sharpe Ratio
(1.7929) is higher than 1, which is commonly considered a great value. This result is to no surprise since the CNN is trained to optimize the Sharpe Ratio through a corresponding asset weight allocation.

To summarize the empirical findings that are described in the last paragraphs. The emn-cnn portfolio offers the best tradeoff between return and risk reflected by a Sharpe Ratio of 1.7929. Considering that most of the volatility is positive because it has the second lowest average drawdown in this time horizon, the emn-cnn portfolio strategy is seen as the dominant portfolio optimization strategy. However, high risk-averse investors who prefer low portfolio fluctuation should choose the min portfolio strategy, since it has the lowest average annual volatility, average drawdown, and max drawdown in the time horizon 2010-2019.

4.2 Empirical findings for the time horizon 2010 – 2022

As the reader might remember, the longer time horizon from 2010 – 2022 brings more uncertainty into the stock market because of COVID-19 acting as an exogenous negative productivity shock spreading its consequences in the years from 2020 until 2022 in the German economy. Leading to a higher unemployment rate, a contracting GDP, and a change in ECB’s monetary policy with a higher ECB’s funds rate and the start of quantitative easing. All these factors influence the German stock market to the down and upside resulting in a higher stock market fluctuation. Table 3 reports the empirical findings for the time horizon 2010 – 2022.
Table 3 reports the empirical results in terms of average annual return, average annual volatility, average drawdown, and max drawdown for each of the four portfolio strategies and the DAX as a benchmark for the time horizon 2010-2022. In parentheses is the p-value of the statistical difference between the performance measurement value of each strategy from that of the DAX benchmark.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Return</th>
<th>Volatility</th>
<th>Sharpe Ratio</th>
<th>Avg. Drawdown</th>
<th>Max Drawdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAX</td>
<td>0.0860</td>
<td>0.1517</td>
<td>0.5276</td>
<td>0.0667</td>
<td>0.3272</td>
</tr>
<tr>
<td>1/N</td>
<td>0.0785</td>
<td>0.1466</td>
<td>0.5171</td>
<td>0.0734</td>
<td>0.3467</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.51)</td>
<td>(0.98)</td>
<td>(0.09)</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>0.1390</td>
<td>0.1242</td>
<td>1.0076</td>
<td>0.0472</td>
<td>0.2810</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.00)</td>
<td>(0.20)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>bs-min</td>
<td>0.1857</td>
<td>0.2004</td>
<td>0.8424</td>
<td>0.1556</td>
<td>0.4895</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.40)</td>
<td>(0.40)</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>emn-cnn</td>
<td>0.2213</td>
<td>0.7056</td>
<td>1.6563</td>
<td>0.0448</td>
<td>0.3072</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td></td>
</tr>
</tbody>
</table>

Comparison of average annual return

As expected due to the macroeconomic events, the average annual return of the DAX deteriorates in this longer time horizon from 0.0969 to 0.0860. Also, a deterioration happens for the 1/N portfolio compared to the shorter time horizon. But also for the time horizon 2010–2022 there is no statistical difference between the average annual returns of the 1/N portfolio and the DAX detectable. For the remaining three portfolio strategies the deterioration in return is not reflected. This can be traced back to the short positions the three strategies are allowed to open, whereas the 1/N portfolio only can open long positions. Compared to the short time horizon the emn-cnn has in this longer time horizon a statistically significant different average annual return of 0.2213 compared to the DAX of 0.0860, which is again very remarkable. The overall ranking of the portfolio strategies in the performance category return stays the same as compared to the time horizon 2010-2019, with emn-cnn portfolio being ranked 1, followed by bs-min, min, and 1/N portfolios. Figure 19 visualizes the cumulative returns for the time horizon 2010-2022 for each portfolio strategy and the DAX.
All, except of one portfolio strategies show a drop during the COVID-19 crash. The DAX and 1/N are more mature, and the min and bs-min are less mature. The emn-cnn strategy shows no impact of the COVID-19 crash, quite the opposite it significantly increases its portfolio value during the crash, see Fig. 19. This shows again the advantage of the CNN to adapt quickly to new market conditions. However, not only the emn-cnn, but also the min and bs-min portfolios took advantage after the COVID-19 crash by a strong increase in their return, see also Fig. 19. One reason to explain the strong increase in return is after the COVID-19 crash there is a stock market rally with low volatility, making it easier for the min portfolio to optimize portfolio weights for an outperformance compared to the DAX. Since the bs-min portfolio shrinks the estimation error of the min portfolio it even outperforms the min portfolio besides the DAX, visualized in Fig. 19. Despite the harsh economic environment all strategies are able to increase average annual return over the years 2020 to 2022. Resulting in a cumulative return in the time horizon from 2010 until 2022 for the DAX of 168%, the 1/N portfolio of 143%, the min portfolio of 454%, the bs-min portfolio of 793%, the emn-cnn portfolio of 1284%.

Comparison of average annual volatility, average drawdown, and max drawdown

Due to the uncertainty in the stock market the volatility of nearly all portfolio strategies increases as expected, the same happens to the average drawdowns. Regarding the average annual volatilities, the min portfolio has with 0.1242 the lowest average annual volatility, followed by the 1/N portfolio with an average annual volatility of 0.1466. In third place is the bs-min portfolio with an average annual volatility of 0.2004.
The emn-cnn portfolio brings up the rear with an average annual volatility of 0.7056. For comparison, the DAX has an average volatility of 0.1517. Compared to the shorter time horizon and besides the 1/N portfolio, also the bs-min portfolio is now statistically indifferent compared to the DAX. Leaving only the min and emn-cnn portfolios with statistically different average annual volatilities compared to the DAX. Regarding the average drawdown as in the shorter time horizon also in the longer time horizon the average drawdown values for min, bs-min, and emn-cnn portfolios are highly significantly different compared to the DAX’s average drawdown. For the time horizon 2010 -2022 the average drawdown for the 1/N portfolio 0.0734 is also statistically significantly higher compared to the DAX 0.0667 at a 10% level. For the remaining three portfolio strategies the average drawdown in this time horizon does not differ substantially compared to the shorter time horizon, compare Table 3 and 2. However, for this time horizon, the emn-cnn portfolio has the lowest average drawdown with 0.0448. Regarding the max drawdown nothing with respect to the ranking of the portfolio strategies changed. The min portfolio has still the lowest drawdown even though it increased from 0.2049 to 0.2810 in the longer time horizon. A further observation is that the bs-min and emn-cnn portfolios have their max drawdowns before the years 2020 – 2022, because their max drawdown values do not change from Table 2 to Table 3. Whereas the max drawdowns of the DAX, 1/N portfolio, and min portfolio deteriorate.

For the time horizon 2010 -2022 the outperformance of the min portfolio is not as clear as it is for the time horizon 2010 – 2019. Being only in the top ranking for the lowest annual average volatility and lowest max drawdown. The top ranking for the lowest average drawdown has to be surrendered to the emn-cnn portfolio. Whereby, the difference in average drawdown values between the min and em-cnn portfolios holds at a 5% significance level.

Comparison of average annual Sharpe Ratio

The ranking of the portfolio strategies does not change for the time horizon 2010 – 2022 compared to the time horizon 2010 – 2019. The lowest average annual Sharpe Ratio has the 1/N portfolio with 0.5171 and the highest average annual Sharpe Ratio has the emn-cnn portfolio with 1.6563. Also, statistical significance draws the same picture in the long time horizon as it has in the short time horizon as only having the emn-cnn portfolio a statistically significant higher Sharpe Ratio compared to the DAX. Table 3 contains, despite being statistically insignificant, a second portfolio that has an average
annual Sharpe Ratio higher than 1, which is the min portfolio with 1.0076. The min and bs-min portfolios can increase their average annual Sharpe Ratios. The increase can be traced back to a relatively higher increase in average annual returns compared to the average annual volatilities. A sophisticated reader might observe one further interesting detail. The emn-cnn portfolio is able to increase its average annual return while decreasing at the same time its average annual volatility for the time horizon 2010 – 2022. But why is the average annual Sharpe Ratio deteriorating in the longer time horizon compared to the shorter time horizon? The reason for deterioration is an increase in ECB’s funds rate because of quantitative easing in order to fight the negative consequences of COVID-19, see Fig. 16. This leads to an increase of the risk-free rate, which is part of the Sharpe Ratio calculation. Since the average annual Sharpe Ratio of the emn-cnn portfolio is deteriorating from time horizon 2010 – 2019 to 2010 – 2022 it can be concluded, that the increase (decrease) in average annual return (volatility) is not sufficiently large enough to compensate for the increase in risk-free rate.

Equally to the short time horizon, in long time horizon the emn-cnn portfolio offers the investor the highest average annual Sharpe Ratio, thus the best tradeoff between return and risk. On the other side, for the time horizon 2010 – 2022 the question is hardly to answer which portfolio strategy a risk-averse investor might choose. For an investor who wants to have a stable portfolio strategy with neither positive, nor negative fluctuations and a low max drawdown the min portfolio might be suitable. For an investor that allocates a higher priority to return and average downside risk, the emn-cnn portfolio might be a better choice.

4.3 Comparison of results with existing literature

Note, the results in performance measurement values cannot be compared directly to the existing literature, since the methodology and, or the datasets used by the literature are different. Thus it is obvious that the results are different. However, there might be similarities in the constellation of the empirical findings of the existing literature and this paper. Exactly these constellations are analyzed in this part. As already mentioned, this paper gets its basic concepts and methodology from the paper “Optimal Versus Naïve Diversification: How inefficient is the 1/N Portfolio Strategy?” by DeMiguel et al. (2009) and from the paper “Portfolio management system in equity market neutral using
reinforcement learning” by Wu et al. (2021), therefore the empirical findings are compared to these papers.

Starting the comparison with the paper by DeMiguel et al. (2009). The same performance measurement used by their paper and this paper is the Sharpe Ratio. Hence the focus of result comparison lies there. For completeness, the first point to name: The 1/N portfolio in the paper by DeMiguel et al. (2009) outperforms the min and bs-min portfolios, which is not congruent with the findings here in my paper. However, DeMiguel et al. (2009) constrained versions of the min portfolio achieve outperformance over the 1/N portfolio. Indicating even in their paper a possible superiority of the min portfolio over the different datasets by introducing further constraints. Let us stick to superiority, the same picture can be observed regarding the ranking of the min and bs-min portfolios in the paper by DeMiguel et al. (2009). Meaning the min portfolio outperforms the bs-min portfolio on average over all datasets and time horizons in their paper which is congruent with the findings here in my paper. In fact also the findings by DeMiguel et al. (2009) stay again in contrast to the statement about the at least equality or even superiority of a shrinkage estimator made Jobson & Korkie (1980). Unfortunately, DeMiguel et al. (2009) do not do a deep dive into the reasons why the bs-min portfolio underperforms. The only explanation they give is that the resulting portfolio weights from the Bayes-Stein shrinkage estimator are more likely to be shrunken to the out-of-sample min portfolio than to the theoretical optimal portfolio weights. Despite these mentioned points, no more similarities in empirical findings can be found between the paper by DeMiguel et al. (2009) and my paper.

Continuing the empirical result comparison with the paper by Wu et al. (2021) which offers more opportunities to compare constellations of performance measurements, because they use as performance measurements Return, Sharpe Ratio, and Max Drawdown. However regarding the used strategies, the paper only tests the emn-cnn strategy for out-of-sample performance. In the paper by Wu et al. (2021) the EMN portfolio strategy implemented by a CNN and with a Sharpe Ratio as a reward function outperforms all other portfolio strategies in terms of Return and Sharpe Ratio. This is congruent with the findings of sections 4.1 and 4.2. Regarding the max drawdown, the emn-cnn portfolio is not able to achieve the lowest max drawdown in their paper and in my paper. In the paper by Wu et al. (2021) another so-called recurrent neural network with a Sharpe Ratio as a reward function achieves the lowest max drawdown, whereas in
my paper the lowest max drawdown in both time horizons is achieved by the min portfolio. Therefore it can be assumed that a CNN with the Sharpe Ratio as a reward function leads to a high return with a high Sharpe Ratio at the same time. Whereby in their paper and my paper most of the exhibited volatility is positive. This can be derived from the fact that the corresponding max drawdown value is relatively low. Since the lack of similar investigated portfolio strategies between the paper by Wu et al. (2021) and my paper further comparison of empirical findings cannot be done.

To summarize the findings of my paper with the ones in the existing literature one can state that most of them are congruent. The only exception in the findings by the existing literature and this paper is the posteriority of the 1/N portfolio compared to the other portfolio strategies.

5. Conclusion

The paper enriches the existing literature by bringing together and evaluating the out-of-sample performance of different portfolio optimization concepts starting with the MPT over a naïve approach and ending in the present day in which neural networks are on the rise. All under the umbrella of modern portfolio strategies. A further contribution is the investigation of the performance of the portfolio strategies by using German stocks to construct optimal portfolios and comparing them to the DAX as a benchmark. This is done in two time horizons. The first one, from 2010 -2019, reflects a growing German economy and therefore a steadily increasing German stock market. The second one, from 2010 – 2022, adds more uncertainty through the years from 2020 – 2022 due to COVID-19 with its short and long-term consequences. Hereby the goal is to check how the portfolio strategies respond to a changing market environment.

Furthermore, the paper is able to answer the following research questions:

(i) Do any of the portfolio strategies presented in this paper achieve a higher annual return than the DAX as a benchmark?

For both time horizons (2010 -2019 and 2010 – 2022) the min, bs-min, and the emn-cnn portfolios are able to outperform the DAX as a benchmark in terms of
average annual return. However, only the emn-cnn portfolio outperforms the DAX statistically significantly with an average annual return of 0.2132 and 0.2213 for the time horizons 2010-2019 and 2010 – 2022, respectively.

(ii) Do any of the portfolio strategies presented in this paper achieve a lower annual volatility than the DAX as a benchmark?

The min portfolio with an average annual volatility of 0.1068 and 0.1242 in the shorter and longer time horizon turns out to be the winner, also on a statistically highly significant level of 1%. This is a surprising result, since the bs-min portfolio is expected to outperform the min portfolio, because it uses the concept of a shrinkage estimator for the variance-covariance matrix. According to the statement made by Jobson & Korkie (1980) the performance of a shrunken estimator is sometimes better, but never worse than that of the original one. The underperformance can be explained by Jorion (1986). First, the underperformance can come from, when the sample size is too large or second when there are significant mean fluctuations of the variance within the time series dataset. Regarding average annual volatility, in both time horizons the 1/N portfolio is able to secure the second lowest variance. However, it is not statistically different from the average annual variance of the DAX. Rank 3 and 4 occupy the min and emn-cnn portfolios.

(iii) Do any of the portfolio strategies presented in this paper achieve a higher annual Sharpe Ratio than the DAX as a benchmark?

With an average annual Sharpe Ratio of 1.7929 and 1.6563 and also a high significance level of 1% the emn-cnn portfolio clearly outperforms the DAX (0.5611 and 0.5276) for the time horizons 2010 – 2019 and 2010-2022, respectively. This is a not surprising result since the CNN is trained to deliver an as high as possible Sharpe Ratio and hence correspondingly optimizes the portfolio weights to achieve this goal. All other portfolio strategies are not able to statistically outperform the DAX in terms of Sharpe Ratio.

(iv) Do any of the portfolio strategies presented in this paper achieve a lower Drawdown than the DAX as a benchmark?
Regarding the max drawdown, the min portfolio has the edge in both time horizons. With a max drawdown of 0.2049 for the time horizon from 2010-2019, deteriorating in the second time horizon from 2010 - 2022 to 0.2810. This deterioration lies in expected limits, because the years from 2020 - 2022 add uncertainty to the German economy resulting a stock market drops that the portfolios must face. Regarding the average drawdown over the two time horizons, the answer to the question which portfolio is best can only be given by splitting it up into the time horizons. For the time horizon 2010 -2019 the min portfolio has on average the lowest drawdown of 0.0450, which is in line with the title for the lowest max drawdown. Also, the result is highly statistically significant. For the time horizon 2010 - 2022 the lowest average drawdown has the emn-cnn portfolio with 0.0448, also highly statistically significant. This is an indication that the emn-cnn portfolio adapts quicker to a changing stock market environment. Comparing the findings of the average annual volatility and average drawdown between the min and emn-cnn portfolios, the min portfolio has the lower average annual volatility. Whereas most of the emn-cnn’s volatility is positive, because the average drawdown is relatively low in comparison.

By analyzing the empirical findings in general, the min and emn-cnn portfolios are the two top-performing portfolios over the chosen performance measurements and time horizons. On the one hand, the emn-cnn portfolio is more tuned towards having a higher return for a given relatively small volatility with low downside risk in both time horizons. On the other hand, the min portfolio is optimized having a low average annual volatility and max drawdown in both time horizons.

At this point one remark on the performance of the 1/N portfolio. Investors should be aware that the outperformance of the 1/N portfolio stated by DeMiguel et al. (2009) might be given if the portfolio consists of sub-portfolios each inheriting further stocks. However, a clear posteriority of the 1/N portfolio compared to the other implemented portfolio strategies is given here in this paper by applying the 1/N portfolio to individual stocks.

Despite the promising results found in this paper it is necessary to mention some critical points. The first one is, that no transaction costs are incorporated in all four implemented portfolio strategies. An incorporation could change the performance results
drastically depending on how often the calculated optimal portfolio weights lead to a reallocation within the investor’s portfolio. In addition, the portfolio strategies in this paper rely on stock prices to calculate the optimal portfolio weights, thus the data quality of the stock prices must be at its highest. If not, it can lead to severe misallocations by non-optimal portfolio weights resulting in bad portfolio performance. A very important assumption this paper makes is that stocks are partially purchasable. An investor must be aware by implementing any of the portfolio strategies from this paper that all of them use partial stock positions for their optimal portfolio weight allocations. None of them is optimized to calculate optimal portfolio weights under the constraint, that only whole numbers of shares are purchasable. Fortunately, many brokers nowadays offer partial stock purchases, therefore the investor can choose an appropriate one. At the end one simple but critical question remains: Which of the proposed portfolio optimization strategies in this paper will be the likeliest implemented by investors in practice? This question is hard to answer, because it depends on the type of investor. A fund manager is more likely to implement the emn-cnn strategy although it is mathematically and technically more complex, because she has the needed knowledge, infrastructure, and time available. A retail investor, especially if she stands at the beginning of her investment journey, is more likely to implement the 1/N portfolio or just buys an ETF of the DAX. Out of her perspective this is simpler and does not require any mathematical or technical background knowledge.

Let me finish this paper by identifying further investigation possibilities. The first further investigation possibility is to use instead of the minimum-variance approach a minimum-semivariance approach as proposed in the paper by Estrada (2015). Briefly explained, the semivariance by Estrada (2015) incorporates negative variance into the variance-covariance matrix, but lets positive variance values be zero entities in the matrix. The results in a new way to evaluate the relative riskiness of assets, which resembles more the attempt to reduce the average drawdown of the portfolio. Such an approach might be very interesting because the empirical findings from this paper conclude that strategies with the lowest average or max drawdown (min or emn-cnn) perform best. In this context it would be interesting to test the minimum-semivariance against the minimum-variance portfolio and the emn-cnn portfolio that is conditioned to optimize its drawdown instead of its Sharpe Ratio. Another further interesting investigation possibility would be to test the proposed portfolio strategies on a higher and lower timeframe, for example monthly and daily, to investigate in the performance changes of each strategy depending on the
timeframe. Similar to a change in timeframe, the stock market can also be changed by applying the four portfolio strategies to the US equity market or an Asian equity market and comparing the findings to the findings of this paper. Lastly, a more technical investigation possibility could be to clearly determine the reason why the bs-min portfolio underperforms the min portfolio.


