Bachelor's Thesis

Ranking Financial Assets

Department of Statistics
Ludwig-Maximilians-Universität München

Jonas Schernich

Munich, Aug. 8th, 2023

Submitted in partial fulfillment of the requirements for the degree of B. Sc.
Supervised by Prof. Daniel Wilhelm, PhD

Abstract

This Bachelor's Thesis delves into uncertainty within portfolio rankings by using both marginal and simultaneous confidence sets based on Mogstad et al. (2020). A range of portfolio strategies, including factor, low beta, and leverage strategies, are ranked. Focus lies on investigating the occurrence of autocorrelation within stock market time series data and the potential ramifications on confidence set construction.

The investigation into various methodologies for constructing a covariance matrix from mean values across different time series under conditions of autocorrelation revealed a complexity. Specifically, accurately estimating confidence intervals becomes intricate when predicting rankings based on expected values in the presence of autocorrelation.

The thesis elucidates the presence of significant ambiguity in the practice of portfolio management back testing, thereby highlighting the fact that rankings based on historical performance do not guarantee superior performance in the future. The leveraged strategy and its gold-incorporating variant were the only portfolios identified as consistently surpassing the performance of the S&P 500 throughout all analyzed time periods. Even promising portfolio theory breakthroughs like factor investing did not significantly outperform the S&P 500 in any investigated time period.

Contents

1 Introduction

Statistical analysis plays a crucial role in various fields, including understanding financial markets and portfolio management. Fundamental concepts like the Capital Asset Pricing Model (CAPM) (Jagannathan et al., 1995), factor models (Fama and French, 1993), and value at risk (VaR) (Jorion and Philippe, 2007) are based on relatively simple statistical methods. However, also more advanced statistical concepts like GARCH models, introduced by Engle (1982) are ubiquitous in econometrics.
According to Thompson (2012)(i, ii) and Draper (1995)(iii), the sources of ranking uncertainty include sampling variability(i), measurement error(ii) and model assumptions(iii). Also subjective judgments effects rankings as elaborated in Statistical uncertainty in Rankings.
Rankings are widely used in finance and economics as they provide an easily understandable rating of populations based on a specific feature, relative to the ratings of populations (Mogstad et al., 2020). Rankings of GDP, portfolio returns, and economic freedom, among others, are commonly used and accepted. However, the uncertainty of rankings is often overlooked, even though many rankings are based on estimates involving statistical uncertainty.

Addressing this issue, Mogstad et al. (2020) have developed an inference procedure to account for uncertainty in rankings. The procedure incorporates marginal and simultaneous confidence sets to provide estimators with a measure of uncertainty. It allows for the estimation of such a confidence set for each rank in a ranking. The authors of the paper applied this new procedure to rankings of the Program for International Student Assessment (PISA) and rankings of neighborhoods in the USA.

In this bachelor thesis, my objective is to utilize the procedure developed Mogstad et al. (2020) to investigate the reliability of rankings regarding expected returns of various investment portfolios. By applying this procedure to the rankings of investment portfolios, I will analyze and quantify the statistical uncertainty associated with these rankings.
The goal is to reevaluate common strategies used in investment portfolios, such as factor investing, low beta investing, evolutionary theories, index investing or leverages strategies. Some of them have been stated to outperform the market historically. A Hypothesis that shall be reviewed using confidence sets. Additionally, I will examine variations in reliability across different time periods.

2 Statistical uncertainty in Rankings

2.1 Introduction to Uncertainty

Rankings provide us with a hierarchical order of units based on certain attributes or criteria. However, how confident can we be in interpreting these rankings? How significantly can the rank order be influenced by different sources of uncertainty? Taking into consideration the uncertainty in rankings is crucial to avoid misinterpretations and to draw sound conclusions. In this section, we will investigate the sources of uncertainty in rankings and discuss methods to quantify and deal with this uncertainty.
Some of the most relevant sources of uncertainty in rankings are

- **Sampling Variability:**
  Rankings are often based on samples, whether it's a survey of opinions or a subset of a larger population. Sampling variability arises due to random variations in the sample selection, leading to potential differences if another sample was selected. This variability can impact the ranking order and introduce uncertainty (Thompson, 2012).

- **Measurement Error:**
  In many rankings, measurements or data are collected to assess the performance or characteristics of the ranked entities. However, measurement error can occur due to various factors such as instrument inaccuracies, human subjectivity, or data collection errors. These errors introduce uncertainty by affecting the reliability and validity of the measurements, consequently impacting the rankings (Thompson, 2012).

- **Model Assumptions:**
  Rankings may be derived using mathematical or statistical models that make certain assumptions about the data or underlying relationships. These assumptions can introduce uncertainty if they do not perfectly align with the reality of the ranked entities. Deviations from the assumptions can introduce uncertainty throughout the ranking process (Draper, 1995).

- **Subjective Judgments:**
  Some rankings involve subjective judgments or expert opinions, such as peer evaluations or expert panels. These subjective judgments introduce a level of uncertainty as different experts may have varying perspectives or criteria for evaluation. The subjective nature of these judgments can lead to fluctuations in rankings and inherent uncertainty.

  Having discussed the multiple sources of uncertainty in rankings, it becomes evident that quantifying and dealing with this uncertainty is crucial. The following section will delve into some methods used to handle such uncertainties.

2.2 Dealing with Uncertainty

Over the last decades, several methods to quantify and deal with such uncertainty in rankings have been developed. Major ones being the bootstrap and stochastic approaches. However, they have distinct drawbacks that must be considered.

The Bootstrap
The Bootstrap is a resampling method introduced by Efron (1979). It estimates the distribution of a statistic (such as a mean or a rank) by drawing many samples (with replacement) from the original data sample and then calculating the statistic for each of these new samples. This procedure offers a way to quantify the uncertainty of estimates, yet it is computationally intensive and not specifically designed for ranking lists.
While bootstrap is a useful tool for reducing uncertainty, it has not been developed specifically for the use in rankings. Applying it in rankings can lead to additional complexity, adding to one of the main disadvantage of bootstrap, which is its computational intensity(Sauerbrei and Schumacher, 1992). In addition, if applied to rankings, the bootstrap method tends performs poorly if the ranked populations ranking attributes are very close to each other(Romano and Shaikh, 2012). Specifically for ranking populations this was also emphasized by Mogstad et al. (2020), who showed that in rankings with more than two populations, bootstrap did not meet the coverage requirements.

Stochastic Approaches
Stochastic approaches, such as the Bayesian method (Gelman et al., 2013) and Monte Carlo simulations (Metropolis and Ulam, 1949), leverage random variations or probabilities to make statistical estimates and predictions. The Bayesian method updates the probability of a hypothesis based on new data, while Monte Carlo simulations draw random samples from a probability distribution to obtain numerical results. These methods are flexible and can accommodate uncertainty, yet they often require strong assumptions and can be computationally intensive.
Despite the availability of various methods to manage uncertainties in rankings, challenges persist. One of the approaches that seek to address these challenges is the inference procedure for ranks developed by Mogstad et al. (2020).

2.3 Inference on Ranks

In the paper *"Inference for Ranks with Applications to Mobility across Neighborhoods and Academic Achievement across Countries"*, (Mogstad et al., 2020) proposes a robust methodology to quantify uncertainty in rankings that are typically computed using estimated values. This methodology, which requires low computational complexity and no tuning parameters, is utilized to construct confidence sets for the rank of each population, effectively incorporating uncertainty into these rankings.
The authors elaborate on two types of confidence sets: marginal and simultaneous. Marginal confidence sets help elucidate uncertainty regarding the rank of a particular population, whereas simultaneous confidence sets are designed to tackle uncertainty when questioning the ranks of all populations collectively. How these confidence sets are built is explained in great detail in section three of Mogstad et al. (2020).

To underscore the practicality of their method, Mogstad et al. re-examine the rankings of U.S. neighborhoods based on intergenerational mobility, and developed countries based on academic achievement (PISA). By computing both marginal and simultaneous confidence sets for the ranks of these populations, they assess the robustness of the rankings and offer insights into the associated uncertainty.

Different types of confidence sets answer distinct economic questions. Marginal confidence sets are instrumental in determining if a specific population is among the best or worst ranked, while simultaneous confidence sets provide a broader understanding of geographic patterns and overall rankings. Mogstad et al. also discuss how to construct confidence sets to identify the top or bottom populations with a certain level of confidence, which they call τ-best or τ-worst populations.

In this thesis, the goal is to apply the marginal and simultaneous confidence sets proposed by Mogstad et al. to answer an economic question. These confidence sets will be calculated for econometric rankings of investment portfolios, focusing mainly on stocks.

3 Portfolio Theory and Research

3.1 Indices

Stock indices serve as a means to monitor and assess the performance of individual stocks or the broader stock market. They consolidate the prices or values of a specific group of stocks into a single metric, providing a snapshot of market performance. Typically, these indices are developed and maintained by financial institutions, stock exchanges, or independent index providers.

The primary role of stock indices is to serve as benchmarks or reference points for evaluating the performance of investment portfolios, individual stocks, or other investment products. They enable investors to gauge the performance of their investments relative to the overall market or a specific market segment. Through the comparison of investment returns to a designated benchmark, investors can evaluate the competence of portfolio managers or the effectiveness of their investment strategies.

The construction of indices employs various methodologies, chosen based on the index's objectives and structure. The most prevalent approach is market capitalization weighting, sticking to the later introduced factors, this would be a one factor model, whereby the weight of each stock in the index is determined by its market value relative to the index's total market value. Other methodologies include equal-weighting, wherein each stock carries an equal weight in the index, and price-weighting, which assigns weights based on stock prices.
(Carhart (1997), Asness et al. (2013) and Fama and French (1992))

3.2 Portfolio Theorie

3.2.1 Introduction to Portfolio Theory

Since Harry Markowitz (Markowitz, 1952) laid the foundations of portfolio theory in the 1950s, the field has rapidly evolved and numerous different concepts have been developed to construct an optimal portfolio. However, even the meaning of an optimal portfolio is not clear, as investors and their objectives are heterogeneous.

Generally speaking, returns should be high while risk should be low, creating a trade-off. Thus, managers have attempted to combine uncorrelated positions in their portfolios to reduce volatility. Portfolios are optimized based on different indicators such as beta, Value at Risk (VaR), or the Sharpe Ratio (Bodie et al., 2013). After the development of these indicators started in the 1960s, the field continued to emerge.

In particular, since Eugene Fama's publication on the Efficient Market Hypothesis in 1970 (Fama, 1970), investing in indices that broadly reflect market returns has become increasingly popular. Building upon this work, Fama and French (1992) published their research on factor premiums in the 1990s, leading to the creation of factor portfolios. Since then, various active trading strategies and new concepts like behavioral finance and evolutionary portfolio theory have further disrupted portfolio management, many of which are somewhat contradictory, like momentum strategies and the evolutionary portfolio strategy.

To compare the performance of portfolios following different concepts of portfolio theory, I build basic portfolios, each created based on a different concept of portfolio theory. These portfolios will be ranked alongside major indices and gold to determine their outperformance and compare the certainty of their performance. To maintain comparability, all portfolios will largely consist of stocks, or leveraged products that are based on stocks portfolios. The main benchmark is the S&P 500.

3.2.2 Risk

The interpretation of risk within the framework of portfolio theory is a fundamental aspect that requires thorough understanding. A range of metrics such as beta (Fama and French, 2004), volatility (Kim et al., 1998), and maximum drawdown, among others, are typically used to quantify the risk associated with a portfolio. Essentially, these metrics aim to measure the likelihood of the short-term return falling below the expected return. However, the relevance of these risk measures to investors is highly contingent on the duration of their investment (Beck, 2021). In the context of a long-term investment strategy, the decision to accept a diminished expected portfolio return for the sake of mitigating short-term risk may not necessarily align with the investor's objectives. This raises the question of whether a focus on default risk ought to be the primary concern for long-term portfolios. It is important to note that within the scope of this thesis, any reference to 'risk' pertains to the likelihood of achieving a short-term return that falls below the anticipated return.

3.3 Strategies

3.3.1 Market cap weighted Portfolios

Following the work by Sharpe (1991), the strategy of broad market participation, as an alternative to individual stock selection or the use of investment vehicles such as funds, has seen an increased popularity. The seminal paper by Sharpe discusses the paradigm that the expected returns of selective investing are commensurate with the average market return, albeit accompanied by an elevated risk attributable to inadequate diversification. This revelation precipitated the advent of Exchange Traded Funds (ETFs), engineered to achieve maximal diversification. ETFs typically adopt a market capitalization weighting strategy, often benchmarked against popular indices of regional stocks.
An alternative approach to index weighting is the 'equal weight' strategy, which maintains equilibrium in the portfolio value of each constituent stock through regular rebalancing. This method has the potential to yield superior returns and mitigate risk, particularly by preventing disproportionate exposure to stocks experiencing inflated valuations, such as during the Japanese asset price bubble of the late 1980s (Malladi and Fabozzi, 2017). However, due to limitations in the availability of historical data and the prevalence of value-weighted indices, the focus of the ranking in this thesis will be confined to portfo-

lios weighted by market capitalization.

The following market cap weighted indices will be included:

- S&P 500
- FTSE 100
- DAX
- Nikkei 225
- Hang Seng
- NASDAQ 100
- MSCI Emerging Markets
- MSCI World
- MSCI Europe

3.3.2 Low Beta

The low beta strategy is predicated on the beta indicator (Fama and French, 2004), a concept rooted in the CAPM (Sharpe, 1964). The beta coefficient serves as a comparative measure of a portfolio's volatility relative to its designated benchmark. The specific definition of beta is as follows:

$$\beta_i = \frac{\text{cov}(r_i, r_M)}{\sigma_M^2} \tag{1}$$

where r_i is the portfolio return and r_M is the benchmark return.

In line with the factor premiums identified by Fama and French (1992), it has been observed that portfolios demonstrating lower beta values tend to generate superior alpha values and Sharpe ratios (Frazzini and Pedersen, 2014). As a result, the pursuit of lower beta has become a prevalent strategy in portfolio management.
One approach to crafting a low-beta portfolio involves the selection of stocks that have historically exhibited a beta value lower than the reference benchmark. The constructed low-beta portfolio in this analysis follows:

Stock	weight
Coca-Cola Company	10%
Procter & Gamble Company	10%
Johnson & Johnson	10%
PepsiCo Inc.	10%
Verizon Communications Inc.	10%
Walmart Inc.	10%
McDonald's Corporation	10%
Consolidated Edison Inc.	10%
Duke Energy Corporation	10%
Kimberly-Clark Corporation	10%
Total	100%

Benchmarked to the S&P 500, the beta of this portfolio is 0.58.

3.3.3 Growth

The growth strategy, such as the S&P 500 growth, is an investment approach that focuses on selecting stocks or assets with the potential for significant capital appreciation. It involves investing in companies that are expected to demonstrate above-average revenue and earnings growth rates compared to the broader market. Growth investors prioritize a company's future prospects and accept higher volatility in exchange for the potential for long-term returns (Cronqvist et al., 2015). S&P 500 Growth defines a growth index by overweighting stocks by the following factors: sales growth, the ratio of earnings change to price, and momentum (*SP500 Growth*, n.d.).
The following growth index will be included:
S&P 500 Growth

3.3.4 Small Cap Value

In 1992 Fama and French introduced the the size and value factor and thus added to the previously common one-factor model of the CAPM. Since then, there has been lots of debate about the significance of the size premium due to its inconsistency and even claims it might have disappeared after its discovery, which would fit into the efficient market hypothesis (Malkiel, 2003). However, new research by Asness et al. (2018) indicates that the the small cap prime indeed exists, if the portfolio is controlled for junk. Such a control for junk can be reached by combining the value and small cap factor, as the S&P Small Cap Pure Value does.
The following Small Cap Value index will be included:
S&P Small Cap Pure Value

3.3.5 Value

A value index, such as the S&P 500 Value, tracks the performance of stocks considered undervalued based on fundamental factors. These stocks are perceived to have lower prices relative to their intrinsic value. Academic research examines the characteristics and performance of value stocks and value indices to assess their potential for generating excess returns compared to other investment strategies. S&P defines a growth index by overweighting stocks by the following three factors: the ratios of book value, earnings, and sales to price (*SP500 Value*, n.d.).
The following value index will be included:
S&P 500 Value

3.3.6 Evolutionary Portfolio Theory

Evolutionary Portfolio Theory, put forth by Hens and Schenk-Hoppé (2001), extends the traditional portfolio selection model proposed by Markowitz (1952) and later extended by Sharpe (1964) and Lintner (1965). While Markowitz's model focuses on mean and variance to assess return and risk, the Evolutionary Portfolio Theory incorporates ideas from evolutionary game theory and emphasizes market interaction.

According to this approach, portfolio strategies compete for market capital, and the endogenous price process acts as a market selection mechanism. The equilibrium concept provided by the Evolutionary Portfolio Theory is a distribution of market capital (wealth shares) that remains invariant under the market selection process. Simple trading strategies, referred to as portfolio rules, compete for dominance, and the unique evolutionary stable strategy is identified as the one that can resist mutations and outperform other strategies (Hens and Schenk-Hoppé, 2001).
The recommended evolutionary stable strategy suggests dividing wealth proportionally to the expected relative returns of assets.
Hens claims that the strategy of the Global Portfolio One (GPO) is according to the evolutionary portfolio theory (Hens, 2021). The Portfolio used for this ranking to represent the evolutionary portfolio is structured similarly. The portfolio consists partly of equities and partly of default-proof bonds, with the base distribution of 20% bonds and 80% stocks.
The stock portfolio is world-wide diversified, the bonds are 10 year US treasuries only, due to data availability and simplicity.
To implement the evolutionary strategy, the portfolio share of stocks is anti cyclically raised to up too 120% in two steps.
80% → 100% → 120%
Compared to the GPO approach, the application of leverage serves to further enhance the countercyclical component of the portfolio. To keep it simple, the interest rate for the leverage is equal to the 10 year treasury yield. Thus, the three Regimes, as they are referred to in the GPO, are as follows:

	Regime 1	**Regime 2**	**Regime 3**
MSCI World	70%	87%	105%
MSCI EM	10%	13%	15%
Fixed Income	20%	0%	-20%
Total	100%	100%	100%

Table 1: Three Regimes of the Evolutionary Portfolio Strategy

Upon a 20% decline from the all-time high of the portfolio, a transition is triggered from Regime 1 to Regime 2. Should the portfolio suffer a further loss, reaching a total of 40% below its all-time high, a subsequent shift from Regime 2 to Regime 3 is occurs. When a new all-time high is achieved, the portfolio reverts back to the initial state, Regime 1.

3.3.7 Leverage for the long run + Gold

The last two portfolio strategies are leveraged strategies based on Bilello and Gayed (2016). Focusing on the S&P 500, they found that daily leverage strategies can enhance performance if used in low volatility environments. The authors found that these patterns primarily occur during periods when the S&P 500 index exceeded its 200-day moving average (SMA200). As a result, they proposed an investment strategy that involves maintaining a daily leveraged position in the S&P 500 when the index is above its SMA200 and holding cash when it falls below.
The ranking will include two portfolios based on this strategy. The first portfolio will be constructed as intended by Bilello and Gayed, while for the second strategy, I will add another component I combined this strategy with a leverage strategy using Gold during bear markets, as gold tends to act as a hedge against them (Baur and Lucey, 2010). This makes the portfolio strategy only slightly more complicated.
To adhere to Beck's terminology of 'Regimes', the portfolios are structured as follows:

	Regime 1	**Regime 2**
S&P 500 3x Lev.	100%	0%
Gold/Cash	0%	100%
Total	100%	100%

Table 2: Two Regimes of the two leveraged investment strategies

If the S&P 500 is over its SMA200 Portfolio is in Regime 1, otherwise in Regime 2.

4 Data and Methodology

4.1 Data origin and structure

The data utilized in this thesis comprise of time series of financial products. The time series comprises daily price data or interest rate data for each financial product. The time series utilized in this thesis were obtained from either the databases provided by LMU Institut for Finance & Banking or publicly available sources.
Specifically the data was sourced as follows:
LMU Institut for Finance & Banking:

- Gold
- MSCI World
- MSCI Emerging Markets
- FTSE 100
- MSCI Europe
- Nasdaq 100
- FTSE 100
- S&P 500
- DAX
- Hang Seng
- Nikkei 225
- S&P 500 Value
- S&P 500 Growth
- S&P Small Cap Pure Value

Investing.com:

- 10 Year US Treasury Yield

Yahoo Fianance:

- Individual Stocks

Due to differing availabilities of length between the different time series the the data has been split into 3 subsets.

January 01, 1970 - December 31, 2022

This subset covers the longest time span and encompasses all time series which have available data back to January 01, 1970.
Those are:

- MSCI World
- S&P 500
- Dax
- Hang Seng
- Nikkei 225
- Gold

January 01, 1988 - December 31, 2022

This particular subset encompasses time series that possess data available since January 01, 1988, also encompassing those with data available for a longer duration.
Those are:

- MSCI World
- S&P 500
- Dax
- Hang Seng
- Nikkei 225
- Gold
- MSCI Emerging Markets
- FTSE 100
- MSCI Europe
- Nasdaq 100
- S&P 500 Value
- S&P 500 Growth
- Low Beta Stocks

January 01, 1996 - December 31, 2022

This particular subset pertains to the briefest duration and encompasses all time series that possess accessible data dating back to January 01, 1996, also encompassing those with data available for a longer period.
Those are:

- MSCI World
- S&P 500
- Dax
- Hang Seng
- Nikkei 225
- Gold
- MSCI Emerging Markets
- FTSE 100
- MSCI Europe
- Nasdaq 100
- S&P 500 Value
- S&P 500 Growth
- Low Beta Stocks
- S&P Small Cap Pure Value

As outlined in Portfolio Theory and Research three portfolios have been manually build using some of the given time series.

4.2 Data manipulation

4.2.1 Leverage for the long run

Two portfolios are built based on Bilello and Gayed (2016). One holding Gold (S&P 500 3x & Gold) under the SMA200 of the S&P 500, the other one holding cash (S&P 500 3x & Cash). To implement this strategy, using the given data, the daily return of the S&P 500 and Gold was calculated, using the following formula:

$$r_t = \frac{p_{t+1} - p_t}{p_t}$$

Where:

$$r_t : \text{daily return at time t}$$
$$p_t : \text{price at time t}$$
$$p_{t+1} : \text{price at time t+1}$$

To achive the leverage that was used in the by Bilello and Gayed, the daily returns of these portfolios where calculated as follows:

$$r_{t,Lev.} = \begin{cases} 3 \cdot r_{t,\text{S\&P 500}}, & \text{if S\&P 500} > \text{SMA200}, \\ r_{t,\text{Gold}}, & \text{if S\&P 500} \leq \text{SMA200 and portfolio "S\&P 500 3x \& Gold"}, \\ 0, & \text{if S\&P 500} \leq \text{SMA200 and portfolio "S\&P 500 3x \& Cash"}. \end{cases}$$

The price of this Portfolio was calculated using a starting value of 100 and compounding the daily returns $r_{t,\text{Lev.}}$ over time.

4.2.2 Evolutionary Portfolio Theory

The Evolutionary Portfolio uses data from the MSCI World, MSCI Emerging Markets and 10 Year US Treasury Yield, structured as explained in Evolutionary Portfolio Theory (3.3.6). The return that such a strategy would have archived by its exposure to the bonds was approximated as follows:

$$r_{t,\text{daily yield}} = \frac{r_{t,\text{yearly yield}}}{365} \tag{2}$$

where $r_{t,\text{daily yield}}$ represents the daily yield at time t, and $r_{t,\text{yearly yield}}$ denotes the yearly return offered by a 10-year US treasury bought at time t.
Again, this is only an approximation and wouldn't be the the exact return if that strategy was applied in the real world. The daily returns for the MSCI World and MSCI Emerging Markets where calculated the same way as for the S&P 500.
The daily return of this portfolio is

$$r_{t,\text{Evolutionary}} = \begin{cases} r_{t,\text{EM}} \cdot 0.1 + r_{t,\text{World}} \cdot 0.7 + r_{t,\text{daily yield}} \cdot 0.2, & \textit{Regime } 1 \\ r_{t,\text{EM}} \cdot 0.13 + r_{t,\text{World}} \cdot 0.87 + r_{t,\text{daily yield}} \cdot 0, & \textit{Regime } 2 \\ r_{t,\text{EM}} \cdot 0.15 + r_{t,\text{World}} \cdot 1.05 + r_{t,\text{daily yield}} \cdot (-0.2), & \textit{Regime } 3 \end{cases}$$

Where
$r_{t,\text{EM}}$ is the daily return of the MSCI Emerging Markets at time t
$r_{t,\text{World}}$ is the daily return of the MSCI World at time t
$r_{t,\text{daily yield}}$ is the daily return of the 10 Year US Treasury at time t
Regimes 1-3 as defined in Evolutionary Portfolio Theory (3.3.6)
The price of this Portfolio was calculated using a starting value of 100 and compounding the daily returns $r_{t,\text{Evolutionary.}}$ over time.

4.2.3 Low Beta

The individual stock data was sourced differently than the data for the indices. Thus there are some differences in trading days present in the data sets which lead to differences in length of the time series. Those have been fixed by replacing missing values with stock value from the previous existing trading day.
This manipulation does not effect overall performance and also doesn't effect stock volatility to a relevant extend, with the change in portfolio standard deviation $< 0.025\%$.

4.3 Additional Information

- Taxes, fees and other factors of this nature that could effect the returns where not taken into account.
- The closing price was always used as the daily price

5 Visualization of Portfolio Performance

In order to obtain a comprehensive understanding of the performance of the different portfolios, visual representations such as performance graphs and charts are provided. These graphical tools facilitate the comparison of average performance and volatility across the portfolios. To improve clarity and organization, the portfolios have been divided into three subsections for the time periods from 1988 to 2022 and 1996 to 2022. This categorization has led to the creation of more coherent and visually appealing graphs. In order to account for the limited availability of portfolios within the time frame spanning from 1970 to 2022, all time series data has been consolidated and visualized in a single plot. The utilization of logistic scales on the y-axis has been employed to account for the exponential growth observed in stock portfolios.

Group one:
Indices
Group two:
Faktor Portfolios
Group three:
Other Portfolios

Each group is displayed with the S&P500 as a benchmark.
More detailed performance comparisons can be found in the Appendix.

5.1 1970 - 2022

According to the data depicted in Figure 1, spanning the period from 1970 to 2022, all portfolios exhibited a return exceeding 1000%. The Portfolio comprising the Nikkei 225 exhibited the poorest performance, whereas both leveraged strategies outperformed the non-leveraged portfolios. The most robust performance was achieved through the amalgamation of the 3x leveraged S&P 500 and Gold. The average annual returns also demonstrate this superior performance. The Nikkei 225 index concluded with a modest average return of 7.3%, whereas the leveraged S&P 500 index, when combined with gold, yielded a considerably higher average annual return of 27.9%. Nevertheless, the enhanced performance was no free lunch.
As Figure 3 shows, the best performing portfolio ended up with one of the highest volatility, which is to be expected from a leveraged strategy.

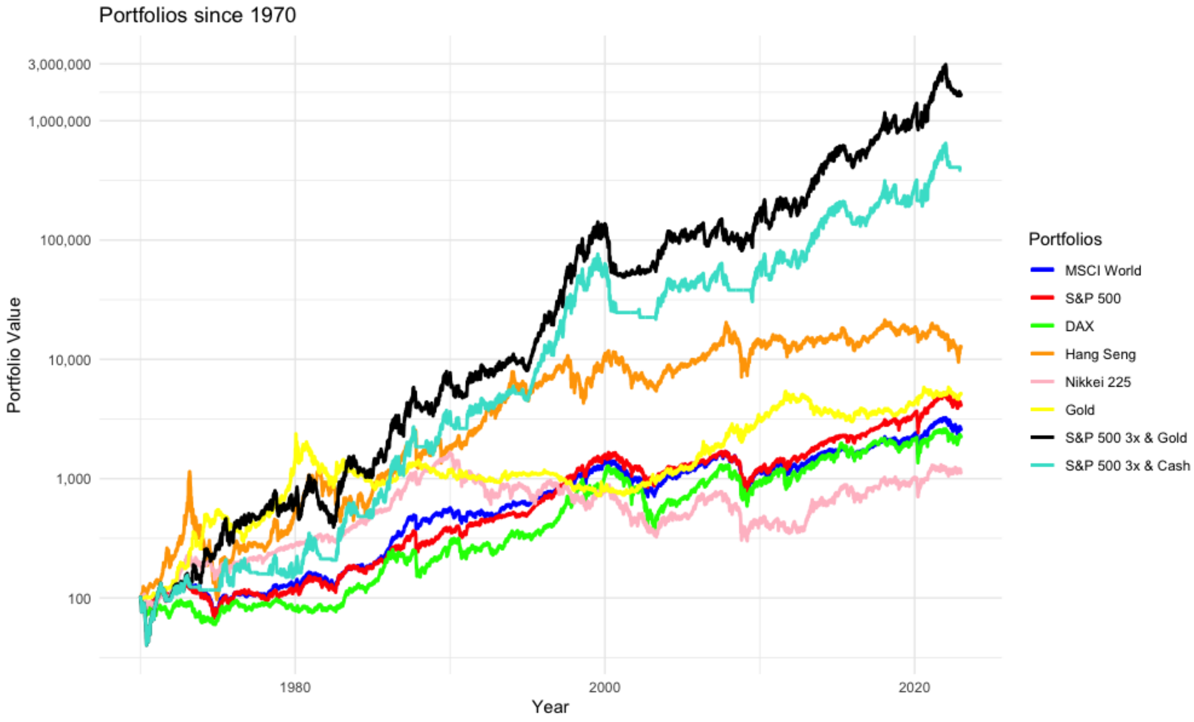

Figure 1: Portfolios 1970-2022

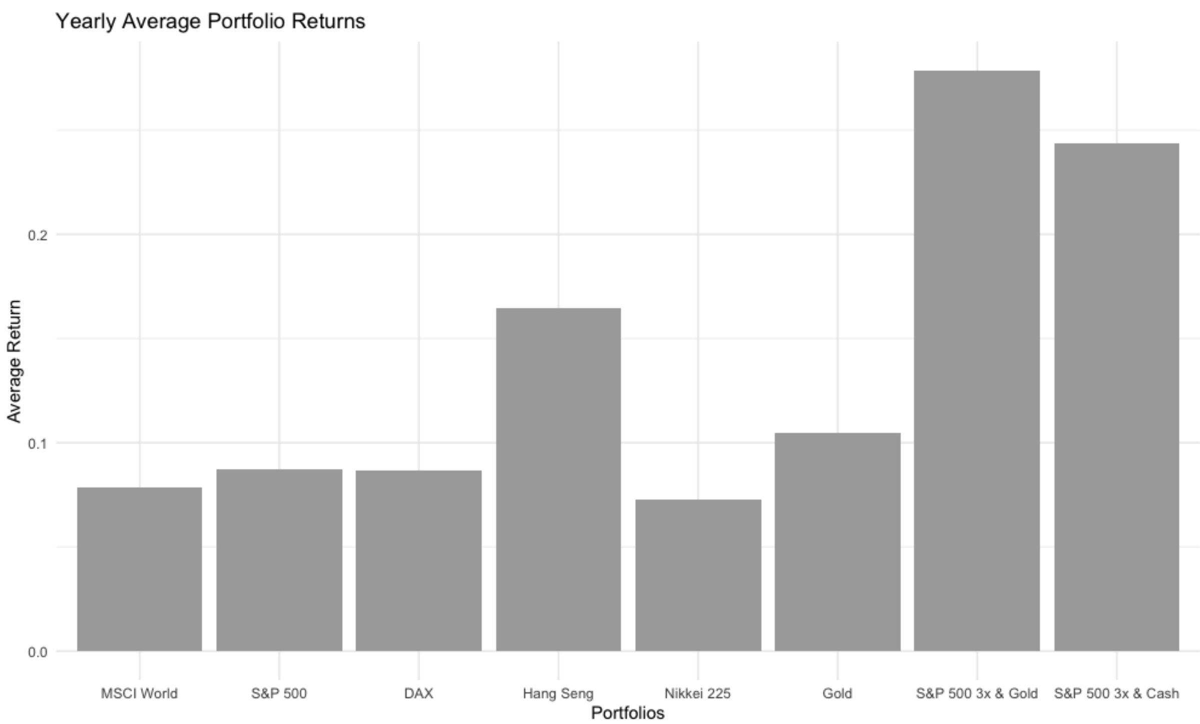

Figure 2: Average Yearly Returns 1970-2022

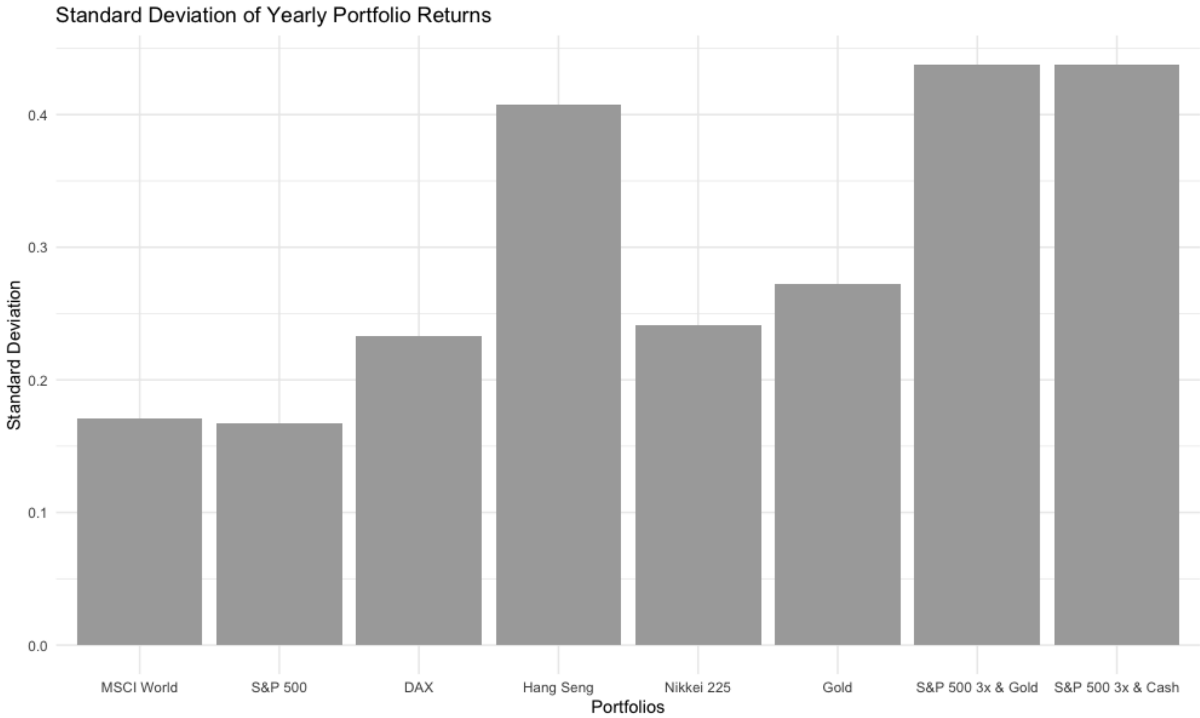

Figure 3: Standard Deviation of Yearly Returns 1970-2022

5.2 1988 - 2022

In a manner akin to the temporal interval spanning from 1970 to 2022, the Nikkei 225 index displayed a notably subpar performance from 1988 to 2022, ultimately achieving a marginal positive outcome over the duration of 32 years (Figure 4). The best performing index in this time frame was the NASDAQ-100, returning an average of 14% per year (Figure 7). It is noteworthy that the value portfolio, which was anticipated to generate superior returns through its factor prime, not only ended up performing worse compared to the growth factor portfolio but also underperformed the S&P 500 benchmark (Figure 5).

The best performance was again archived by the leveraged S&P 500 and gold combination (Figure 6) while the lowest volatility was unsurprisingly measured at the Low Beta portfolio (Figure 8).

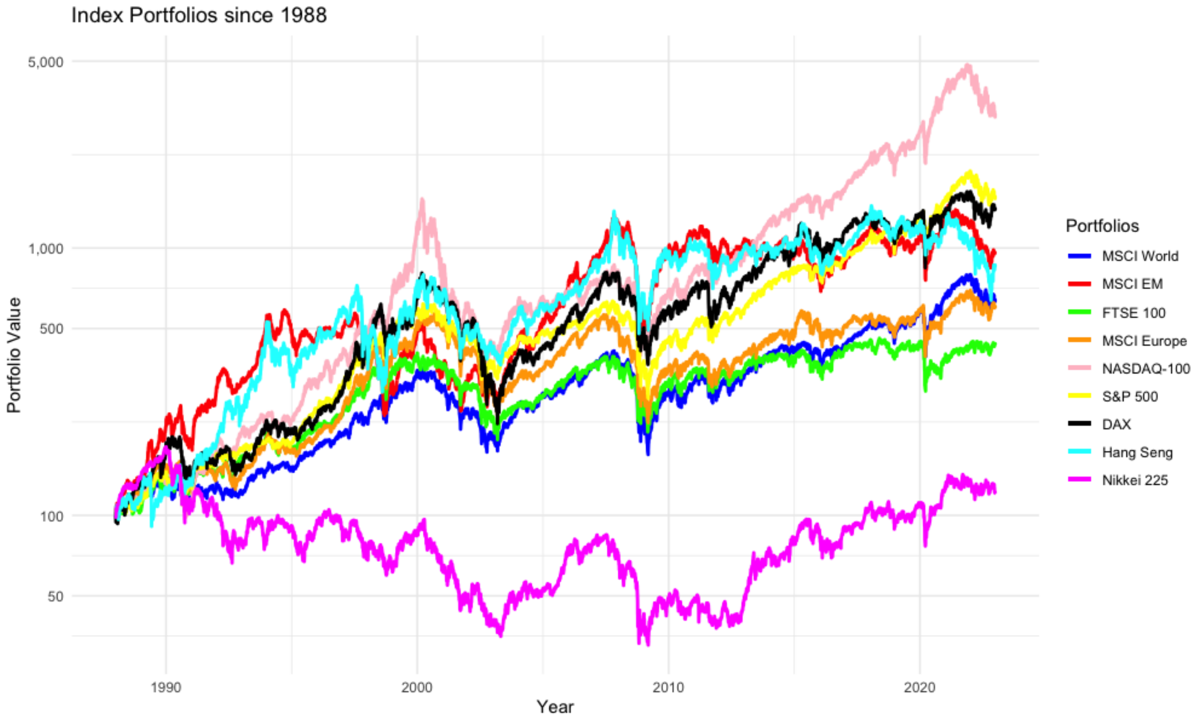

Figure 4: Portfolios 1988-2022

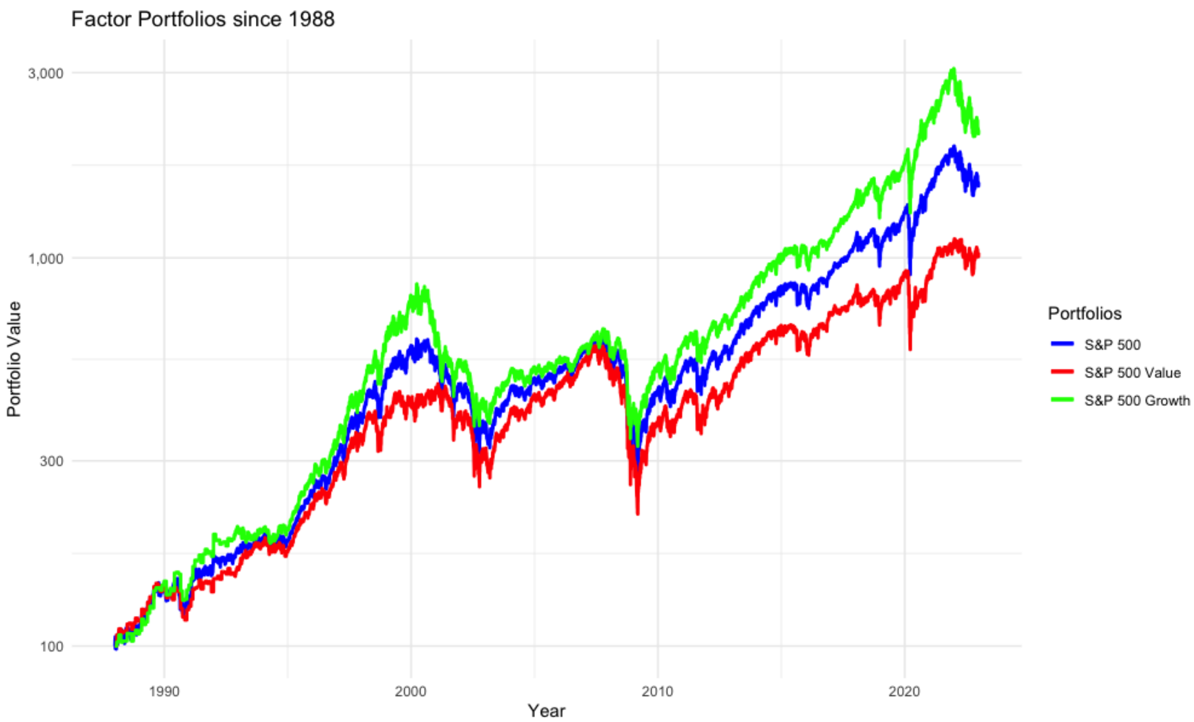

Figure 5: Portfolios 1988-2022

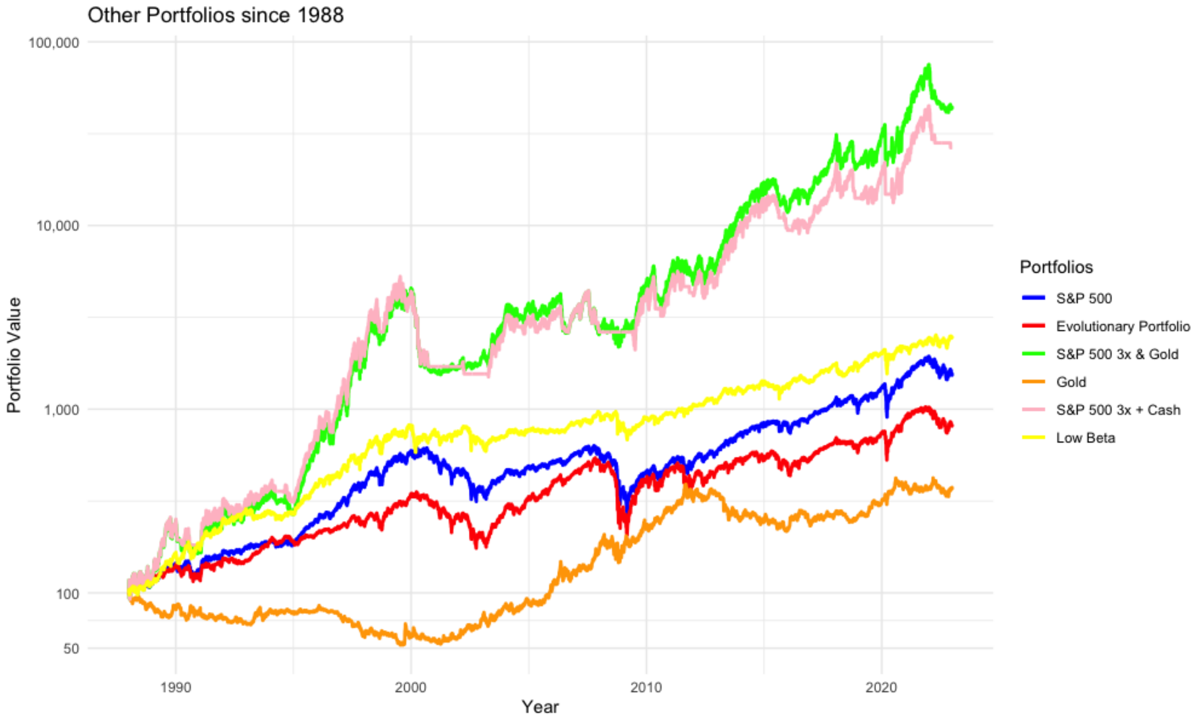

Figure 6: Portfolios 1988-2022

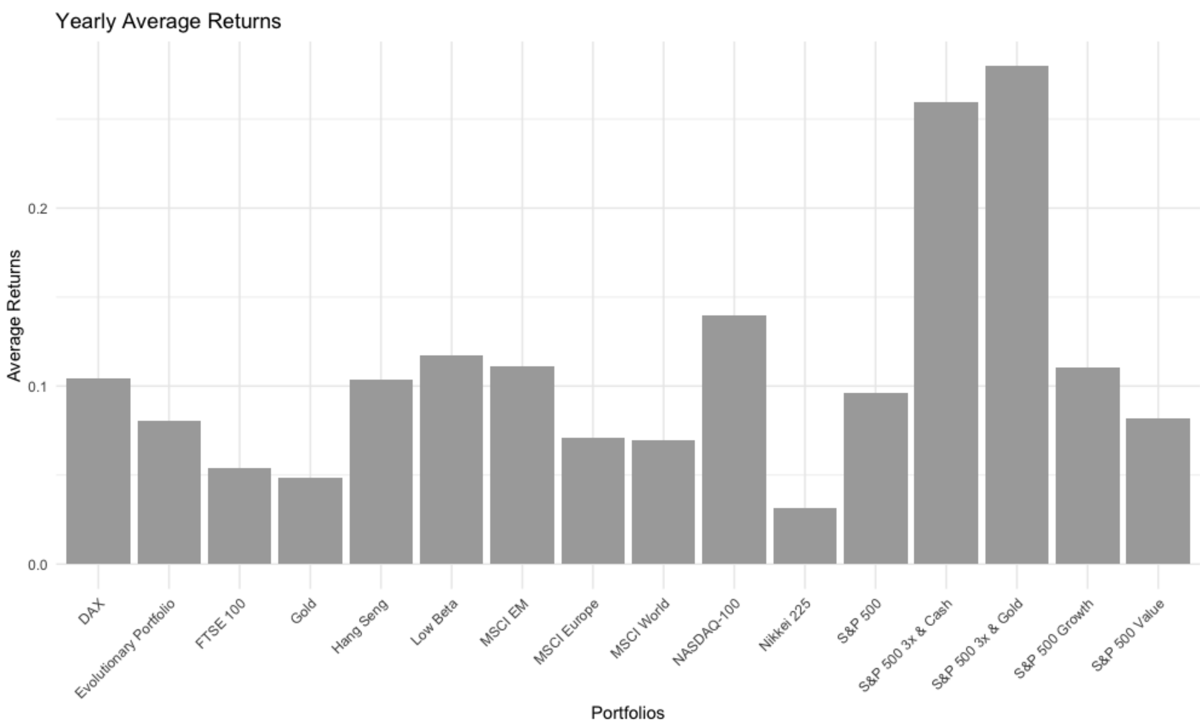

Figure 7: Average Yearly Returns 1988-2022

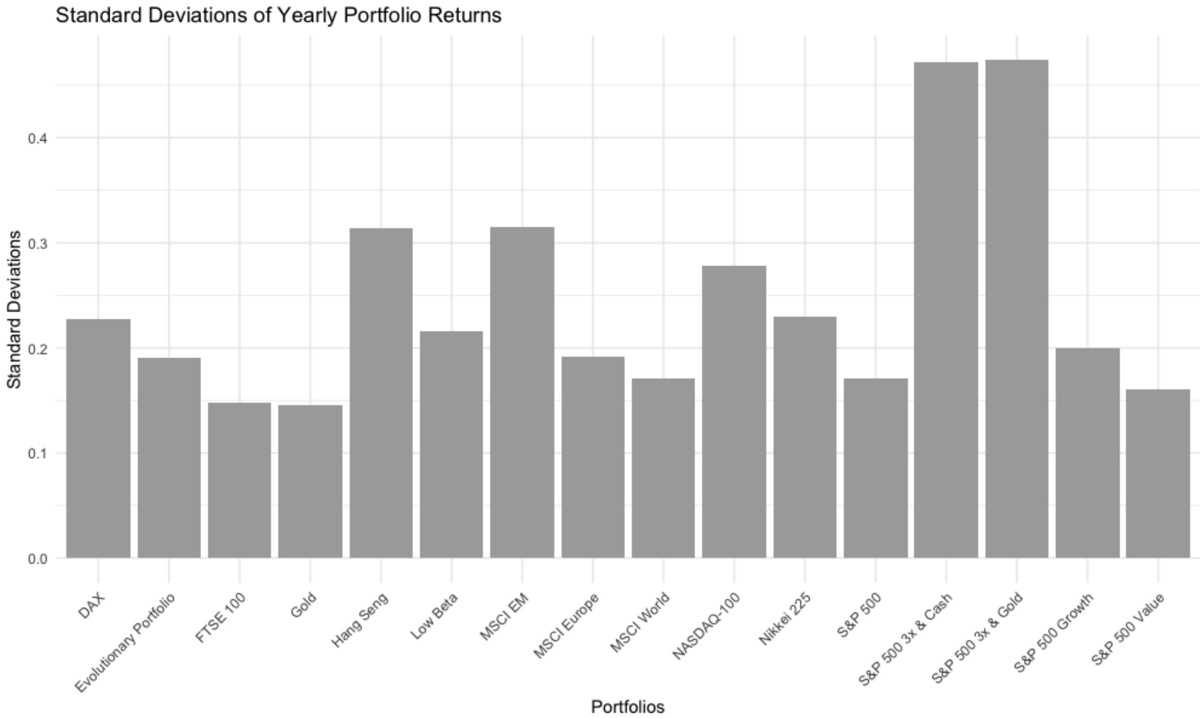

Figure 8: Standard Deviation of Yearly Returns 1988-2022

5.3 1996 - 2022

From 1996 - 2022 the head and the tail of the index performances didn't change compared to 1988 - 2022. As Figure 9 shows the NASDAQ-100 continues leading the ranking while the Nikkei 225 is again lagging behind. The greatest performance was again archived by the combination of the leveraged S&P 500 and gold, though with providing 25.5% less yearly return compared to the longer perspective since 1970. The S&P Small Cap Pure Value portfolio has demonstrated superior performance within the given time frame, making it the most effective factor portfolio, (Figure 10) with an average return of 10.9% (Figure 12), while the lowest volatility was again archived by the Low Beta portfolio (Figure 13).

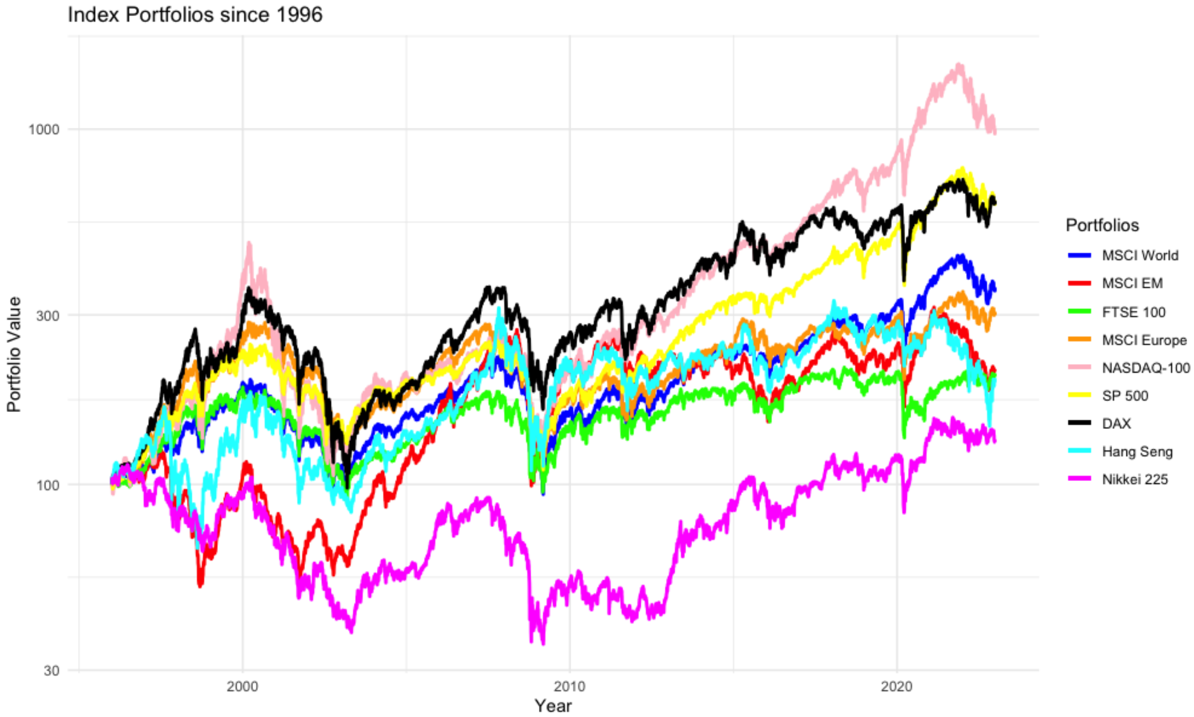

Figure 9: Portfolios 1988-2022

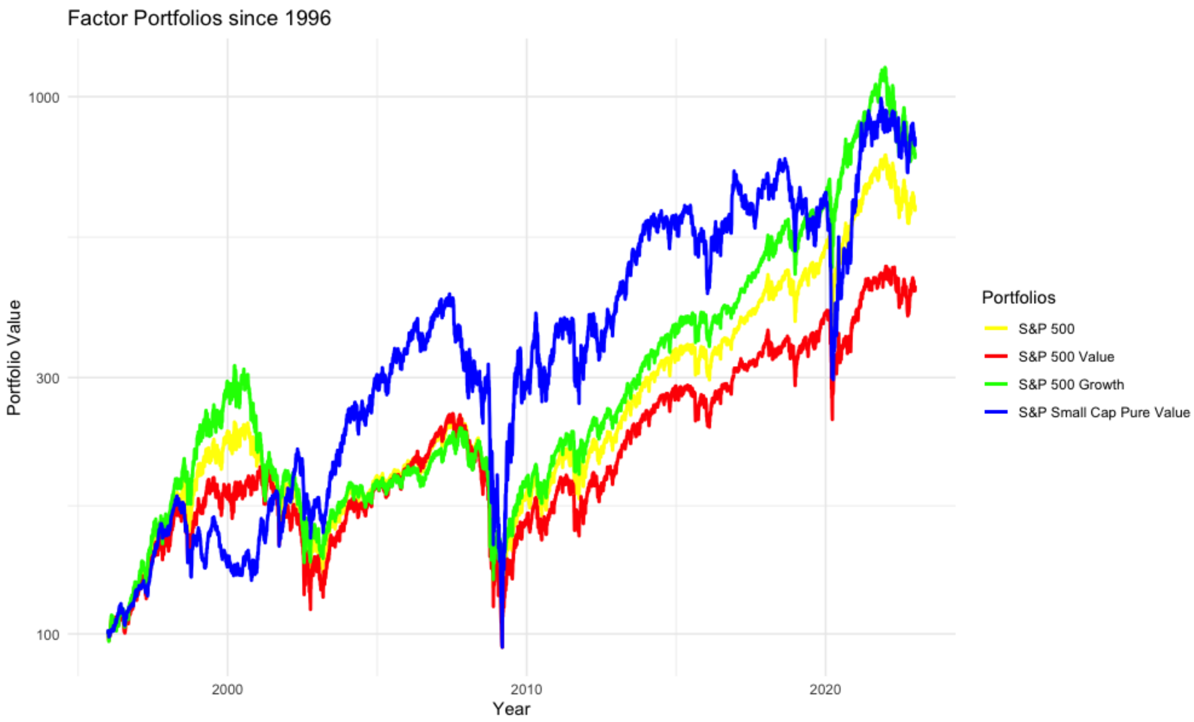

Figure 10: Portfolios 1996 -2022

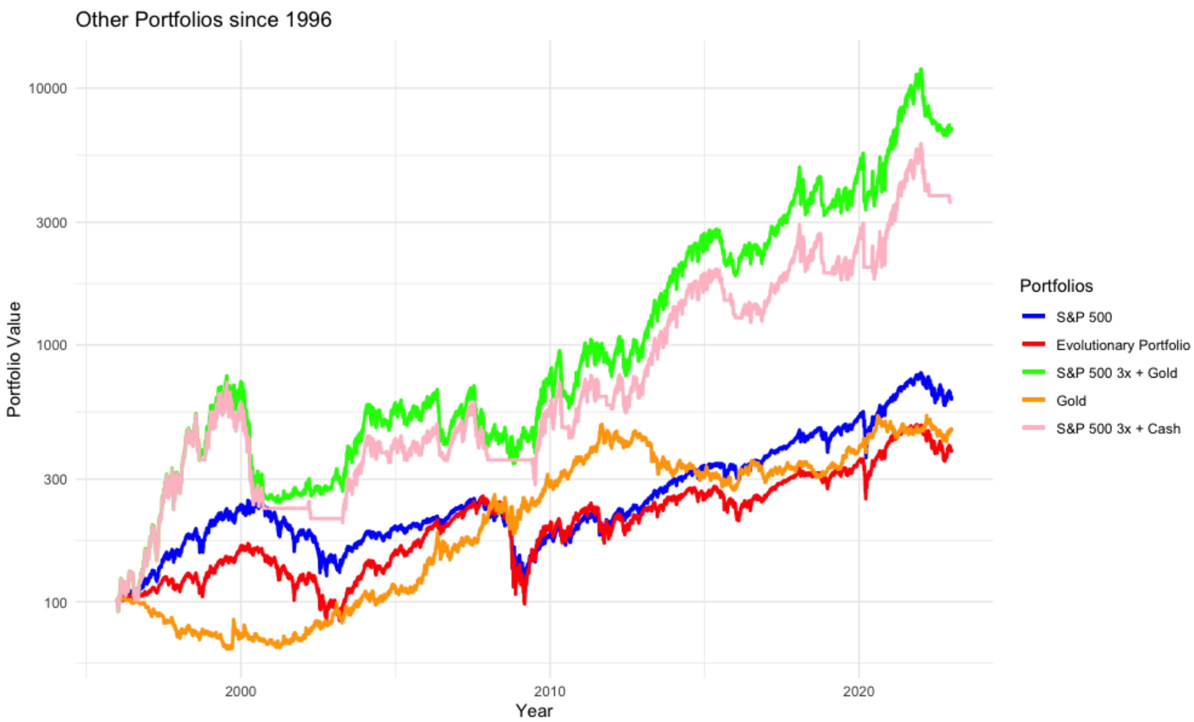

Figure 11: Portfolios 1996 -2022

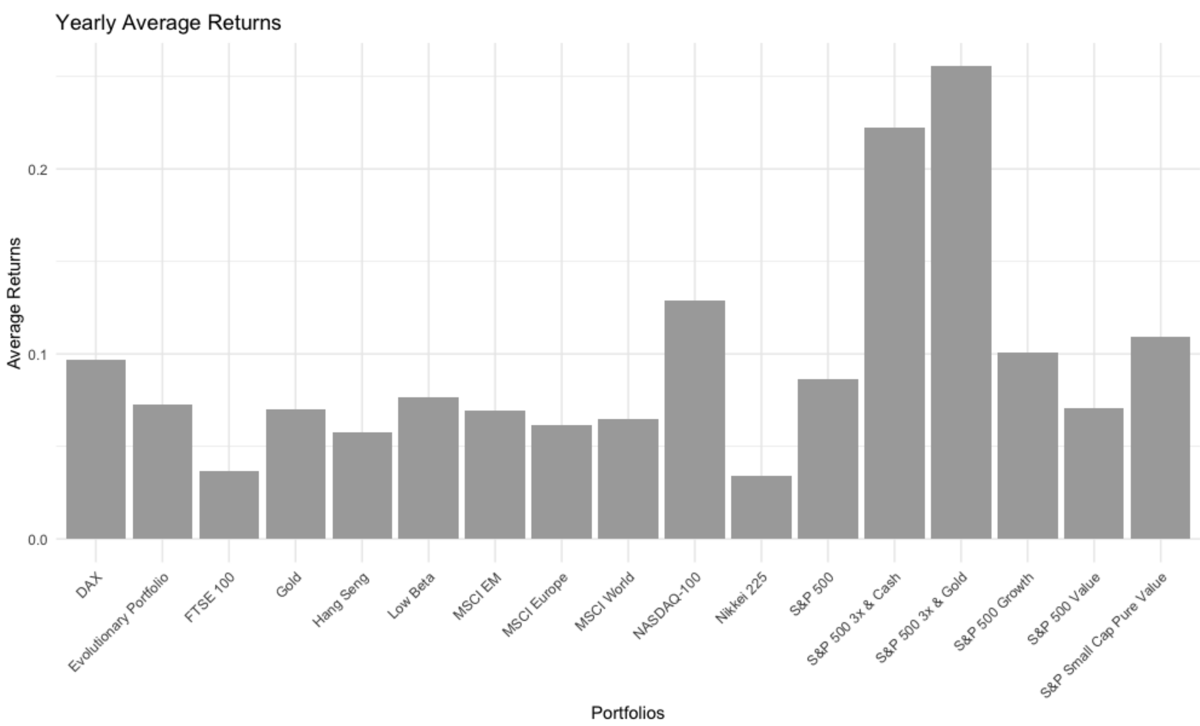

Figure 12: Portfolios 1996 -2022

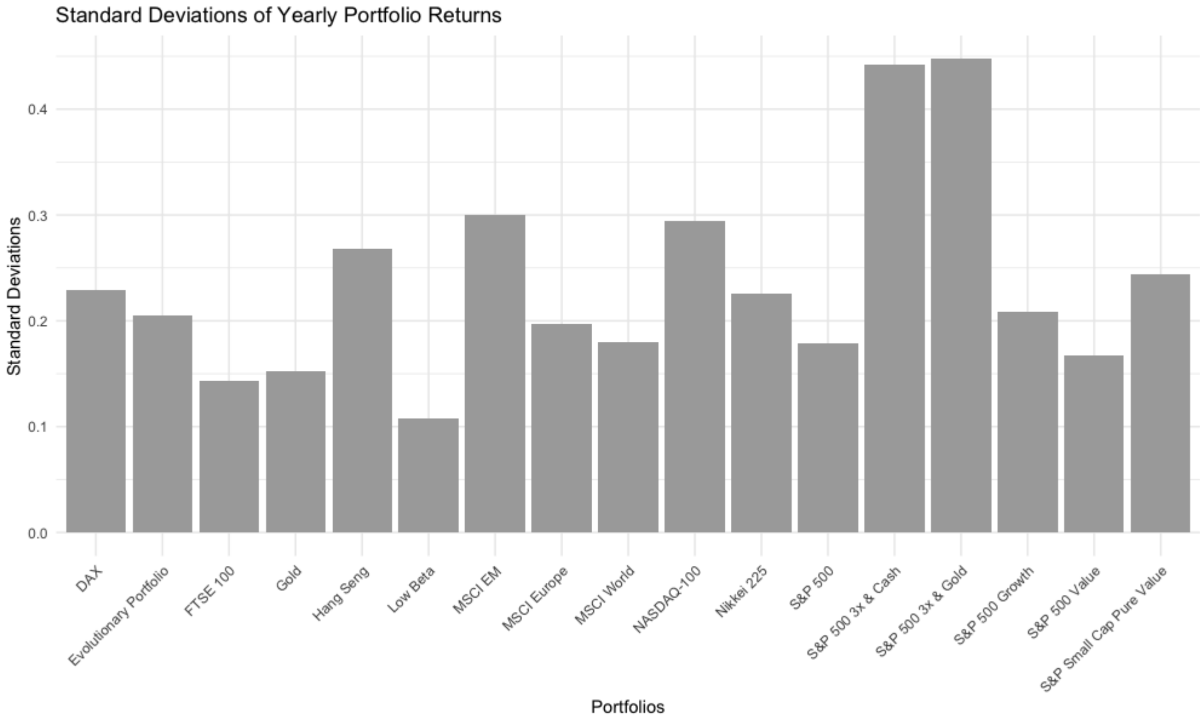

Figure 13: Portfolios 1996 -2022

6 Application of Ranking Analysis

6.1 Covariance Matrix

A fundamental for the application of Mogstad et al. (2020) method to quantify ranking uncertainty is the covariance matrix of the different ranked populations, in our case the different portfolios in each time span.

The covariance matrix quantifies the covariance between pairs of variables. It's a square matrix of dimension $p \times p$ for a dataset with p variables, where each entry at position i, j represents the covariance between the i^{th} and j^{th} variables. The formula for covariance is given by:

$$\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] \tag{3}$$

where E denotes the expectation, X_i and X_j are the i^{th} and j^{th} variables, and μ_i and μ_j are their respective means.

The covariance matrix Σ is then constructed as follows:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{4}$$

The diagonal entries of the covariance matrix (σ_{ii} for all i) represent the variances of each variable. Therefore, one can view the covariance matrix as consisting of a diagonal matrix of variances:

$$D = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix} \tag{5}$$

and off-diagonal matrices that capture the covariances.(Ludwig Fahrmeir, 1999)

6.2 Auto correlation

However, the necessary prerequisite for this calculation of the covariance matrix is that the data sets are independent and identically distributed (i.i.d.). In time series of stock portfolio data it should be questioned whether the distribution of the returns is i.i.d. since autocorrelation is common in such data. Autocorrelation is a relationship between values separated from each other by a given time lag. Thus it is advisable to conduct an autocorrelation test such as the Durbin-Watson Test or the Ljung-Box test. (Shumway et al., 2000)

6.2.1 Durbin-Watson (DW) Test

The Durbin-Watson test is a statistical test used to detect the presence of autocorrelation in the residuals (prediction errors) of a regression analysis. It is named after statisticians James Durbin and Geoffrey Watson (Durbin and Watson, 1950).
The Durbin-Watson statistic is defined as:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \tag{6}$$

Where:

- e_t represents the residual at time t,
- T is the number of time periods.

The Durbin-Watson statistic ranges between 0 and 4. A value of 2 indicates no autocorrelation in the sample. Values from 0 to less than 2 suggest positive autocorrelation, and those from 2 to 4 suggest negative autocorrelation.
The null hypothesis of the Durbin-Watson test is that there is no autocorrelation in the sample. More formally, we denote the null hypothesis as:

$$H_0 : \rho = 0 \tag{7}$$

Where:

- ρ is the population autocorrelation.

Rejection of this hypothesis suggests that the errors are autocorrelated.

6.2.2 Ljung-Box (LB) Test

The Ljung-Box test is a type of statistical test of whether any of a group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags (Ljung and Box, 1978).
The Ljung-Box statistic is defined as:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k} \tag{8}$$

Where:

- n is the sample size,
- $\hat{\rho}_k$ is the sample autocorrelation at lag k,
- h is the number of lags being tested.

In the context of this test, a p-value less than a chosen significance level (i.e. 5%) suggests that the data are probably not independently distributed. As Hurra (2022) showed, the Ljung-Box test has more power for smaller sample sizes, such as some of those used in this thesis, due to the restricted available time frames. To gain more certainty both tests have been performed.

6.2.3 Autocorrelation Test Results

After running the Durbin-Watson test using the Car package and the Ljung-Box Test using the Stats package in R, it seems evident that there is no relevant autocorrelation within the time series. Table 3 shows the test results for the portfolios from 1970 - 2022. The results for the other time spans are available in the Appendix(Table 7, 8). All P-Values are greater than 0.05, most of them far greater and the DW-Statistic from the Durbin-Watson test is always relatively close to 2.
As Hurra (2022) predict, the P-Values of the Ljung- Box tend to be smaller than those of the Durbin-Watson test. Accordingly it is assumed that it is viable to use the previously described covariance matrix (6.1) for this data.
It should be mentioned, that this is somewhat lucky, since autocorrelation would have made the problem significantly more difficult.

Portfolios since 1970	DW Statistic	P Value DW	P Value LB
Monthly Return MSCI World	1.986	0.840	0.220
Monthly Return S&P 500	1.988	0.878	0.326
Monthly Return DAX	1.999	0.968	0.881
Monthly Return HANG SENG	2.000	0.990	0.833
Monthly Return Nikkei 225	1.995	0.924	0.725
Monthly Return Gold	2.000	0.984	0.877
Monthly Return S&P 500 Lev Gold	1.986	0.800	0.679
Monthly Return S&P 500 Lev Cash	1.943	0.872	0.271
Yearly Return MSCI World	1.960	0.866	0.324
Yearly Return S&P 500	1.965	0.938	0.184
Yearly Return DAX	2.035	0.944	0.206
Yearly Return HANG SENG	2.018	0.976	0.217
Yearly Return Nikkei 225	1.880	0.680	0.739
Yearly Return Gold	1.877	0.592	0.250
Yearly Return S&P 500 Lev Gold	1.936	0.860	0.173
Yearly Return S&P 500 Lev Cash	1.943	0.848	0.271

Table 3: Results of Durbin-Watson Test & Ljung-Box Test for Portfolios from 1970 - 2022

6.3 The difficulties with autocorrelation

Applying Mogstad et al. (2020) strategy for on populations with autocorrelated data is difficult when they should be ranked by their expected values. Common econometrical methods such as the Heteroskedasticity and Autocorrelation Consistent Method (HAC) or multivariate Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models allow us to calculate the volatility of time series. It is possible to calculate a covariance matrix of the different time series based on that co-volatility. This covariance matrix however, displays the covariance structure of the time series themselves, instead of their means. While one could simply divide the basic sample covariance matrix (6.1) by the number of observations to calculate the covariance matrix of the mean, in a non i.i.d. scenario like this autocorrelated one, the computation becomes considerably more complex. Indeed, there is no straightforward approach to derive the covariance matrix of the mean solely based on the sample covariance matrix of the mean when dealing with autocorrelated data. One common option to calculate a covariance matrix in such a scenario would be the use of resampling procedures. In this case one could calculate parameters of the given data structure using ARIMA and GARCH models or, too keep it simpler, use multivariate moving block bootstrap (MMBB) and thus re-sample new, structurally equivalent data sets. By calculating the mean for each time series in each sampled data set, one gets a distribution of mean values and thus could calculate the covariance matrix of those mean values.
Yet there arises a new problem using this strategy. The calculation of uncertainty is now based on a distribution of mean values, which is influenced more by the number of re-sampling iterations rather than the variance in the original time series.
As a result, varying the number of sampling iterations leads to differing confidence sets

for the ranks, as shown in Figure 14. Although, as described in Dealing with Uncertainty (2.2), the statistical method employed to quantify uncertainty is indeed valid, it is important to acknowledge the limitations that have been previously discussed. Furthermore, it is worth noting that the dependence of the confidence sets on the number of iterations, may not align with the typical objectives one seeks to achieve using Mogstad et al. method. Hence, it appears that the application of these confidence sets on autocorrelated data lacks a straightforward approach.

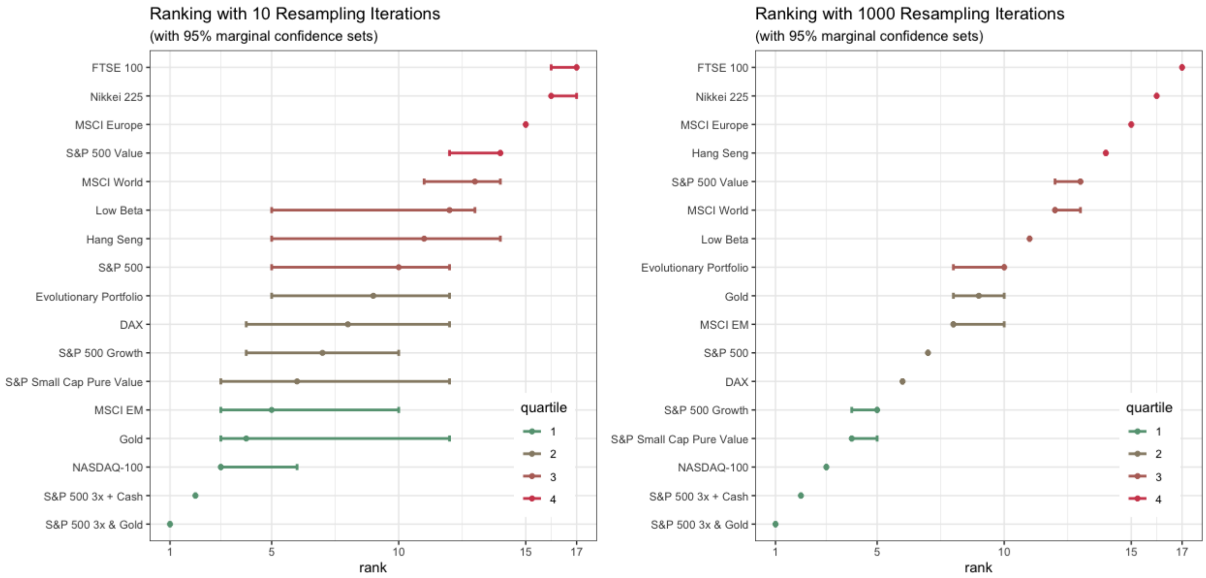

Figure 14: Marginal confidence sets of means with 10 vs. 1000 iterations, using block length of 3 (following Romano (1992))

6.4 Portfolio Rankings

By utilizing the provided covariance matrix (6.1), it is possible to compute the uncertainties associated with the portfolio rankings using the marginal, as well as the simultaneous confidence sets.

$\hat{Rank}_i$ ranks $Portfolio_i$ based on the average annual return $\overline{r}_{y,i}$ or average monthly return $\overline{r}_{m,i}$, which are estimators for the true expected returns $r_{y,i}$ or $r_{m,i}$. The average returns are used to estimate the rank, hence $Rank_i$ is also just an estimator for the true $Rank_i$. I want to emphasize the different definition of the notation r_i in this thesis and the paper from Mogstad et al. (2020), where r_i denotes the rank of population i.

The explicit reasoning of the construction of the marginal and simulatenous confidence sets can be looked up in Mogstad et al. paper. In this thesis the method is applied, using the available R package, that belongs to the paper.

While here the rankings for yearly portfolio returns are presented, the results for monthly returns are available in the Appendix.

The shown rankings use 95% confidence sets, rankings with 50% and 5% confidence sets are also available in the Appendix. Figure 15 shows the ranking of average yearly portfolio returns of the years 1970 - 2022. Even when basing the ranking estimation on this longest time frame of 52 years there seems to be a relatively high amount of uncertainty about the true rank of each portfolio. Unsurprisingly, the amount of uncertainty grows with shorter time spans and more ranked portfolios. While in the ranking from 1988 - 2022 (Figure 16) the confidence sets for most portfolios cover a large span of possible places and in the time span between 1996 and 2022 (Figure 17), most confidence sets cover all possible ranks.

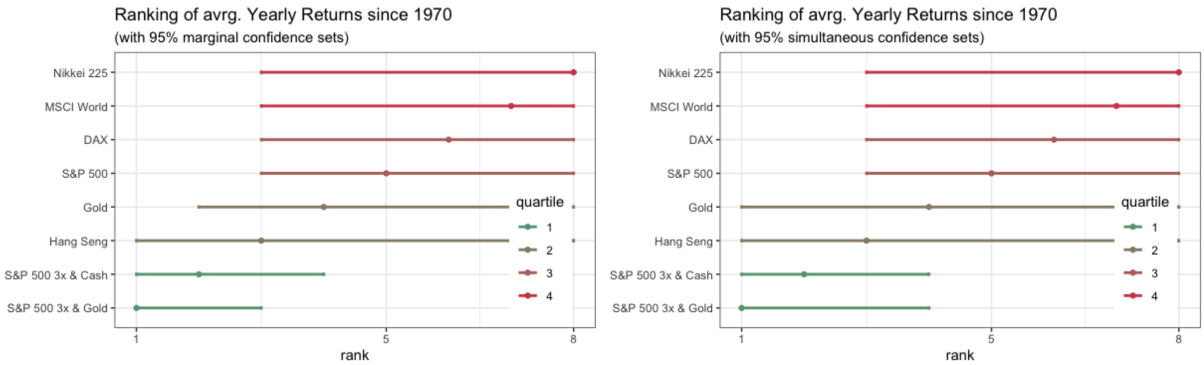

Figure 15: Rankings of avrg. Yearly Returns 1970 - 2022, with marginal and simultaneous confidence sets

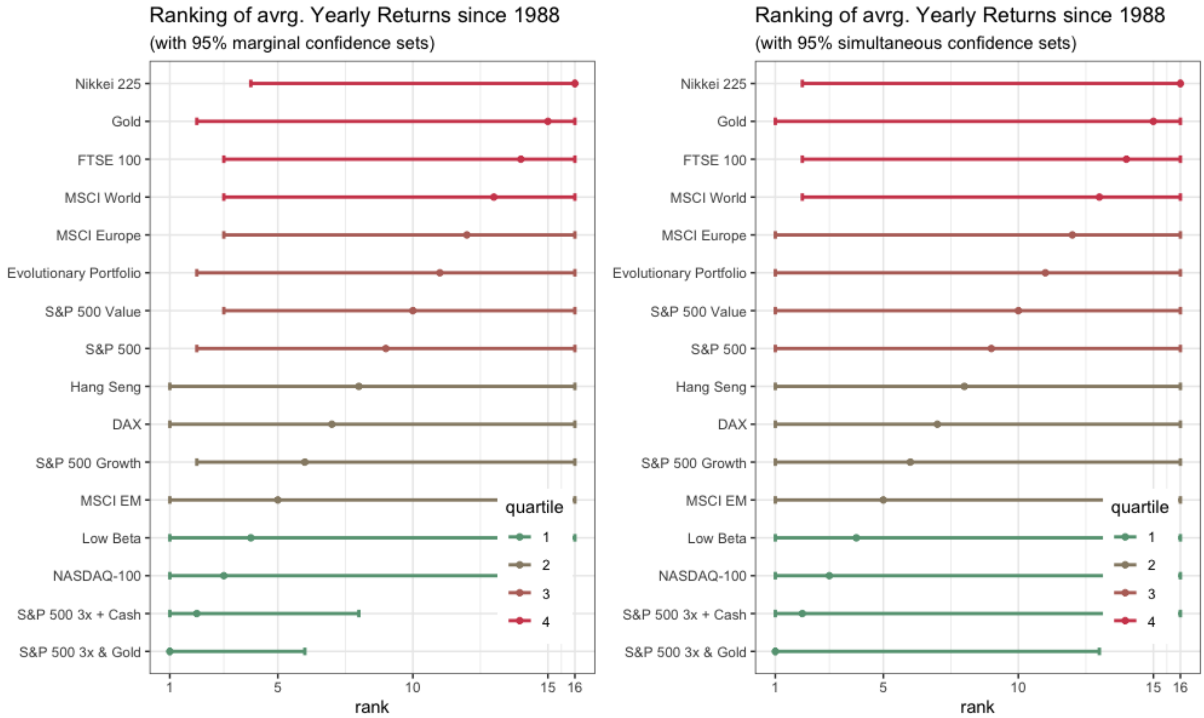

Figure 16: Rankings of avrg. Yearly Returns 1988 - 2022, with marginal and simultaneous confidence sets

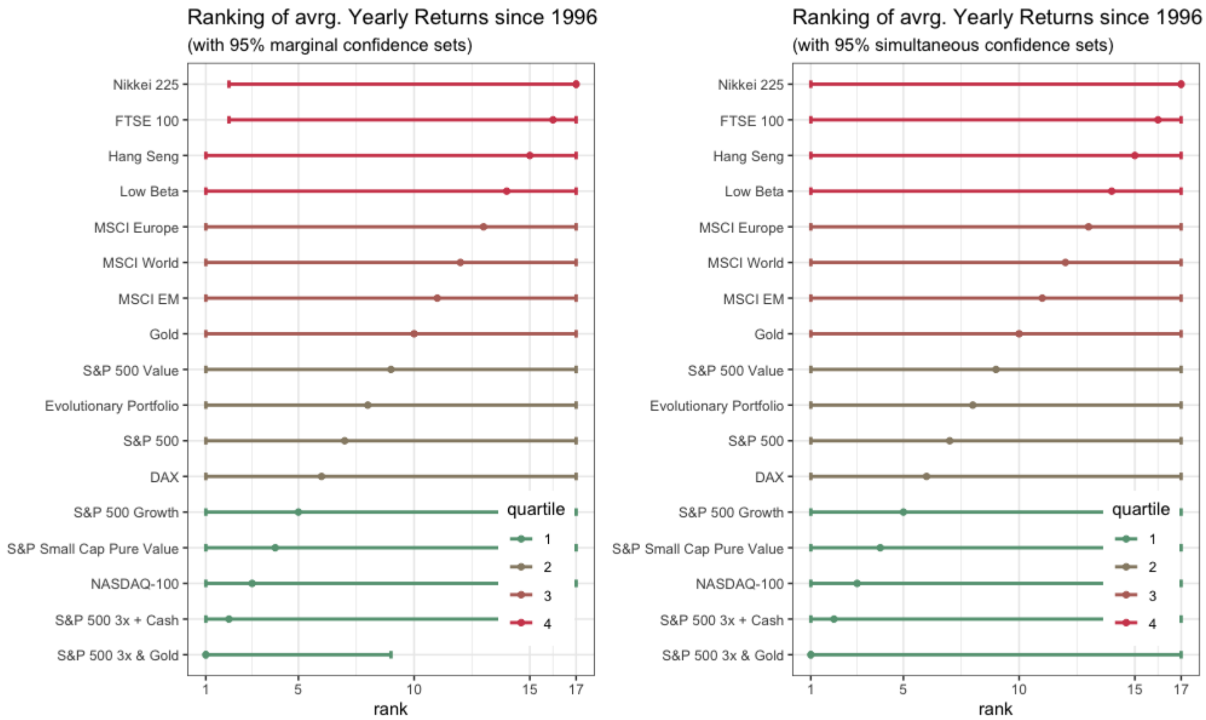

Figure 17: Rankings of avrg. Yearly Returns 1996 - 2022, with marginal and simultaneous confidence sets

Mogstad et al. (2020) stated that one reason that might make their method particularly interesting to researches is the ability to rank the populations by their concreteness. They define a population to be among the top if its rank is less than or equal to a pre-specified value τ. They introduce a confidence set for what they call τ- best populations. They define it to be a random set that contains the identities of these populations with probability approximately no less than some pre-specified level.
Table 4, 5 and 6 display the rankings of the portfolios, as well as their simultaneous confidence sets and the τ-best and τ-worst populations for different confidence levels for all time frames. It is evident once more that the majority of confidence sets exhibit a considerable degree of magnitude, particularly within the time frame spanning from 1996 to 2022. Even if the confidence level is reduced to 5% the τ- best place for the fourth rank S&P 500 Small Cap Pure Value, is place 15 of 17. It seems that in all three time frames the certainty of the ranks sharply reduces after rank 2 and especially after rank three, there is few difference in the τ- best for all other ranks.
Generally, if one wants to base portfolio management decisions on such rankings, it seems low confidence has to be accepted, to conclude any superiority of portfolios. It might also imply that combining different metrics rather than just the performance is useful for portfolio decision making. Combining different metrics is discussed in the ***Conclusion and Discussion*** (7).

$\hat{Rank}$	Portfolio	$\overline{r}_y$	Confidence Sets (CS)			τ-best		
			95% CS	50% CS	5% CS	95%	50%	5%
1	S&P 500 3x + Gold	27.9%	[1, 4]	[1, 2]	[1, 1]	4	2	1
2	S&P 500 3x + Cash	24.3%	[1, 4]	[2, 3]	[2, 2]	4	3	2
3	Hang Seng	16.4%	[1, 8]	[1, 6]	[3, 3]	8	6	3
4	Gold	10.5%	[1, 8]	[3, 8]	[4, 8]	8	8	8
5	S&P 500	8.7%	[3, 8]	[3, 8]	[4, 8]	8	8	8
6	DAX	8.7%	[3, 8]	[3, 8]	[4, 8]	8	8	8
7	MSCI World	7.8%	[3, 8]	[4, 8]	[4, 8]	8	8	8
8	Nikkei 225	7.3%	[3, 8]	[4, 8]	[4, 8]	8	8	8

Table 4: Overview on average yearly portfolio returns between 1970 and 2022. Confidence sets shows the simultaneous ranking confidence sets at a give confidence level. τ-best shows the confidence set of the τ-best portfolios, at a given ranking and confidence level.

$\hat{Rank}$	Portfolio	$\overline{r}_y$	Confidence Sets (CS)			τ-best		
			95% CS	50% CS	5% CS	95%	50%	5%
1	S&P 500 3x Gold	28%	[1, 13]	[1, 4]	[1, 1]	13	4	1
2	S&P 500 3x Cash	25.9%	[1, 16]	[1, 5]	[2, 2]	16	5	2
3	NASDAQ-100	14%	[1, 16]	[2, 13]	[3, 8]	16	13	8
4	MSCI EM	11.1%	[1, 16]	[1, 16]	[3, 14]	16	16	14
5	S&P 500 Growth	11.1%	[1, 16]	[3, 13]	[3, 8]	16	16	13
6	DAX	10.4%	[1, 16]	[3, 13]	[3, 12]	16	16	13
7	Hang Seng	10.4%	[1, 16]	[1, 16]	[3, 15]	16	16	15
8	Low Beta	10.3%	[1, 16]	[3, 15]	[3, 12]	16	16	15
9	S&P 500	9.6%	[1, 16]	[3, 14]	[5, 11]	16	16	15
10	S&P 500 Value	8.3%	[1, 16]	[3, 16]	[6, 14]	16	16	15
11	Evolutionary Portfolio	8%	[1, 16]	[3, 16]	[5, 15]	16	16	15
12	MSCI Europe	7.1%	[1, 16]	[4, 16]	[7, 16]	16	16	16
13	MSCI World	7%	[2, 16]	[6, 16]	[8, 15]	16	16	16
14	FTSE 100	5.4%	[2, 16]	[8, 16]	[10, 16]	16	16	16
15	Gold	4.8%	[1, 16]	[3, 16]	[6, 16]	16	16	16
16	Nikkei 225	3.1%	[2, 16]	[6, 16]	[13, 16]	16	16	16

Table 5: Overview on average yearly portfolio returns between 1988 and 2022. Confidence sets shows the simultaneous ranking confidence sets at a give confidence level. τ-best shows the confidence set of the τ-best portfolios, at a given ranking and confidence level.

$\hat{Rank}$	Portfolio	$\overline{r}_y$	Confidence Sets (CS)			τ-best		
			95% CS	50% CS	5% CS	95%	50%	5%
1	S&P 500 3x Gold	25.5%	[1, 17]	[1, 5]	[1, 1]	17	5	1
2	S&P 500 3x Cash	22.3%	[1, 17]	[1, 11]	[2, 2]	17	11	2
3	NASDAQ-100	12.9%	[1, 17]	[1, 15]	[3, 8]	17	15	8
4	S&P Small Cap Pure Value	10.9%	[1, 17]	[2, 17]	[3, 15]	17	17	15
5	S&P 500 Growth	10%	[1, 17]	[2, 15]	[3, 13]	17	17	15
6	DAX	9.7%	[1, 17]	[2, 15]	[3, 13]	17	17	15
7	S&P 500	8.6%	[1, 17]	[3, 15]	[4, 13]	17	17	15
8	Low Beta	7.3%	[1, 17]	[2, 17]	[3, 16]	17	17	16
9	Evolutionary Portfolio	7.3%	[1, 17]	[2, 17]	[4, 16]	17	17	16
10	S&P 500 Value	7.1%	[1, 17]	[3, 16]	[4, 16]	17	17	16
11	Gold	7%	[1, 17]	[1, 17]	[3, 17]	17	17	17
12	MSCI EM	6.9%	[1, 17]	[1, 17]	[4, 17]	17	17	17
13	MSCI World	6.5%	[1, 17]	[5, 17]	[7, 16]	17	17	17
14	MSCI Europe	6.2%	[1, 17]	[4, 17]	[7, 16]	17	17	17
15	Hang Seng	5.8%	[1, 17]	[2, 17]	[4, 17]	17	17	17
16	FTSE 100	3.7%	[1, 17]	[8, 17]	[14, 17]	17	17	17
17	Nikkei 225	3.4%	[1, 17]	[4, 17]	[8, 17]	17	17	17

Table 6: Overview on average yearly portfolio returns between 1996 and 2022. Confidence sets shows the simultaneous ranking confidence sets at a give confidence level. τ-best shows the confidence set of the τ-best portfolios, at a given ranking and confidence level.

7 Conclusion and Discussion

7.1 Statistical take away of building confidence sets for time series rankings

The seminal paper by Mogstad et al. (2020) significantly advanced the estimation techniques for uncertainty in rankings. Its utility appears particularly profound when differentiating between simultaneous and marginal confidence sets. Although the data incorporated within this thesis did not exhibit substantial autocorrelation, I embarked upon an investigation into the applicability of Mogstad et al.'s method within the realm of autocorrelated data.

This investigation elucidated that autocorrelation introduces formidable challenges when ranking based on the mean values of datasets, a commonly desired criterion, especially within the context of market returns. The mean is often leveraged to approximate the expected return, mandating the construction of a covariance matrix of these expected returns. This is a straightforward process in the context of i.i.d. data. However, when dealing with autocorrelated data, there exists no established method for computing such a covariance matrix.

Several potential techniques for constructing such a matrix were explored in this thesis, including resampling through block bootstrap and parameterizing a dataset utilizing a VARIMA and MGARCH model. Nevertheless, I illustrated the challenges inherent to these methods when seeking to reconcile the goal of estimating the covariance matrix based on the actual variability inherent in the time series.

Although I postulate that the computation of such a covariance matrix is theoretically plausible, it is also likely to involve immense computational requirements, reflecting the intricate nature of autocorrelated data. However, the task of developing a reliable methodology for generating these matrices extends beyond the scope of this Bachelor's thesis and, therefore, demands further investigation.

In the absence of such an advanced method, I conclude that in autocorrelated data, the construction of reliable confidence intervals according to Mogstad et al. (2020) is likely to be imprecise when ranking is intended to be performed by the expected value. Therefore, until such a breakthrough occurs, I recommend that the preferred application of Mogstad et al.'s method be confined to i.i.d. data, to ensure the accuracy and reliability of results.

7.2 Implications of Rankings and their Uncertainty for Portfolio Management

7.2.1 Uncertainty

Backtesting is a fundamental element in portfolio management. Nonetheless, when Mogstad et al. (2020)'s method is utilized to quantify the uncertainty inherent in such backtests, it becomes evident that these rankings can be uninstructive when high confidence is desired. Given the constraints of historical stock market data, with a scarcity of indices extending back more than 50 years, constructing portfolios based on historical performance does not guarantee future outperformance. Upon evaluation of the portfolios discussed in this thesis, only the leveraged strategy developed by Bilello and Gayed (2016) and its amalgamation with gold, introduced in this thesis, appear to be viable portfolio choices for surpassing the S&P 500 throughout all three periods under investigation. Yet, even this assertion holds only if the requisite uncertainty level is maintained at 5%; at a 50% confidence level, the confidence sets of the S&P 500 and these two portfolios begin to intersect.

7.2.2 Factors

Factor premiums are a compelling concept and align with economic logic. However, this analysis does not corroborate the premise that these factors outperform the S&P 500. The rankings also suggest a temporal shift among different factors, such as the diminishing significance of the value factor. This observation could be coherent, given that the recognition of these premiums might channel more capital towards these enterprises. Still, the inherent prime resulting from higher returns remains an open question. Recent research does suggest the continued existence of the small-cap prime, but it seems to have been confined to quality small-cap stocks.
Understanding the value prime, however, remains a challenge. I hypothesize that the recognition of this prime was a consequence of an era characterized by an industrial economy, and it was relevant within that context. However, the global economy has undergone considerable change. The earnings capacity of industrial companies is confined by physical demand and resources, leading to a convergence in the growth of large firms and making high valuations of top-tier companies frequently unjustifiable. Currently, we are observing the phenomenon of platform capitalism, with tech giants demonstrating unbounded growth. As their dependence on physical resources is minimal, these firms' earnings capacity grows exponentially with their value. The increased usage of a software or platform enhances its value to others, notably in the case of social media platforms. Consequently, the marginal benefit of each new user for these firms grows, while the marginal cost for every additional user approximates zero. This dynamic supports higher valuations for the most successful companies, with companies exhibiting lower P/E ratios increasingly perceived as less competitive.
Such theories could potentially explain a perceived fading of factors, as indicated by the lack of consistent outperformance in these rankings at any reasonable confidence level. However, it remains uncertain whether the factors identified by Fama and French (2015) would have demonstrated significance using Mogstad et al. (2020)'s method.

7.2.3 The 'optimal' portfolio

The results of this thesis indicate that the development of an ideal portfolio is a personalized process that should not excessively depend on past performance due to the inherent uncertainty of back tests. However, the ranking suggests that individuals with a greater investment horizon might be attracted to a leverage strategy like the 'S&P 500 3x & Gold' strategy, as the increased volatility could be manageable over a longer period of time. For individuals who experience unease with high levels of volatility or possess a shorter investment time frame, adopting a portfolio characterized by lower volatility, such as the low-beta portfolio, or a combination of bonds and stocks, as exemplified by the evolutionary portfolio, seems to offer a more beneficial approach.

The findings may also generally hold promise for individuals who do not wish to construct a portfolio strictly based on annual returns. If an investor, owing to preference or accessibility, chooses to emphasize particular regions or factors, from a statistical perspective, the certainty of lower expected returns based on diminished historical performance is not high.

A Appendix

CAPM: Capital Asset Pricing model
VaR: Value at Risk
PISA: Program for International Student Assessment
Lev: Leveraged
GPO: Global Portfolio One
SMA200: 200-day moving average
i.i.d.: independent and identically distributed
DW: Durbin-Watson
LB: Ljung-Box
HAC: Heteroskedasticity and Autocorrelation Consistent Method
ARIMA: Autoregressive Integrated Moving Averag
GARCH: Generalized Autoregressive Conditional Heteroskedasticit
MMBB: multivariate moving block bootstrap
S&P 500 3x & Gold: Portfolio holding a 3x daily leveraged S&P 500 if the S&P 500 is above its SMA200 and gold if the S&P 500 is below
S&P 500 3x & Cash: Portfolio holding a 3x daily leveraged S&P 500 if the S&P 500 is above its SMA200 and cash if the S&P 500 is below
CS: Confidence Sets
r: Return
$\bar{\text{r}}$: Average Return
$\widehat{\text{Rank}}_i$: Estimator for the rank of $Portfolio_i$
Rank_i: True rank of $Portfolio_i$

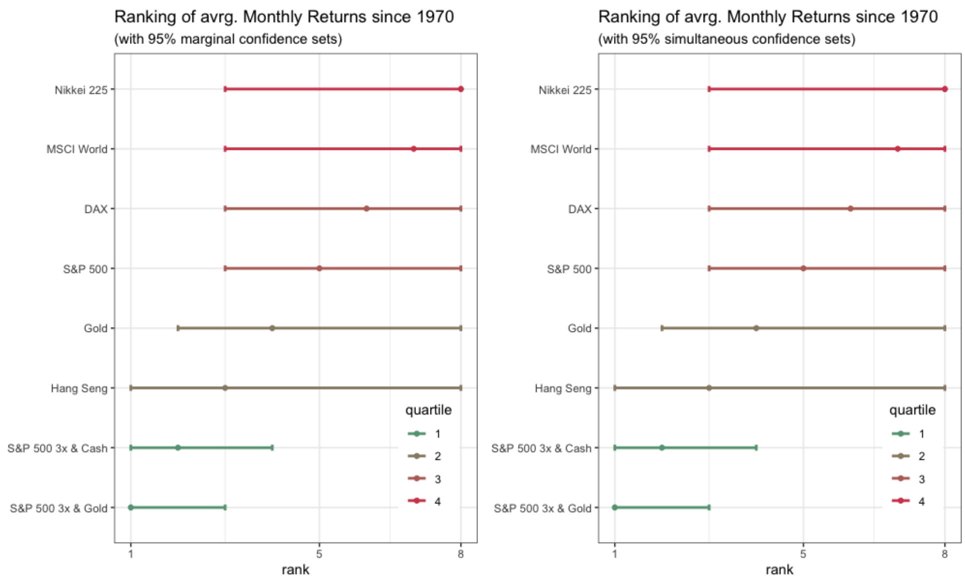

Figure 24: Rankings of avrg. Monthly Returns 1970 - 2022, with marginal and simultaneous 95% confidence set

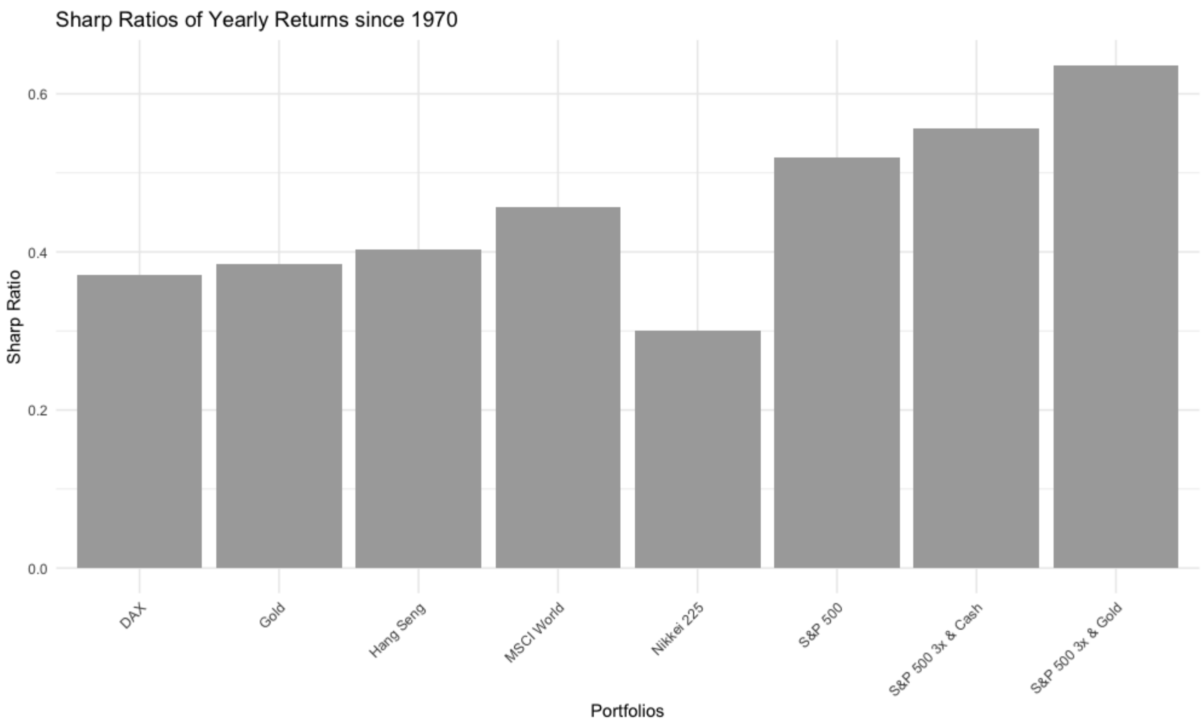

Figure 18: Yearly Return Sharp Ratio 1970 - 2022

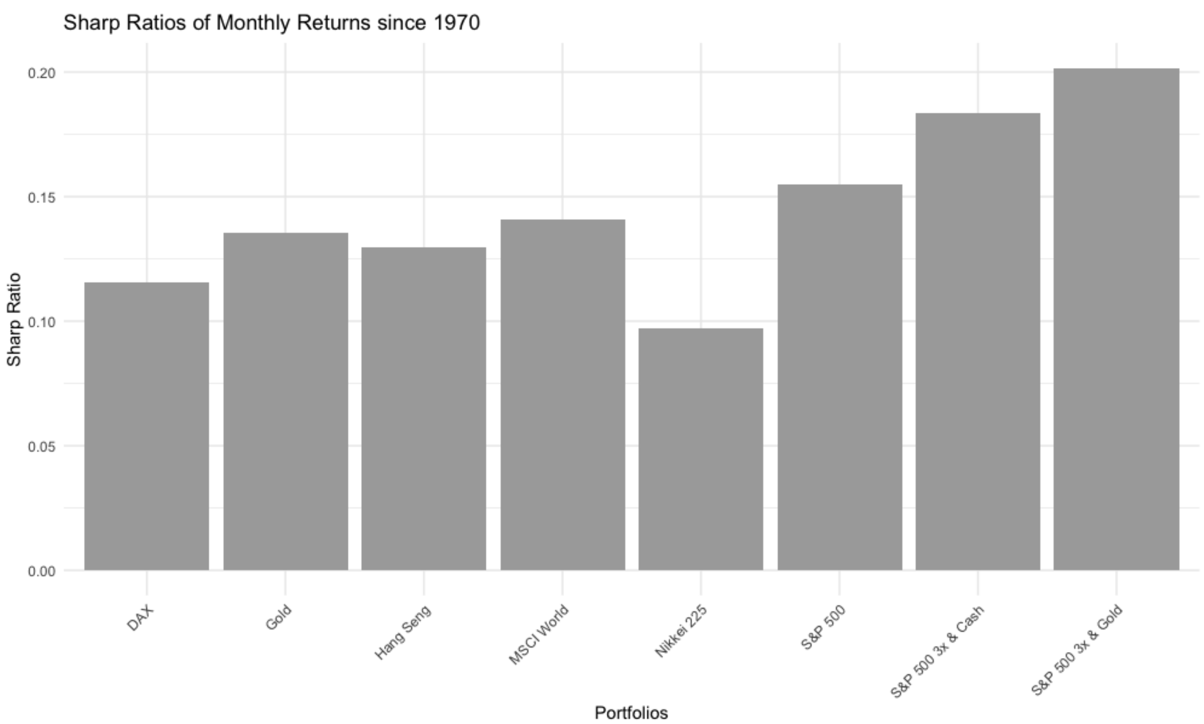

Figure 19: Monthly Return Sharp Ratio 1970 - 2022

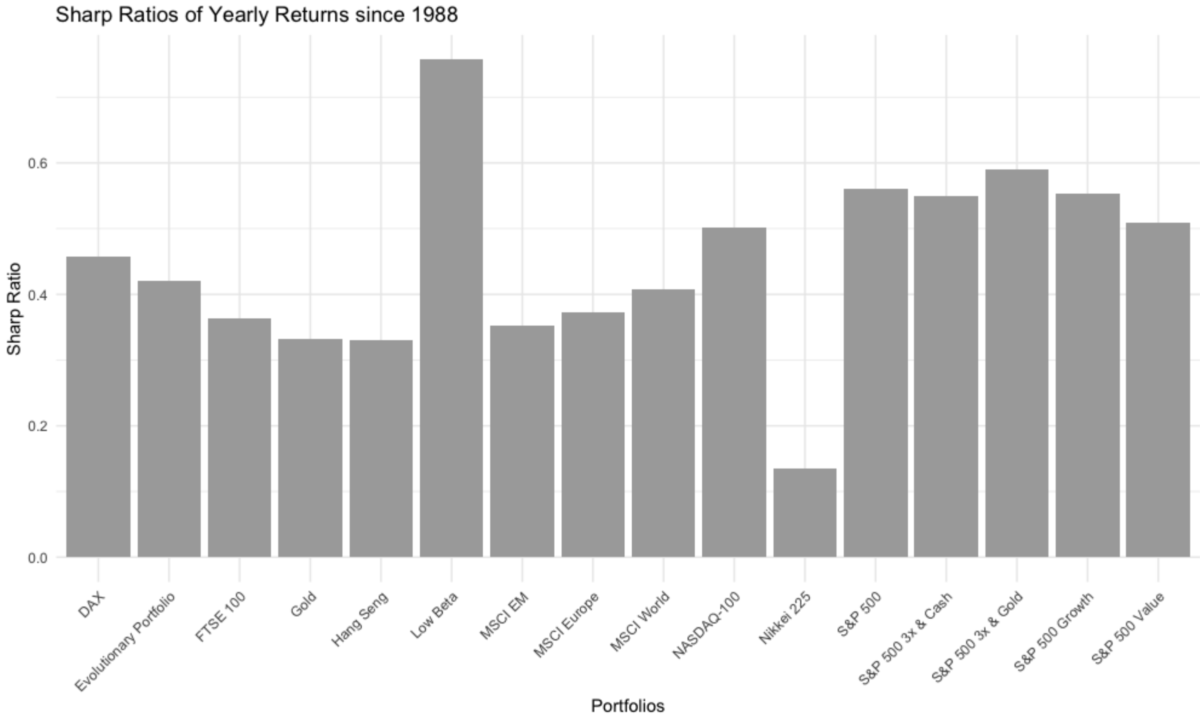

Figure 20: Yearly Return Sharp Ratio 1988 - 2022

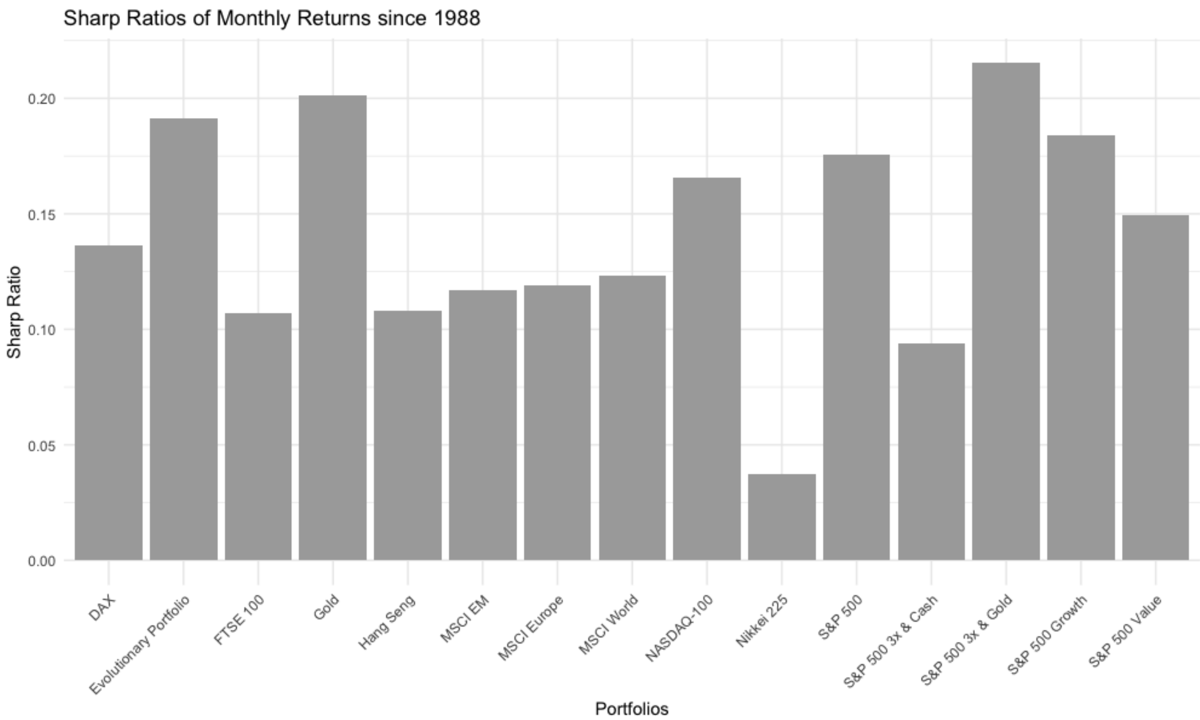

Figure 21: Monthly Return Sharp Ratio 1988 - 2022

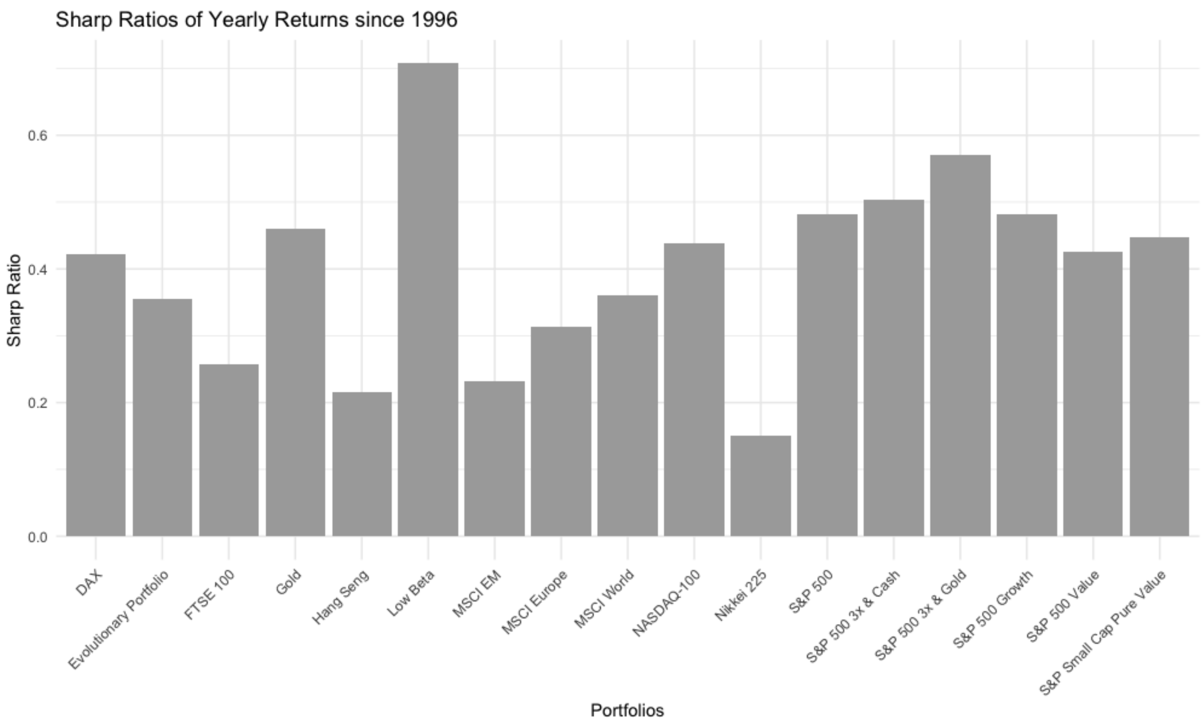

Figure 22: Yearly Return Sharp Ratio 1996 - 2022

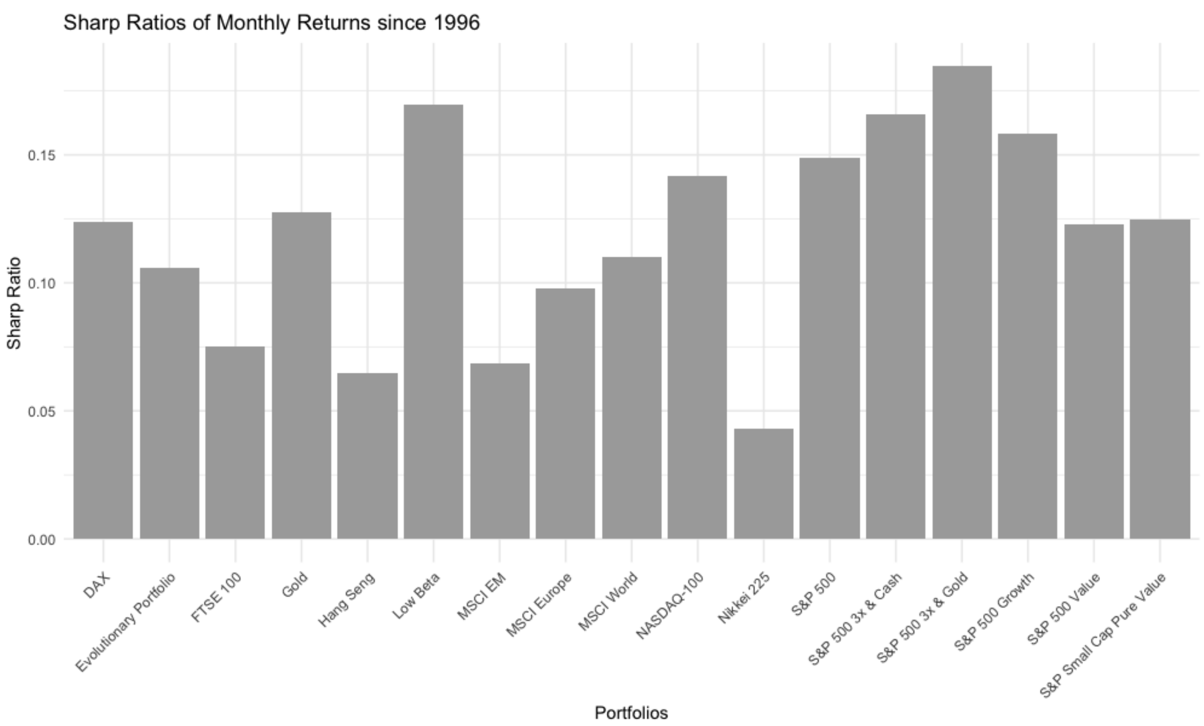

Figure 23: Monthly Return Sharp Ratio 1996 - 2022

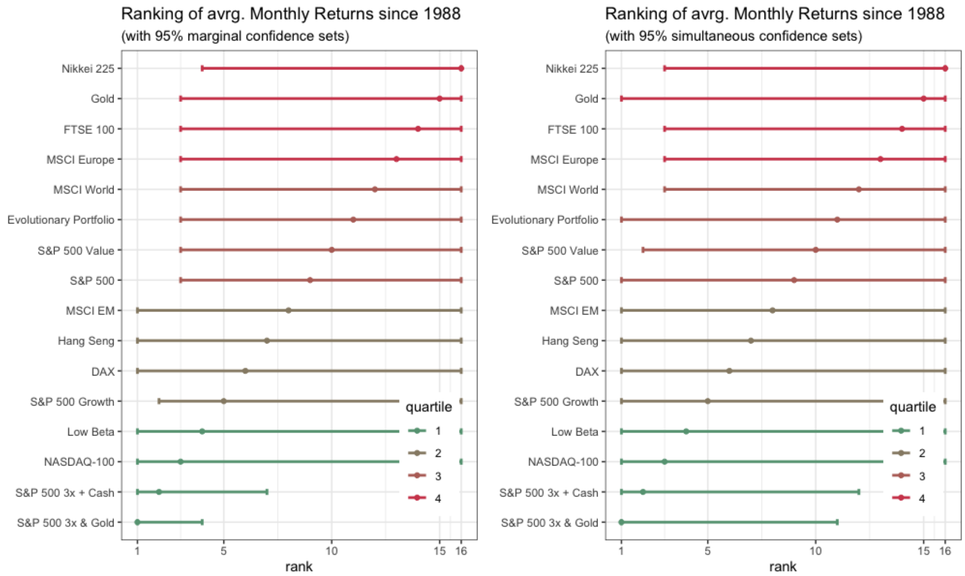

Figure 25: Rankings of avrg. Monthly Returns 1988 - 2022, with marginal and simultaneous 95% confidence set

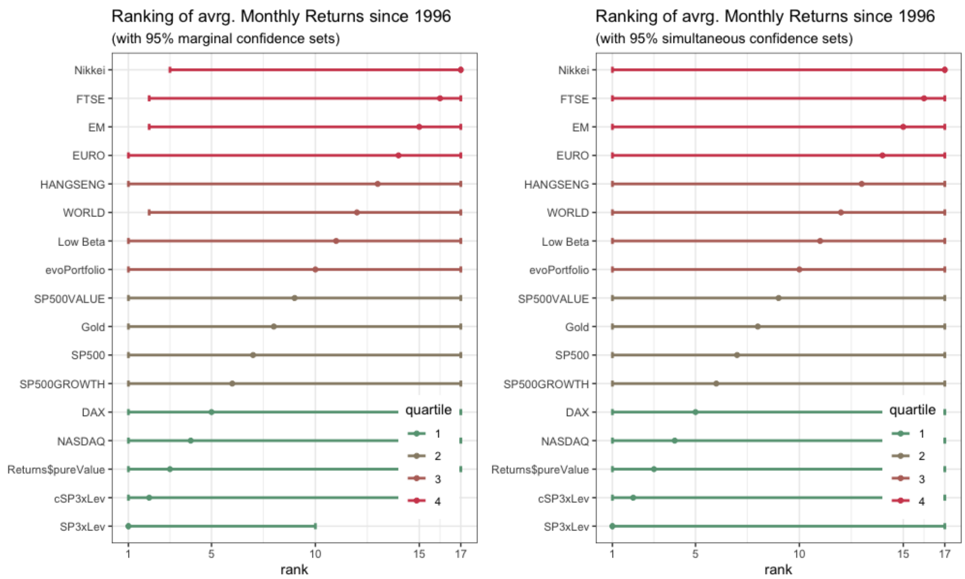

Figure 26: Rankings of avrg. Monthly Returns 1996 - 2022, with marginal and simultaneous 95% confidence set

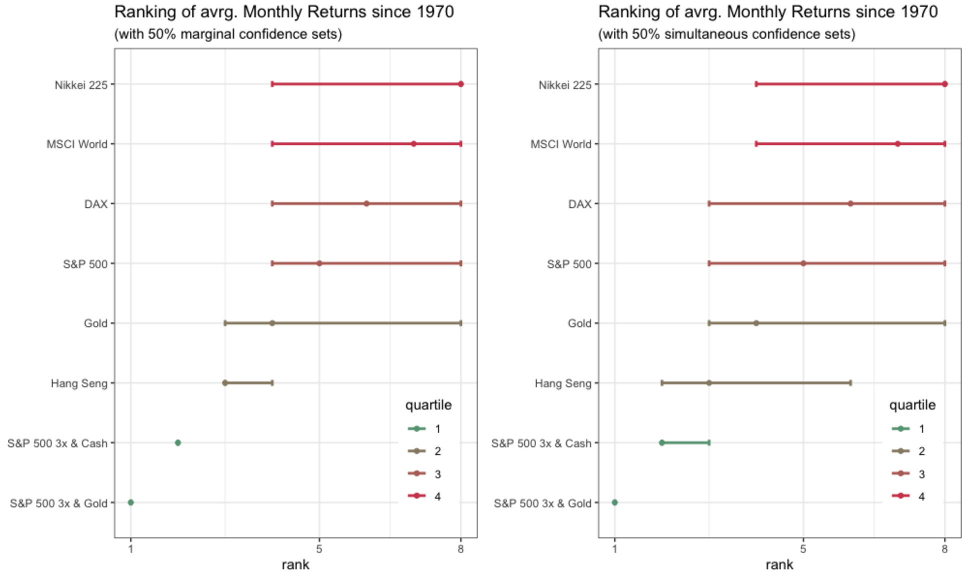

Figure 27: Rankings of avrg. Monthly Returns 1970 - 2022, with marginal and simultaneous 50% confidence set

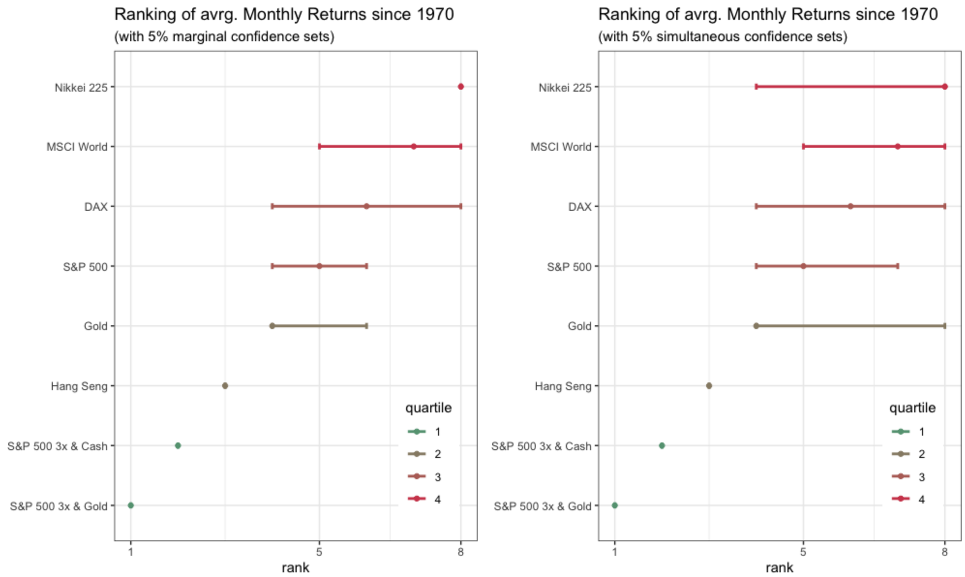

Figure 28: Rankings of avrg. Monthly Returns 1970 - 2022, with marginal and simultaneous 5% confidence set

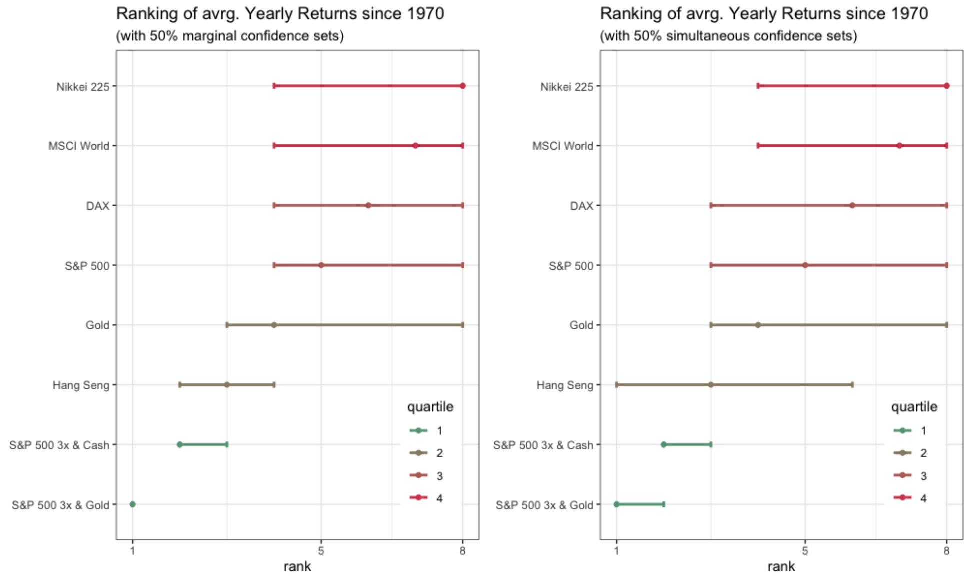

Figure 29: Rankings of avrg. Yearly Returns 1970 - 2022, with marginal and simultaneous 50% confidence set

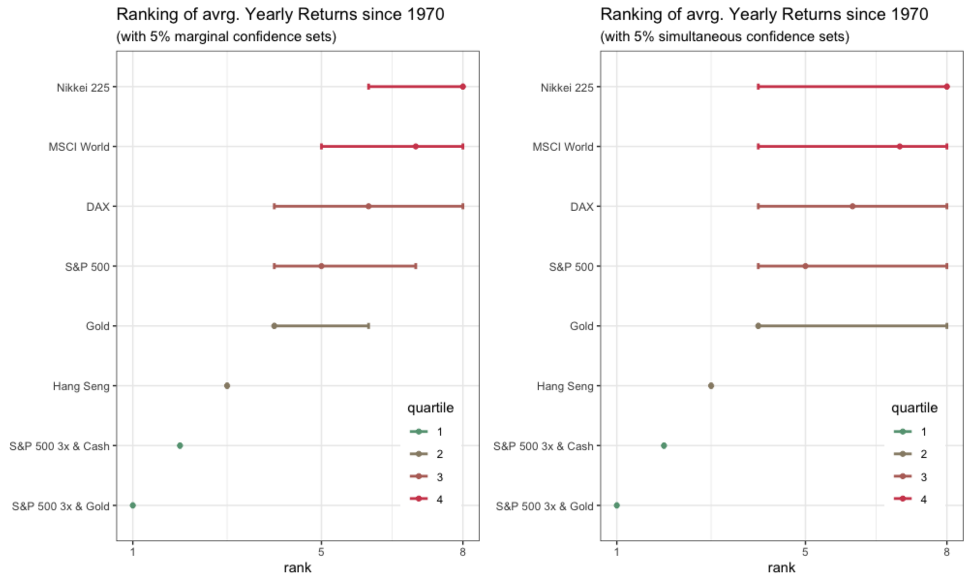

Figure 30: Rankings of avrg. Yearly Returns 1970 - 2022, with marginal and simultaneous 5% confidence set

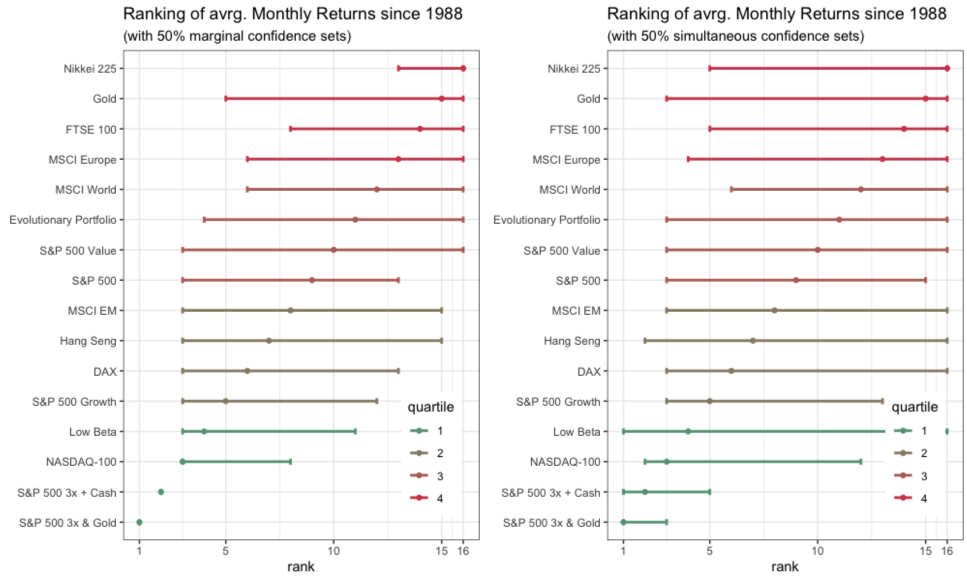

Figure 31: Rankings of avrg. Monthly Returns 1988 - 2022, with marginal and simultaneous 50% confidence set

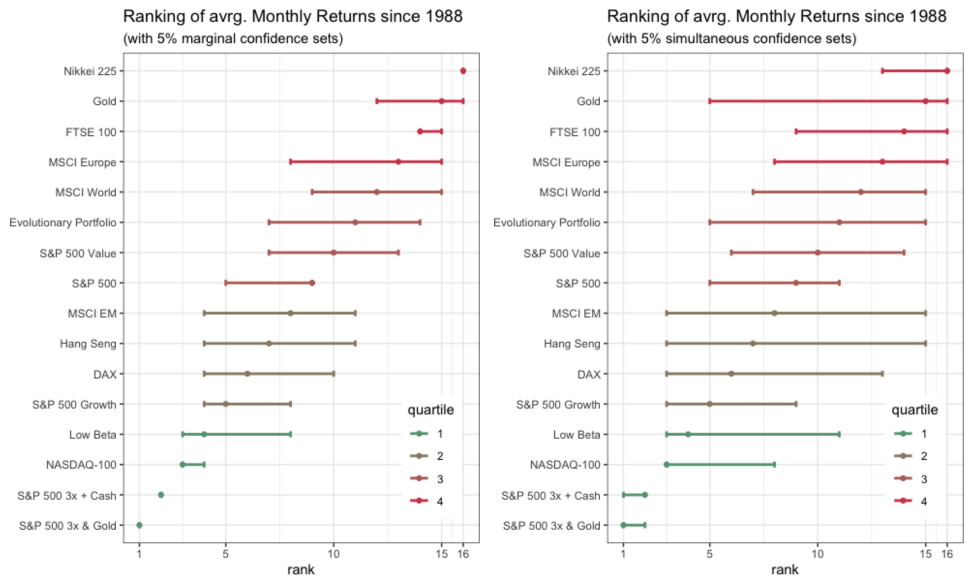

Figure 32: Rankings of avrg. Monthly Returns 1988 - 2022, with marginal and simultaneous 5% confidence set

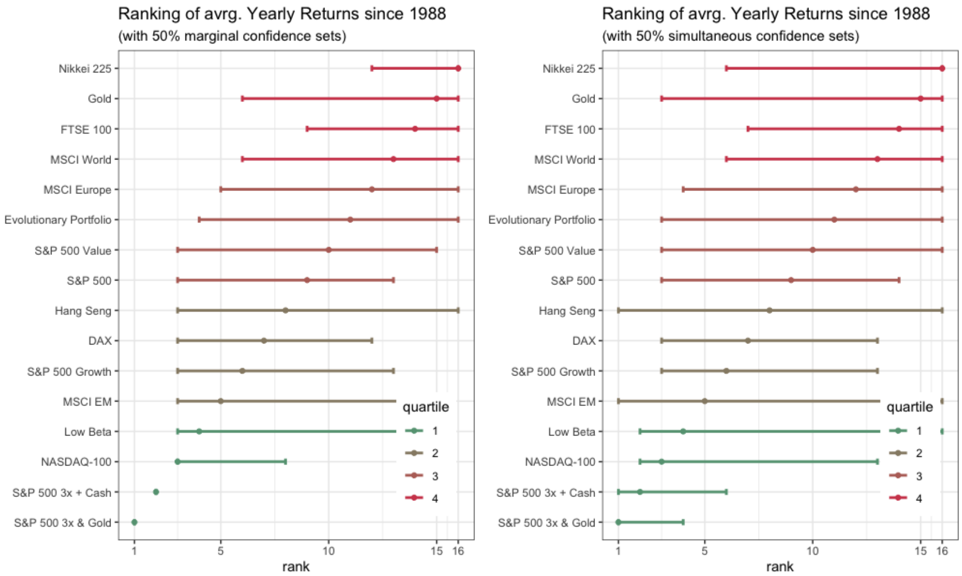

Figure 33: Rankings of avrg. Yearly Returns 1988 - 2022, with marginal and simultaneous 50% confidence set

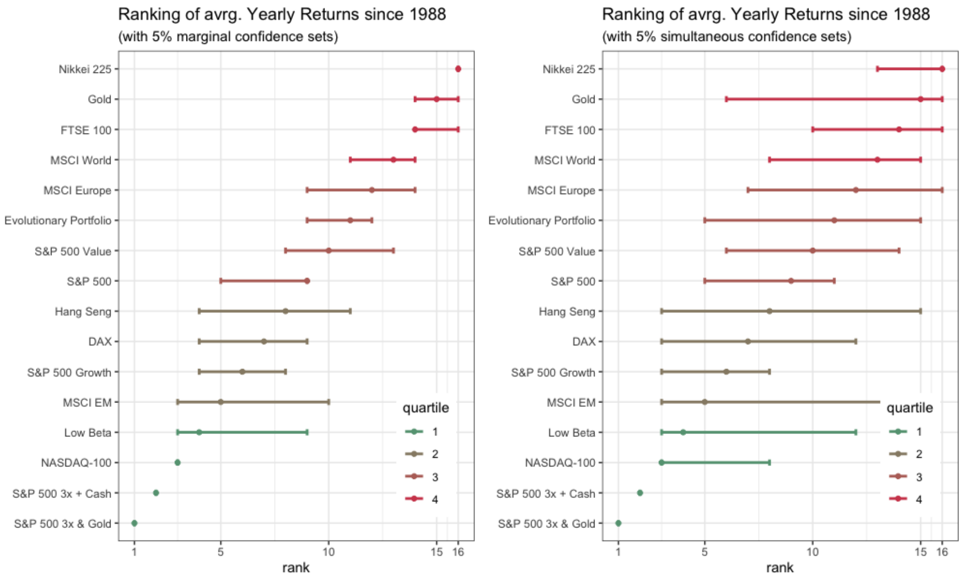

Figure 34: Rankings of avrg. Yearly Returns 1988 - 2022, with marginal and simultaneous 5% confidence set

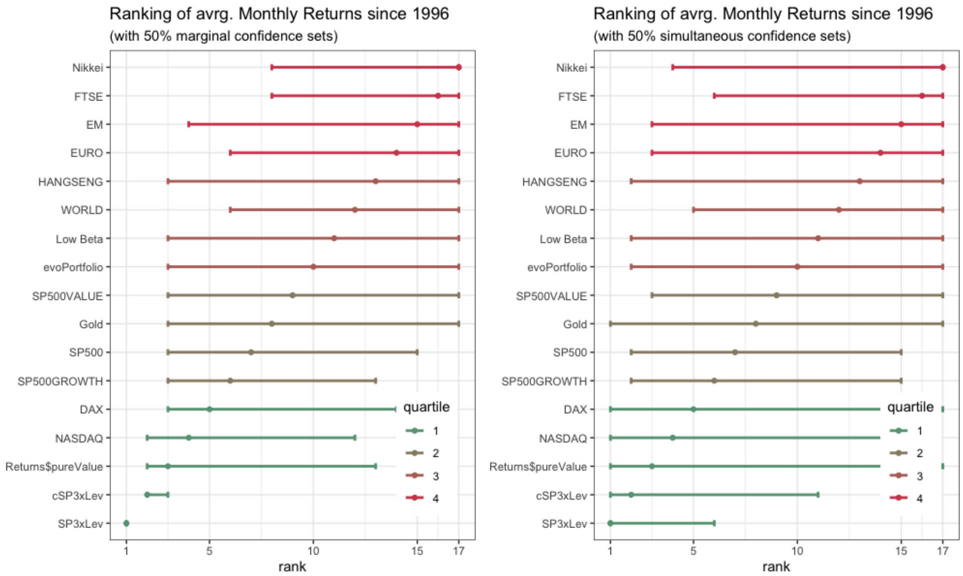

Figure 35: Rankings of avrg. Monthly Returns 1996 - 2022, with marginal and simultaneous 50% confidence set

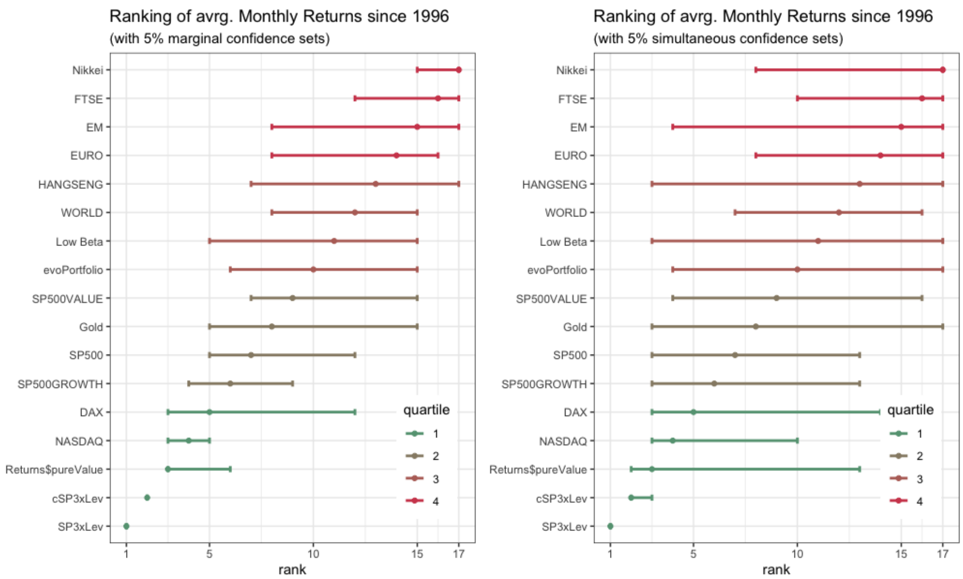

Figure 36: Rankings of avrg. Monthly Returns 1996 - 2022, with marginal and simultaneous 5% confidence set

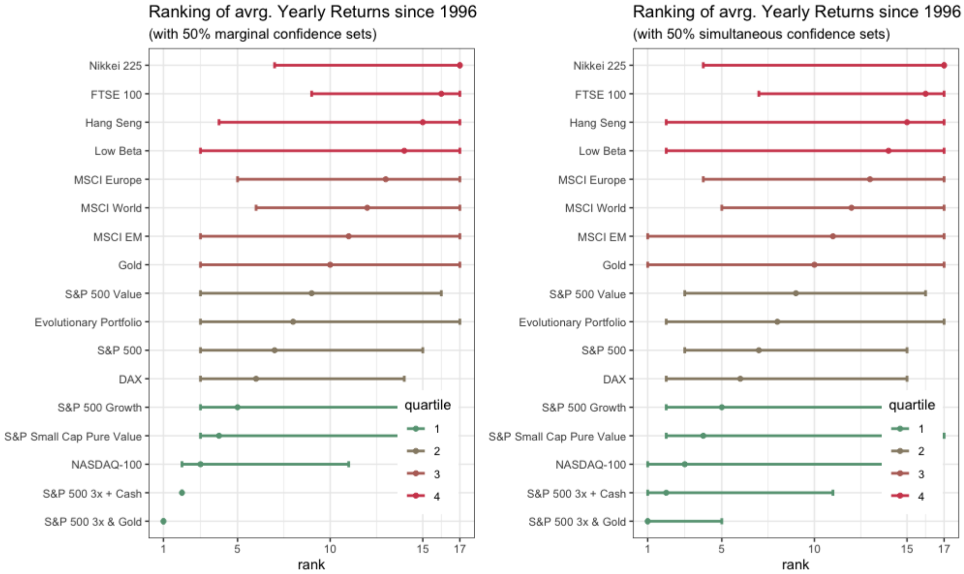

Figure 37: Rankings of avrg. Yearly Returns 1996 - 2022, with marginal and simultaneous 50% confidence set

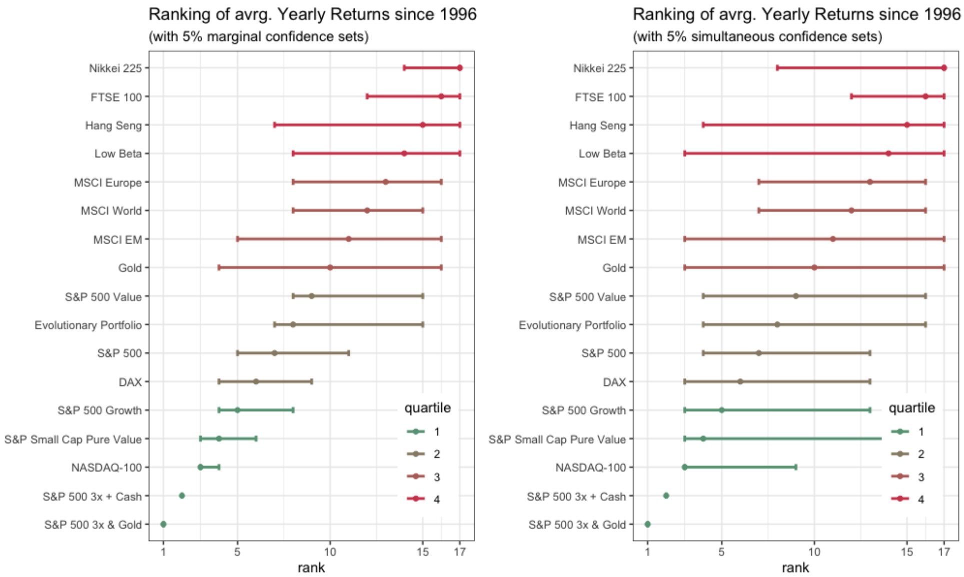

Figure 38: Rankings of avrg. Yearly Returns 1996 - 2022, with marginal and simultaneous 5% confidence set

Portfolios since 1988	**DW Statistic**	**P Value DW**	**P Value LB**
Monthly Return MSCI World	1.993	0.918	0.617
Monthly Return EM	2.011	0.908	0.661
Monthly Return FTSE	1.997	0.984	0.236
Monthly Return EURO	1.989	0.900	0.574
Monthly Return NASDAQ	1.992	0.996	0.951
Monthly Return S&P 500	1.997	0.914	0.572
Monthly Return DAX	1.968	0.728	0.856
Monthly Return HANG SENG	1.999	0.984	0.793
Monthly Return SP500 Value	1.998	0.940	0.649
Monthly Return S&P500 Growth	1.993	0.992	0.496
Monthly Return Nikkei 225	2.001	0.946	0.866
Monthly Return Low Beta	2.012	0.930	0.028
Monthly Return S&P500 Lev Gold	2.001	0.956	0.933
Monthly Return Gold	2.010	0.884	0.780
Monthly Return S&P500 Lev Cash	1.999	0.974	0.926
Monthly Return Evolutionary Portfolio	1.992	0.938	0.673
Yearly Return MSCI World	2.021	0.952	0.209
Yearly Return EM	1.930	0.892	0.766
Yearly Return FTSE	1.866	0.754	0.462
Yearly Return EURO	1.988	0.992	0.350
Yearly Return NASDAQ	1.966	0.906	0.247
Yearly Return S&P500	1.899	0.798	0.197
Yearly Return DAX	2.011	0.976	0.164
Yearly Return HANG SENG	2.025	0.944	0.608
Yearly Return S&P500 Value	2.011	0.962	0.128
Yearly Return S&P500 Growth	1.827	0.572	0.257
Yearly Return Nikkei 225	2.044	0.920	0.559
Yearly Return Low Beta	1.864	0.632	0.356
Yearly Return S&P500 Lev Gold	1.880	0.736	0.221
Yearly Return Gold	2.049	0.966	0.954
Yearly Return S&P500 Lev Cash	1.877	0.670	0.258
Yearly Return Evolutionary Portfolio	2.132	0.680	0.204

Table 7: Results of Durbin-Watson Test & Ljung-Box Test for Portfolios from 1988 - 2022

Portfolios since 1996	**DW Statistic**	**P Value DW**	**P Value LB**
Monthly Return MSCI World	1.989	0.890	0.470
Monthly Return EM	2.007	0.972	0.838
Monthly Return FTSE	1.999	0.958	0.467
Monthly Return EURO	1.997	0.992	0.807
Monthly Return NASDAQ	1.990	0.932	0.980
Monthly Return S&P500	1.992	0.946	0.465
Monthly Return DAX	1.996	0.938	0.998
Yearly Return HANG SENG	1.992	0.936	0.911
Monthly Return S&P500 Value	1.991	0.892	0.603
Monthly Return S&P500 Growth	1.991	0.982	0.353
Monthly Return Nikkei 225	1.993	0.934	0.928
Monthly Return Low Beta	2.012	0.934	0.067
Monthly Return S&P500 Lev Gold	2.000	0.992	0.975
Monthly Return Gold	2.003	0.990	0.827
Monthly Return S&P500 Lev Cash	1.999	0.998	0.973
Monthly Return Evolutionary Portfolio	1.992	0.948	0.572
Monthly Return S&P500 Small Cap Pure Value	1.990	0.924	0.915
Yearly Return MSCI World	1.988	0.990	0.205
Yearly Return EM	2.029	0.872	0.639
Yearly Return FTSE	1.983	0.952	0.238
Yearly Return EURO	1.960	0.904	0.322
Yearly Return NASDAQ	1.943	0.892	0.500
Yearly Return S&P500	1.906	0.778	0.389
Yearly Return DAX	1.962	0.936	0.075
Yearly Return HANG SENG	2.136	0.716	0.182
Yearly Return S&P500 Value	2.027	0.902	0.098
Yearly Return S&P500 Growth	1.826	0.646	0.535
Yearly Return Nikkei 225	1.944	0.826	0.174
Yearly Return Low Beta	1.953	0.874	0.331
Yearly Return S&P500 Lev Gold	1.855	0.684	0.299
Yearly Return Gold	1.991	0.842	0.997
Yearly Return S&P500 Lev Cash	1.889	0.804	0.390
Yearly Return Evolutionary Portfolio	2.106	0.730	0.180
Yearly Return S&P500 Small Cap Pure Value	2.131	0.716	0.125

Table 8: Results of Durbin-Watson Test & Ljung-Box Test for Portfolios from 1996 - 2022

B Electronic appendix

Data, code and figures are provided in electronic form on Github.

References

Asness, C., Frazzini, A., Israel, R., Moskowitz, T. J. and Pedersen, L. H. (2018). Size matters, if you control your junk, *Journal of Financial Economics* **129**(3): 479–509.

Asness, C. S., Moskowitz, T. J. and Pedersen, L. H. (2013). Value and momentum everywhere, *The journal of finance* **68**(3): 929–985.

Baur, D. G. and Lucey, B. M. (2010). Is gold a hedge or a safe haven? an analysis of stocks, bonds and gold, *Financial review* **45**(2): 217–229.

Beck, A. (2021). Erfolgreich wissenschaftlich investieren.

Bilello, C. and Gayed, M. (2016). Leverage for the long run: A systematic approach to managing risk and magnifying returns in stocks., *Journal of Technical Analysis* (70).

Bodie, Z., Kane, A. and Marcus, A. (2013). *Ebook: Essentials of investments: Global edition*, McGraw Hill.

Carhart, M. M. (1997). On persistence in mutual fund performance, *The Journal of finance* **52**(1): 57–82.

Cronqvist, H., Siegel, S. and Yu, F. (2015). Value versus growth investing: Why do different investors have different styles?, *Journal of Financial Economics* **117**(2): 333–349.

Draper, D. (1995). Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **57**(1): 45–70.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I, *Biometrika* **37**(3/4): 409–428.

Efron, B. (1979). *Bootstrap methods: another look at the jackknife*, Springer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica: Journal of the econometric society* pp. 987–1007.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work, *The journal of Finance* **25**(2): 383–417.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns, *the Journal of Finance* **47**(2): 427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds, *Journal of financial economics* **33**(1): 3–56.

Fama, E. F. and French, K. R. (2004). The capital asset pricing model: Theory and evidence, *Journal of economic perspectives* **18**(3): 25–46.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model, *Journal of financial economics* **116**(1): 1–22.

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta, *Journal of financial economics* **111**(1): 1–25.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian data analysis*, CRC press.

Hens, T. (2021). Die evolution für sich arbeiten lassen, *Portfolio Journal* **10-2021**.

Hens, T. and Schenk-Hoppé, K. R. (2001). An evolutionary portfolio theory.

Hurra, T. (2022). Ljung-box or durbin watson—which test is more powerful.

Jagannathan, R., McGrattan, E. R. et al. (1995). The capm debate, *Federal Reserve Bank of Minneapolis Quarterly Review* **19**(4): 2–17.

Jorion and Philippe (2007). *Value at risk: the new benchmark for managing financial risk*, The McGraw-Hill Companies, Inc.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models, *The review of economic studies* **65**(3): 361–393.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics* **47**(1): 13–37.
URL: *http://www.jstor.org/stable/1924119*

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models, *Biometrika* **65**(2): 297–303.

Ludwig Fahrmeir, Rita Künstler, G. T. I. P. C. H. (1999). Statistik: Der weg zur datenanalyse.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics, *Journal of economic perspectives* **17**(1): 59–82.

Malladi, R. and Fabozzi, F. J. (2017). Equal-weighted strategy: Why it outperforms value-weighted strategies? theory and evidence, *Journal of Asset Management* **18**: 188–208.

Markowitz, H. (1952). Portfolio selection, *The Journal of Finance* **7**(1): 77–91.
URL: *http://www.jstor.org/stable/2975974*

Metropolis, N. and Ulam, S. (1949). The monte carlo method, *Journal of the American statistical association* **44**(247): 335–341.

Mogstad, M., Romano, J. P., Shaikh, A. and Wilhelm, D. (2020). Inference for ranks with applications to mobility across neighborhoods and academic achievement across countries, *NATIONAL BUREAU of ECONOMIC RESEARCH* .

Romano, J. P. (1992). A circular block-resampling procedure for stationary data, *Exploring the Limits of Bootstrap* **270**: 263.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap.

Sauerbrei, W. and Schumacher, M. (1992). A bootstrap resampling procedure for model building: application to the cox regression model, *Statistics in medicine* **11**(16): 2093–2109.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *The journal of finance* **19**(3): 425–442.

Sharpe, W. F. (1991). The arithmetic of active management, *Financial Analysts Journal* **47**(1): 7–9.

Shumway, R. H., Stoffer, D. S. and Stoffer, D. S. (2000). *Time series analysis and its applications*, Vol. 3, Springer.

SP500 Growth (n.d.).
URL: *https://www.spglobal.com/spdji/en/indices/equity/sp-500-growth/overview*

SP500 Value (n.d.).
URL: *https://www.spglobal.com/spdji/en/indices/equity/sp-500-value/overview*

Thompson, S. K. (2012). *Sampling*, Vol. 755, John Wiley & Sons.

Declaration of authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the Thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the report as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future Theses submitted. Further rights of reproduction and usage, however, are not granted here. This thesis was not previously presented to another examination board and has not been published.

Munich, 08/08/2023

Jonas Schernich