Machine Beta, Statistical Factors, Non-Linear Mechanisms And The [3N] Methodology

Mukul Pal, Ciprian Radu Tiric, Florina Pal

December 24, 2023

Abstract

The authors introduce the concept of Machine Beta as an advanced approach to address the challenges faced by traditional Smart Beta strategies. This methodology pivots on the use of Statistical factors and Non-Linear Mechanisms to counteract the biases inherent in market capitalization-weighted benchmarks. The primary focus of Machine Beta is to achieve lower tracking errors while outperforming these benchmarks. The document underscores the potential of Machine Beta in reshaping index construction, steering towards benchmarks that are more aligned with performance rather than market cap concentration. Notably, the paper details an idealized case study demonstrating how Machine Beta, underpinned by Statistical factors, can lead to risk-weighted excess returns near 400 bps above S&P500. This case study serves as a practical illustration of the effectiveness of the Machine Beta approach across various asset classes and regions, showcasing its potential to redefine investment strategies and advancing the research using machine learning.

The authors wish to express their profound gratitude to Alexandru George Berciu, Prof. Eva Dulf, Patricia Ratiu, and Dan Todor for their invaluable support, guidance, and invaluable inputs, which significantly contributed to the completion of this research.

I – Introduction

-Bias

Today's Indexes which drive the $20 T passive asset management industry [1] are predominantly based on the Market Capitalization (MCAP) weighting methodology, which drives it calculation from the 1871 [2] work of Ernst Louis Etienne Laspeyres and the chain method [3]. The Laspeyres is pegged to a starting point while the chain method allows for updating the components and quantities. Morgan Stanley Capital International (MSCI) credits the Laspeyres method in its Indexing methodology [4]. The Laspeyres method was adopted because of calculation convenience and because of its intrinsic positive bias [5], which inflated Index value and hence gathered mass popularity owing to increasing inflated value [6]. The Laspeyres Index tends to overstate value if prices increase for goods that consumers continue buying in large quantities. It understates value if consumers substitute expensive things with cheaper alternatives, as it does not account for the reduced consumption of the more expensive things, a problem similar to what is observed in MCAP Indexes [7].

The historical context for the influence of the Laspeyres index's methodology and biases on modern equity indexes is rooted in the broader evolution of financial measurement tools. Initially, price indices like the Laspeyres provided a way to track economic trends through a fixed basket of goods [8], making them easier to calculate and understand. This approach, despite its bias towards overstating value due to fixed quantities, was appealing for its simplicity. As financial markets evolved, there was a need for averaging a more dynamic value, leading to the development of MCAP method for equity indexes [9]. These methods, while more representative of real-time market values, indirectly inherited the notion of emphasizing increased value from Laspeyres, seen as advantageous in market contexts where growth perception was key.

However over the last 100 years it has been clear that MCAP methodology is weight obsessive [10] as it changes its weight with every change in price and creates extremely concentrated benchmarks (currently referred to as magnificent seven), which misrepresents the market, increases herding, consequently increasing risks, and hence challenging capital market integrity. Above this leading Indexing companies propagate this accidental [11] winner's bias, concentration, as outperformance against Active managers creating a myth about the Index's invincibility (unbeatability).

-Governance

The Laspeyres Price Index, known for measuring price changes over time based on a fixed basket of components, can exhibit biases in certain market conditions. For instance, if prices rise for components that are heavily consumed, the index might overstate value changes. Conversely, in situations where consumers opt for cheaper alternatives, leading to reduced consumption of more expensive components, the index can understate these value shifts.

This aspect becomes particularly pertinent in the context of stock indices like the S&P 500. The decision in 2014 [12] by S&P Dow Jones Indices to include both classes of Google shares (GOOG and GOOGL) in the index, thereby expanding the number of components to 501, is a prime example. Typically, companies with multiple stock classes are represented only once in such indices, based on the most liquid stock class. This decision effectively doubled Google's representation in the index, increasing its concentration from 2% to 4%.

This scenario underscores the broader issues of bias and governance in financial indexing. By increasing Google's weight in the S&P 500, the index potentially exacerbates concentration and winner's bias, possibly leading to a distorted market representation. Such decisions not only raise questions about the governance standards of indexing companies but also about the implications for market integrity. They can force the industry to disproportionately invest in these more heavily weighted components, further amplifying the existing biases in the index.

-Fisher’s Index and Time Reversal Test

The knowledge about the bias has been published before Irving Fisher published, “The Making of Index Numbers” in 1922 [13]. Fisher articulated it in his masterwork how Laspeyres fails the Indexing Time Reversal test. Below the author’s explain the superiority of the Fisher’s Index and the calculation of how when you introduce multiple stocks of the same company, MCAP index just like Laspeyres fail the time reversal test. Fisher's price index, named after the economist, is a measure used to calculate the changes in the overall price level of a basket of goods and services over time. It is one of the methods employed in economics to calculate inflation or deflation.

The Fisher Price Index, as represented in Equation (1) from the document, combines two well-known price indices - the Laspeyres Price Index and the Paasche Price Index - into a single index, known as the Fisher Price Index. The Fisher Price Index is essentially the geometric mean of these two indices.

$$Fisher\ Price\ Index = \sqrt{Laspeyres\ Price\ Index\ \times Paasche\ Price\ Index} \qquad - 1$$

The Laspeyres Price Index in Equation (2) looks at how the cost of a fixed basket of goods and services, selected from a base period, would change over time. It tells us how much more or less expensive it would be to buy the same items today as compared to the past.

$$Laspeyres\ Price\ Index = \frac{\sum P_{i,t} \times Q_{i,0}}{\sum P_{i,0} \times Q_{i,0}} \times 100 \qquad - 2$$

The Paasche Price Index in Equation (3), on the other hand, calculates how much it would cost to buy a current basket of goods and services (reflecting current tastes and preferences) at past prices. This gives us an idea of how price changes affect our current shopping habits.

$$Paasche\ Price\ Index = \frac{\sum P_{i,t} \times Q_{i,t}}{\sum P_{i,0} \times Q_{i,t}} \times 100 \qquad - 3$$

However, unlike the Laspeyres Index, Fisher's price index takes both price changes and quantity changes of items in the basket of goods. By considering both factors, it provides a more accurate measure of inflation compared to Laspeyres and some other methods that only consider price changes. Fisher's Price Index is unique because it addresses a key limitation of some other price indices, such as the Laspeyres index and the Paasche index.

The Laspeyres index uses fixed quantities from the base period, ignoring changes in quantities consumed over time. The Paasche index, on the other hand, uses current quantities, ignoring the base period quantities. Fisher's index addresses this limitation by using a geometric mean of the Laspeyres and Paasche indices, thus accounting for both price changes and quantity changes.

In summary, Fisher's Price Index is a single index formula that incorporates elements from both the Laspeyres and Paasche indices, making it a more accurate measure of changes in the cost changes in purchasing power. of living over time. It is widely used in economics and finance to calculate inflation rates and adjust for changes in purchasing power.

Table (1) shows the base year quantities and prices for certain components or commodities. In the following context, "Year 0" acts as the reference or base year against which future changes in prices and quantities are measured.

Year 0	*$ Price*	*Quantity*
Item A	20	50
Item B	8	80
Item C	17	100

Table 1 – Price and quantity of different items in Year 0

Equation (4) represents the Laspeyres price index calculation where the numerator and the denominator are the same, leading to a result of 100. The values are calculated based on the set of prices and quantities in the base Year 0 when prices remain unchanged. Therefore, the index value stays at a constant 100, indicating no change in the price level.

$$Laspeyres\ Price\ Index = \frac{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)}{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)} \times 100 = 100 \qquad - 4$$

Equation (5) represents the Paasche price index calculation, where, similarly to Equation iv, the result is 100. This outcome is derived using the set of prices and quantities from the base year (Year 0) data.

$$Paasche\ Price\ Index = \frac{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)}{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)} \times 100 = 100 \qquad - 5$$

Equation (6) is the square root of the product of two numbers, both 100, to calculate the Fisher Price Index. The result of this operation is 100.

$$Fisher\ Price\ Index = \sqrt{100 \times 100} = 100 \qquad - 6$$

Table (2) is used to compare the prices and quantities of these items from the base year (Year 0) to Year 1. This comparison is essential for calculating the updated price indices such as the Laspeyres, Paasche, and Fisher indices.

Year 1	*$ Price*	*Quantity*
Item A	22	60
Item B	15	70
Item C	18	105

Table 2 – Price and quantity of different items in Year 1

Equation (7) calculates the Laspeyres Price Index for Year 1. The Laspeyres Index measures the cost of purchasing a specific basket of goods at the current year's prices relative to the base year's prices. It uses the base year quantities and the current year prices. In this case, the index value is calculated as 122.75, indicating an increase in the cost of the base year basket when priced at Year 1 prices compared to the base year. This increase suggests inflation or a general rise in prices for these items.

$$Laspeyres\ Price\ Index = \frac{(22\$ \times 50)+(15\$ \times 80)+(18\$ \times 100)}{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)} \times 100 = 122.75 \quad -7$$

Equation (8) represents the calculation of the Paasche Price Index for Year 1. The Paasche Index compares the cost of the current year's basket of goods at current year prices to the base year's prices. It uses the current year quantities and prices. The resulting index value here is 127.54, which also indicates an increase in the cost of purchasing the current year's basket compared to the base year. This higher index value, compared to the Laspeyres Index, might suggest changes in consumption patterns or a shift towards more expensive items in the current year.

$$Paasche\ Price\ Index = \frac{(22\$ \times 60)+(15\$ \times 70)+(18\$ \times 105)}{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)} \times 100 = 127.54 \quad -8$$

Equation (9) calculates the Fisher Price Index for Year 1. The Fisher Index is the geometric mean of the Laspeyres and Paasche indices. In this case, it is the square root of the product of the Laspeyres index (122.75) and the Paasche index (127.54). The result is 125.12, reflecting a balanced measure of price changes considering both the base year and current year quantities and prices. The Fisher Index is often considered a more accurate reflection of price changes as it mitigates the biases inherent in the Laspeyres and Paasche indices.

$$Fisher\ Price\ Index = \sqrt{122.75 \times 127.5} = 125.12 \quad -9$$

Table (3) provides the prices and quantities of different items in Year 2.

Year 2	$ Price	Quantity
Item A	24	55
Item B	14	87
Item C	19	103

Table 3 – Price and quantity of different items in Year 2

Equation (10) calculates the Laspeyres Price Index for Year 2. This index measures the cost of purchasing the base year's basket of goods at Year 2 prices. It uses base year quantities with Year 2 prices, providing insight into how the prices have changed from the base year to Year 2. The index value is calculated as 126.34, suggesting an increase in the cost of the base year basket when priced at Year 2 prices, indicative of inflation or price increases for these items.

$$Laspeyres\ Price\ Index = \frac{(24\$ \times 50)+(14\$ \times 80)+(19\$ \times 100)}{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)} \times 100 = 126.34 \qquad - 10$$

Equation (11) represents the Paasche Price Index for Year 2. This index compares the cost of purchasing the current year's (Year 2) basket of goods at Year 2 prices to the base year's prices. It uses Year 2 quantities and prices, showing how the cost of the current basket has changed compared to the base year. The index value is 134.58, which also points to an increase in the cost of purchasing the current basket compared to the base year. This could reflect changes in consumption patterns or a shift towards more expensive items.

$$Paasche\ Price\ Index = \frac{(24\$ \times 55)+(14\$ \times 87)+(19\$ \times 103)}{(20\$ \times 50)+(8\$ \times 80)+(17\$ \times 100)} \times 100 = 134.58 \qquad - 11$$

Equation (12) calculates the Fisher Price Index for Year 2. The Fisher Index is a balanced measure, being the geometric mean of the Laspeyres and Paasche indices. It is calculated as the square root of the product of the Laspeyres index (126.34) and the Paasche index (134.58). The result, 130.39, reflects a comprehensive view of price changes, considering both base year and current year quantities and prices. The Fisher Index is often considered a more accurate reflection of price changes as it mitigates the biases of both the Laspeyres and Paasche indices.

$$Fisher\ Price\ Index = \sqrt{126.34 \times 134.58} = 130.39 \qquad - 12$$

Table (4) provides a comparative view of three different price indices – Laspeyres, Paasche, and Fisher – over three consecutive years: Year 0, Year 1, and Year 2.

Price Index	*$ Year 0*	*$ Year 1*	*$ Year 2*
Laspeyres	100.00	122.75	126.34
Paasche	100.00	127.54	134.58
Fisher	100.00	125.12	130.39

Table 4 - Price and quantity of different items in Year 0, Year 1 and Year 2.

Fisher made a careful study of the various proposals for computing index numbers in his 1922 masterwork and had suggested various tests to be applied to any formula to indicate whether or not it is satisfactory. The two most important of these he called the Time Reversal Test and the Factor Reversal Test.

Time Reversal Test is a test to determine whether a given method will work both ways in time, forward and backward. The test is that the formula for calculating the index number should be such that it will give the same ratio between one point of comparison and the other, no matter which of the two is taken as base. In other words, when the data for any two years are treated by the same method, but with the bases reversed, the two index numbers secured should be reciprocals of each other so that their product is unity. Symbolically, the following relation should be satisfied.

Equation (13) in the document relates to the Time Reversal Test. This test checks whether a price index method is symmetrical over time. Symbolically, it's expressed as the product of two indices, calculated with reversed bases, being equal to 1. Neither the Laspeyres nor the Paasche methods meet this test, indicating a potential time bias in these methods.

$$P_{0,1} \times P_{1,0} = 1 \qquad - 13$$

Equation (14) shows how multiplying $P_{0,1}$, the index for time “1” on time “0” as base and $P_{1,0}$, the index for time “0” on time “1” as base. If the product is not unity as shown in Equation (15), there is said to be a Time bias in the method. The Time Reversal Test, as proposed by Professor Irving Fisher, is not met by both the Laspeyres and Paasche methods.

$$P_{0,1} = \frac{\sum p_1 q_0}{\sum p_0 q_0}, P_{1,0} = \frac{\sum p_0 q_1}{\sum p_1 q_1} \qquad - 14$$

$$P_{0,1} \times P_{1,0} \neq 1 \qquad - 15$$

When the Laspeyres method is employed with a case where quantities of one component like Google are increased from one class of shares to two classes (GOOG and GOGL), the method fails the Time

Reversal Test. Table 5 presents an idealized example showing changes in units (quantities) and prices of different components including two respective share classes addition over two periods.

Component	*Price 0*	*Quantity 0*	*Price 1*	*Quantity 1*
A	100	3	110	4
B	110	5	120	4
Google (A, C)	130	2	140	4

Table 5 – Idealized example of increase in units with GOOG and GOGL

Tables (6A) and (6B) present valuations of the Index in different periods.

p_1q_0	p_0q_0
330	300
600	550
280	260

p_0q_1	p_1q_1
400	440
440	480
520	560

Table 6 A, 6 B – Different Period Valuations

Equation (16) calculates the Laspeyres Index value based to price and quantities in year 0.

$$\mathrm{P}_{0,1} = \frac{330+600+280}{300+550+260} = 1.09 \qquad - 16$$

Equation (17) modifies the approach by reversing the calculation, using prices and quantities from Year 1 as the base. This reversal, combined with the original value as demonstrated in equation xviii, serves as an alternative method to test the Laspeyres Index against the Time Reversal Test. In this scenario, the result as calculated in Equation (18) deviates from unity, suggesting that increasing quantities in the given example with GOOG and GOGL negatively impacts the Index calculation. This outcome highlights a bias within the Index, which is further exacerbated in this instance due to inadequate Governance in managing the Index.

$$P_{1,0} = \frac{400+440+520}{440+480+560} = 0.918 \qquad -17$$

$$P_{0,1} \times P_{1,0} = 1.09 \times 0.918 = 1.00062 \neq 1 \qquad - 18$$

Fisher's Index, though valuable to showcase the failure and bias of the Laspeyres' method and hence as a consequence the MCAP method, is not well suited to build equity Indexes. To build a better Index, more understanding is needed regarding factor investing, non-linear mechanisms, failure of Smart Beta Indexes and then move to explore the statistical nature of MCAP and other factors.

-Factor Investing

A winner's bias is good for the popularity contest because it has an inflated value, but it's not good for investors owing to its skewed concentration. Hence, it's imperative to build a better benchmark, which can represent the market better and hence nurture the next generation of index funds that increase returns and reduce risk. The only way to do that is to first statistically understand MCAP bias, know how it belongs to a bias complex and use that knowledge to build a bias agnostic new method.

Factors emerged from Investing styles. Value was the first investing style, which was followed by other investing styles. Though the styles showed persistence, they still failed to prevail over the MCAP bias [14]. As statistical methods evolved, the research moved to linear regression for factor explainability [15]. The post Ball and Brown (1968) (BB68) [16] period changed the perception that there is no predictability in accounting data owing to information drift and recently commemorated its 50th year anniversary. Despite the information drift, factors have not been able to establish themselves. They are hard to time. More factors are added to the factor zoo everyday [17]. Many times, factor affiliations have had disaster consequences as factors have shown persistence and then dramatically failed to work [18]. Rather naïve factor Investing has shown better outcomes [19] and the more the number of factors, easier it is to find statistical significance in a few of them [20].

Factors are like any other biases. They persist and fail. Even Eugene Fama talked about periods of unexplainability and that factors could be a proxy for something else [21]. Scoring, ranking or sorting a set of factors is the easier part, the harder part is to explain how they work, why they work and fail and where does the complexity come from. Moreover, most factors are a function of price, especially fundamental factors driven by fundamental ratios. The failure of factor investing leads us back to Granger's challenge famously articulated by Clive Granger, when he said that the only way to bury EMH was to have an open methodology to deliver excess returns for a long period of time [22].

-Non-Linear Mechanisms

In 1992, Clive Granger, a Nobel Prize-winning economist, issued a challenge to the economics community to demonstrate that financial market prices contain information not already reflected in past prices and other available economic data [23]. The challenge aimed to test the Efficient Market Hypothesis (EMH) [24], which posits that financial market prices fully incorporate all available information, making it impossible to consistently outperform the market average with any trading strategy. Simultaneously, Granger highlighted the potential of non-linear models that alternate between forecastable and non-forecastable regimes, hinting at a possible solution to the stock market forecasting conundrum.

Granger emphasized the possibility of disproving the Efficient Market Hypothesis if a method that generates consistent profits, even after accounting for risk and transaction costs, remains effective over a prolonged period, especially when widely known. He suggested that advantages might be gained from longer-term perspectives, disaggregated data, outlier removal, and, notably, the use of non-linear models.

He noted the introduction of various models claiming to forecast stock market price changes, particularly nonlinear models characterized by regime switching based on factors like volatility levels, earnings/price ratios, company size, and calendar effects.

In his later work, Granger shifted focus to ‘stochastic trends’, the smooth, low-frequency components of processes generated by stochastic mechanisms [25]. He recognized that many models contain numerous parameters or economically unintelligible shapes, viewing them as approximations to more complex true generating mechanisms. He critiqued deterministic models and conventional econometric thinking for their limited utility in this context.

Despite his exploration of causality through Granger causality (GC) [26], which measures the causal effect by assessing prediction error reduction rather than identifying underlying stochastic mechanisms, Granger's work didn't fully embrace the concept of factors as a 'zoo', nor did he approach forecasting from an indexing bias perspective.

His work, while acknowledging that model parameters and shapes lack economic interpretation and are likely doomed to fail, did not delve deeply into understanding the inherent bias in Indexing.

- Underperforming Smart Beta to outperforming Machine Beta

Smart Beta strategies, a hybrid of active and passive investment approaches, have grown in popularity due to their promise of higher risk-adjusted returns at a lower cost compared to traditional active management [27]. These strategies typically focus on factors such as value, momentum, size, and volatility, and more recently, there's been a trend towards multi-factor and customized strategies, including ESG integration.

The use of advanced analytics and machine learning has become integral in identifying and optimizing these factors. Additionally, there's a notable increase in adoption by institutional investors and an expansion into global markets, including emerging economies and the fixed income sector.

Despite these advancements, Smart Beta strategies have faced challenges in consistently outperforming traditional market-cap-weighted indices [28]. Several factors contribute to this underperformance.

First, the potential for overcrowding in popular trades can dilute returns, as many investors chase the same factors. Secondly, the risk of data mining in identifying factors can lead to strategies based on historical anomalies that may not persist in the future.

Another critical challenge is the dynamic nature of markets; factors that have historically shown outperformance may not necessarily do so under different market conditions. Furthermore, the increased costs associated with more complex Smart Beta strategies, such as multi-factor or ESG-integrated funds, can erode the net returns for investors.

Lastly, there's the fundamental challenge of timing and market cycles – Smart Beta strategies, like any investment approach, can go through periods of underperformance depending on the prevailing economic environment and market cycles.

Machine Beta emerges as a potential solution to the challenges faced by Smart Beta strategies, offering a theoretically more sophisticated and nuanced approach. To effectively address the inherent biases and risks associated with market capitalization-weighted benchmarks, Machine Beta would need to employ Statistical factors and Non-Linear Mechanisms.

The adoption of advanced analytics and machine learning is pivotal for Machine Beta, with the explicit objective of maintaining low tracking errors while striving to outperform market cap-biased benchmarks. The success of this innovative approach would depend on its applicability and effectiveness across various asset classes and regions.

To establish its credibility and robustness, Machine Beta must navigate the Granger causality challenge, thereby potentially marking a significant advancement in index construction.

The aim is to shift towards more equitable and performance-oriented benchmarks, diverging from the traditional skewed concentration of market cap-weighted indices.

II – Literature Review

-Mean Reversion as a mechanism

In 1886, Francis Galton introduced the concept of mean reversion, focusing on its relative aspects over absolute ones [29]. Despite the concept being publicly known for over 130 years, there has been little progress in evolving Galton's original definition for natural systems into a comprehensive model that enhances our understanding and management of these systems, or in explaining the shortcomings of mean reversion. Galton failed to explore how variables interact with their sample averages, highlighting the equilibrium between converging and diverging forces, and demonstrated how mean reversion manifests and fails across various domains. An ideal representation of Galtonian reversion should be straightforward, relative, and universally applicable. A 2014 paper proposed a new framework, using the stock market as an example, that expands upon Galton's notion of a natural system [30].

Figure 1 illustrates the percentage of a group that moves from one boundary (100 and 0) inwards wards to the other boundary of a relative ranking system. While at the same time the middle level (50) sees a divergence out to both sides, outwards. This mechanism of convergence and divergence happened at the same time suggesting that the concept of mean reversion did not work alone and was accompanied by its opposite force, where things did not revert, but diverge.

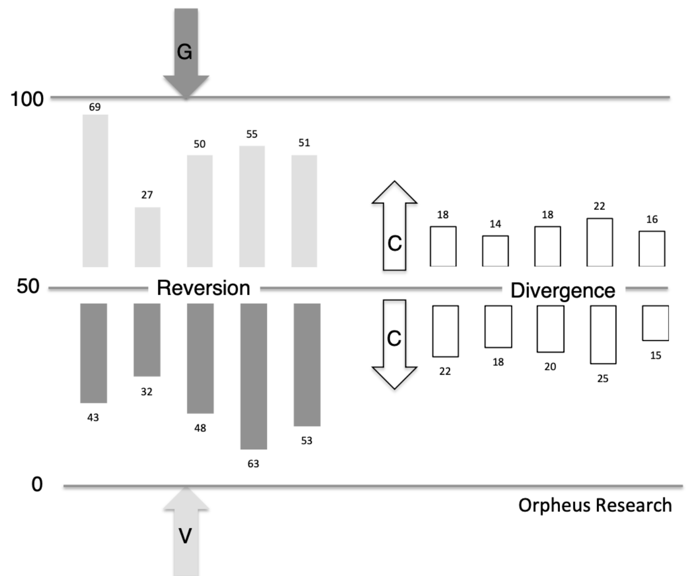


Figure 1 - Mean reversion framework, paper illustrated a mechanism functioning

-Stochastic And Forecastable Trends As States

Mean reversion fails and works. This dual states of predictability of events are like stochastic states. A similar perspective can also be applied to entropy, representing states of disorder (D) or order (O) and how one transforms (powers) the other.

The paper titled "The [3N] Model of Life" [31] elaborates on Erwin Schrödinger's seminal work, "What is Life?" [32] In this influential book, the pioneer of quantum mechanics delved into the workings of life, particularly how nature consistently transforms disorder into order. The paper presents the [3N] model as a framework for understanding how nature navigates the challenges posed by the second law of thermodynamics, enabling the emergence of order from chaos.

Figure 2 is the [3N] Model to decode complexity and offer insights into intricate systems such as stock markets. It suggests various ways to perceive order and chaos, like distinguishing between signal and noise, explainable and unexplainable phenomena, or predictable and unpredictable outcomes. By adopting the concept of probabilistic states, this model opens avenues for exploring multiple facets within the same dataset.

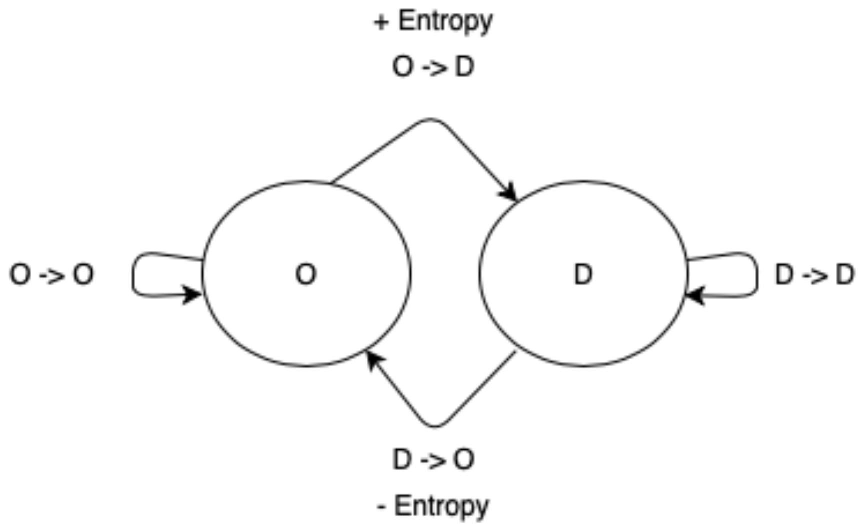


Figure 2 – Entropy Distropy State System

-Bias Generating Mechanism

The paper titled "The [3N] Method" [33] builds upon the concepts introduced in the "The [3N] Model of Life" by expanding the [3N] Mechanism. It seeks to harmonize the discrepancies between informational states. These states are defined by the nature of information, where the principles of BB68 coexist with the EMH, suggesting that information does not offer a long-term advantage.

The present paper reinterprets the [3N] mechanism, merging it with the MCAP weighting method, which is akin to the 'Rich-Get-Richer' (RGR) (Preferential Attachment) [34] model, behaving in a non-normal way owing to its biased methodology, leading to extremity and skewed concentration. This Non-Normal approach [NN] contrasts with Value Investing, or the 'Poor-Get-Richer,' Normal [N], mean reversion approach, a dual state of Mean Reversion and its Failure.

Figure 3 illustrates the [3N] Method, proposing stock markets as complex informational systems, recognizing MCAP as one type of statistical bias within a broader spectrum of biases. And idealized e.g. like RGR, RGP (Rich-Get-Poor), PGR (Poor-Get-Rich), PGP(Poor-Get-Poorer). It explores how this complex of biases not only elucidates the RGR mathematical concept but also paves the way for a novel methodology built as a composite of biased to outperform traditional more biased S&P 500 Index. This innovative approach offers a fresh perspective on building Indexes and challenging conventional market theories.

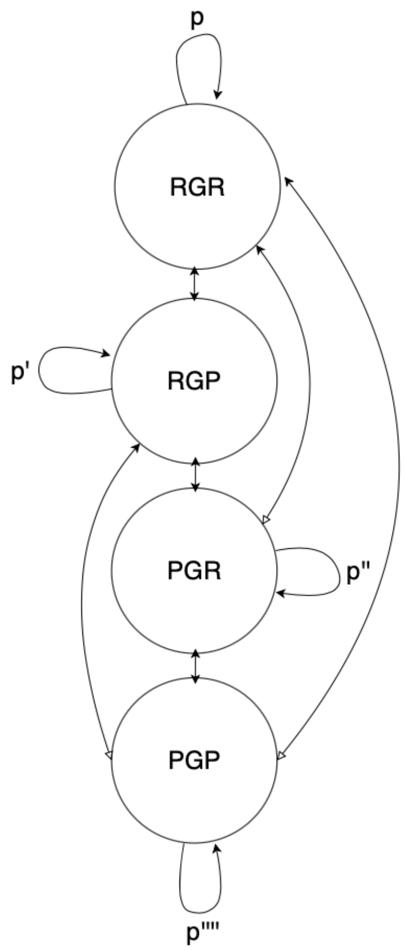


Figure 3 - RGR-RGP-PGR-PGP Markov Chain of Statistical Biases

A bias complex can be seen as a Markov chain, a perpetually vibrating mechanism, driven by a set of statistical persistent behaviors which are not fleeting patterns but robust in their expression. Without a mechanism it's hard to fathom how different statistical behaviors can co-exist. How can RGR coexist and emerge from an interaction of many other statistical behaviors?

How there is no preferential attachment without preferential detachment [35]? A problem where physicists have accused other physicists of indulging in rediscovering America [36]. At the heart of complexity was always a simple data generating mechanism. Nature and its ensuing complexity were never completely random or stochastic.

-The Rebalance Time

Any Investing process that beats the winner's bias in MCAP must be less obsessive in weight like the MCAP Index. The only way to be weight unobsessive is to have a systematic rebalance among different bias states. A true non-linear mechanism needs a temporal character for stable functioning.

One way a multi-state mechanism can become forecastable and stochastic at the same time is when it's a combination of many clocks, which synchronize and desynchronize in a predictably unpredictable manner. Physicists have long referred to this phenomenon as entrainment [37]. Disentrainment [38] or de-synchronization, therefore, refers to a loss of this synchronization, where systems or processes that were once operating in unison begin to operate independently of each other.

In the context of financial markets, entrainment might refer to the phenomenon where market participants' behaviors become synchronized, leading to herding behavior. The opposite would then be a market where participants act independently, breaking away from the collective trend. Hence achieving the impossibility of a dual character of order and disorder, forecastable and stochastic system, the very definition of deterministic disordered chaos. Stock market forecasting works through a chaotic system which modern finance sometime calls, efficiently inefficient, an essence of Granger's 1992 challenge.

A plausible reason this model thinking has not been discovered is because, statistical distributions were never seen as a temporal mechanism. An oscillating clock, that disturbs itself, creates energy from disorder, entrains itself and starts the process all over again, chaotically, moving between disorder and order.

Temporal connections are very important to understand complexity. In "The Beta Maths" paper [39], the author explained how mean reversion (RTM) and failure to revert to mean (FRTM) are connected in time. And how portfolios made with such a differentiating temporal structure can outperform a system that works singularly.

In other words, working in a bias agnostic system using time as a refreshing approach to reestablish biases is better than a singular bias system without any time. Such a system transforms forecasting from absolute approach into a probabilistic mechanistic function, that is capable to beat simple reference benchmarks. Equation xix indicates that a set (S) of portfolios built from mean reversion (MR) working in relatively longer Time (T) along with Failure of Mean Reversion in relatively shorter time (t) would always match or outperform a Universe (U) of portfolios built from any factor without a systematic temporal feature.

$$S\ [\ P\ \{\ (FMR)_t\ (MR)_T\ \}]\geq \mathrm{U}\ [\ \mathrm{P}\ (F)_{ij}] \qquad \text{– xix}$$

This differential behavior confirmed the working of the [3N] model and also the potential of a building a dynamic bias agnostic method to beat the MCAP methodology. Complexity is the expression of an intelligent system because it has to extract energy from noise. A few other papers have been published explaining stationarity in relative ranking data [40].

III – Statistical Factors

- Broken Clock

Redefining the problem is essential to reach an effective solution. A biased benchmark that inflates value is bound to attract popularity among market participants. The more the participants, the more it is likely to grow, a vicious cycle that exaggerates the concentration. While the value and popularity of a benchmark increases, it is naturally going to become a more prominent reference, where every investment alternative is going to be measured against it. And in a certain way, any alternative that competes with the reference, can be seen as oscillating performance around the reference. The alternatives may for a short time go above the reference, but if they ignore the bias associated with the reference, the bias of inflating value in a few winners, the alternative may falter against the reference and underperform. The singular focus of inflated value is going to keep the reference at the center of the investment universe. All this seems like a logical explanation for the benchmark, barring one question. What is the benchmark reverting too? Benchmarks are not immune to statistical laws and cannot be considered like the sun in the solar system.

There is something that a MCAP benchmark should revert too and that is a function of its extremity, the concentration and the inflated value. The same features that take it to the top of the pile also are the very drivers for its reversion, as a top winner's multiyear underperformance can consequently drag the benchmark along with it, for multiple years, owing to its self-reinforcing inertia. A heavy ball as it gets heavier is harder to lift at increasing speed against gravity. The same heavy ball when it falls under water, is harder to bring out owing to the same inertia. Normality, reversion is a reality for the MCAP benchmark too, even if it is at the center of the investment universe. In some sense, benchmark (references) can operate like a broken clock, that shows time correct twice a day, a false sense of functioning.

Modern finance teaches the world how to explain risk and returns of global investments against MCAP benchmarks. In its true sense the comparison between two performances is the comparison between their true biases. The invincibility of the MCAP benchmarks is owing to its positive bias which prevails occasionally above the other intrinsic biases of most investment strategies. When you give something the status of a reference, the opportunity for an investment strategy to track the reference benchmark and leave it surely and steadily behind (outperformance) is a miracle of modern finance because it involves an understanding of the reference and its risks. This creates a monumental hurdle for any market innovation that is designed to disprove the thesis and demystify the myth.

- Market as a composite of statistical factors

An alternative to beating winner's bias cannot be another kind of bias, like a loser's bias. Because just like factors, even biases can come in and out of forecastable states. To build a bias agnostic process that subsumes the MCAP RGR bias requires an understanding of bias generation and bias complex. Where do biases come from? Where do they go?

These questions pertain to the subject of mechanisms. At the heart of a true mechanism are probabilities that define the forecastability and noise inherent in a set of biases. This mechanism is complex and defines complexity by combining stochastic and deterministic trends. Solving the stock market forecasting problem hence cannot be done without articulating complexity as multi state system that exhibits both deterministic and stochastic trends.

Figure 4 illustrates the statistical factors that are forecastable and non-forecastable at the same time, that are persisting and failing to persist at the same time, that have signal and noise at the same time, they are prominent but never singular, they are always expressing in probabilistic degrees. MCAP hence becomes one of the numerous expressions of the system. These statistical factors bring explainability to conventional factors, why they work and fail in an endless cycle.

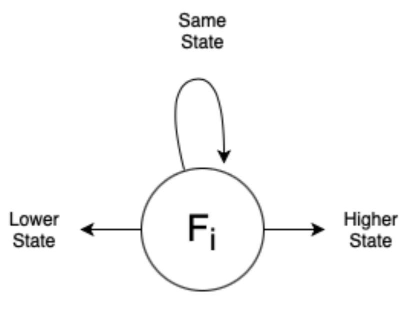


Figure 4 - A conventional Factor lives in many statistical states at the same time

Figure 5 adds some complexity to the Statistical Factors by combining two different factors. A lot more complexity can be added to the Statistical Factors when more variables like weights, rebalance time, asset class, ranking period, basket size etc. are added to the mix. The complexity can be increased into numerous dimensions and degrees of freedom. The more the constraints, the more the complexity. Consequently, we are looking at a whole complex of biases which can only be understood by machines and not humans. If Factors are seen in their true form, in their Long-Short nature, the number of Statistical Factors can increase to millions.

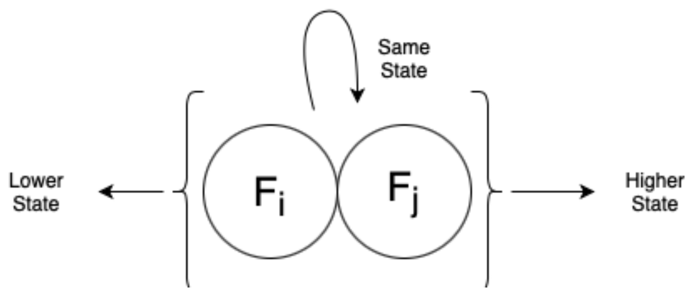


Figure 5 - A multi conventional Factor also lives in many statistical states at the same time

IV – Results

The Exceptional & Rich (E&R) Machine Beta Model was developed with the goal of constructing an improved version of the S&P 500 using the same set of 500 stocks. This model aimed to achieve low turnover, minimal tracking error, and higher risk-weighted annualized returns, following a systematic, scientific, and replicable methodology. The first step involved collecting data, specifically Yahoo End of Day (EOD) closing price data. Next, the model ranked the stocks based on their relative performance across nine different rolling quarterly periods, ranging from 1 to 20 quarters.

A key differentiator of the E&R Model was its approach to stock selection and weighting. Unlike the S&P 500, which relies on MCAP weighting, the E&R Model adopted a more diversified strategy. It utilized quintiles instead of deciles as boundary conditions for statistical styles. This strategy classified stocks into three categories: Statistical Value (V) for stocks in the bottom quintile, Statistical Growth (G) for those in the top quintile, and Statistical Core (C) for stocks in the middle quintiles, excluding the extremes.

The model then moved on to creating various portfolios with different compositions and starting points. These included individual portfolios based on the Value (V), Core (C), and Growth (G) categories, a combined Value-Growth (VG) portfolio consisting of stocks from the top and bottom quintiles, and a comprehensive Value-Core-Growth (VCG) portfolio that included stocks from all quintiles. In the VCG portfolio, Value (V) and Growth (G) components were each allocated a 40% weight, while the Core (C) component received a 20% weight. Beginning in 2016, a total of 5000 unique portfolios were generated. The dynamic nature of stock rankings meant that each new selection of a starting point resulted in a different composition of the portfolio, ensuring the uniqueness of each construction.

Figure 6 bar charts illustrate the annualized excess returns of five investment strategies, labeled C, G, V, VCG, and VG vs. S&P500. The G strategy significantly outperforms the others with the highest return, followed by VG and VCG with average positive annualized excess returns above 400 basis points above the S&P 500. The V strategy shows a small positive return, while the C strategy barely registers above zero, indicating minimal excess return over the benchmark.

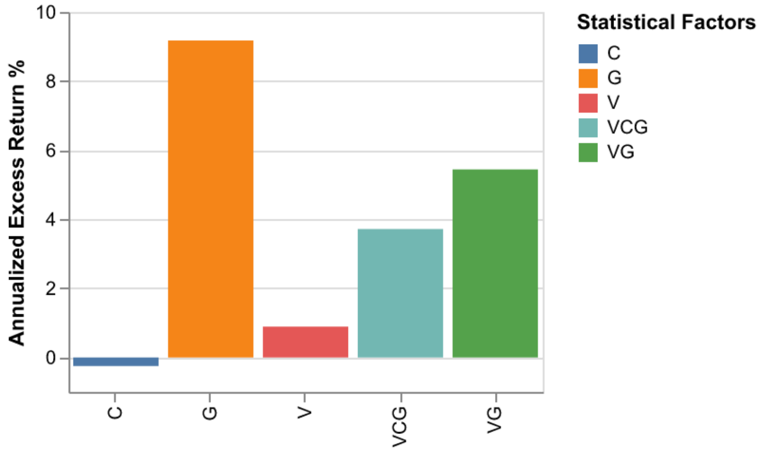


Figure 6 - Average Annualized Excess Returns of since 2016

Figure 7 scatter plot depicts the relationship between annualized excess returns and annualized excess volatility for the VCG strategy across different portfolios and holding periods. Most portfolios cluster around a region of low to moderate excess volatility and positive excess returns, indicating a consistent performance. The plot is divided into quadrants indicating the percentage of portfolios that fall into high and low volatility and return categories, with the majority (67.92%) being in the quadrant of low volatility and high returns. 71.81% of data was clustered in the positive quadrants (Q1+Q2).

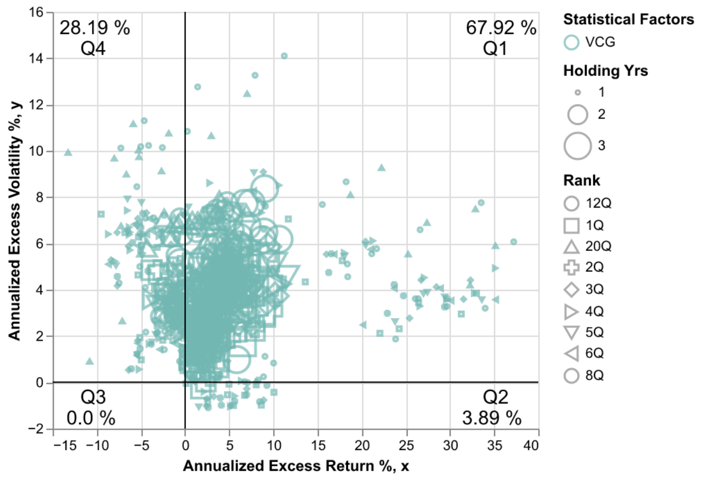


Figure 7 - Annualized Excess Returns vs. Annualized Excess Returns

Figure 8 scatter plot shows the distribution of the VCG strategy portfolios plotted against annualized excess returns and the information ratio. A large proportion of the portfolios, about 46.81%, are concentrated in the top right quadrant, indicating a high information ratio and high excess returns, which is an ideal outcome for performance.

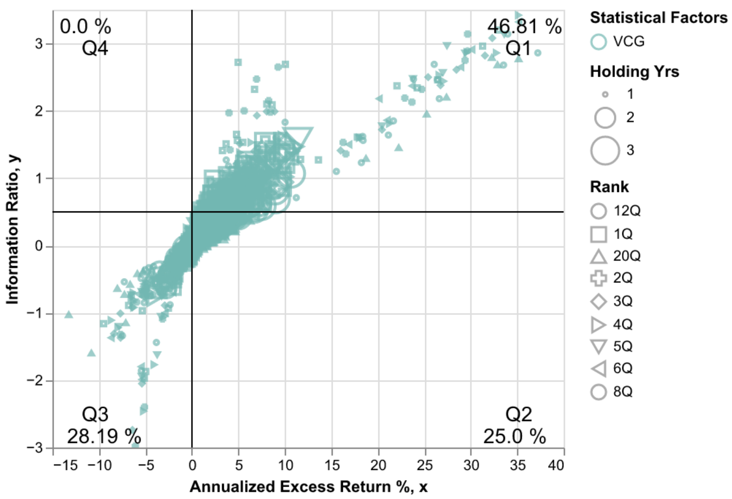


Figure 8 - Annualized Excess Returns vs. Information Ratio

Figure 9 scatter plot maps the VCG strategy portfolios based on their tracking error versus information ratio. The majority of the portfolios, representing 47.08%, are clustered around the lower tracking error and lower information ratio, indicating a conservative approach with returns close to the benchmark. The Cartesian plot has a smaller portion, 9.58%, lies in the desirable upper right quadrant with high information ratios despite higher tracking errors, suggesting some portfolios are achieving superior risk-adjusted performance relative to their deviation from the benchmark. 83.81% of data was clustered in less than 10% Tracking Error Quadrants (Q3+Q4).

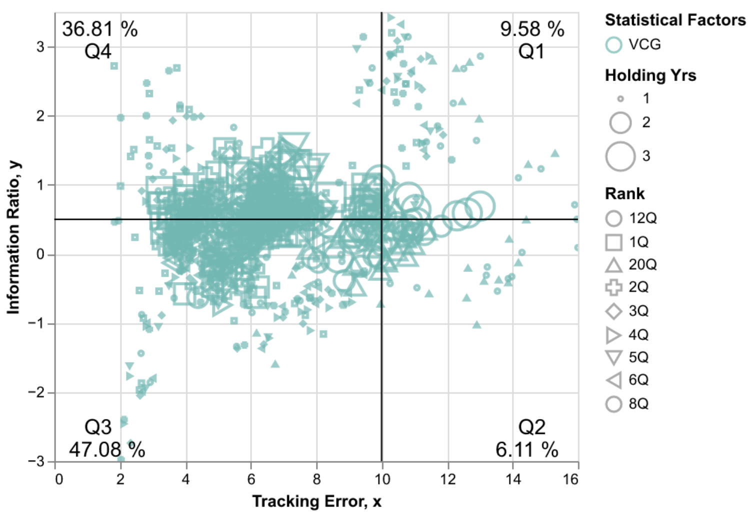


Figure 9 - Annualized Excess Returns vs. Information Ratio.

Table 7 is a table with the Modern Portfolio Theory statistics for various investment strategies over one to three-year periods, revealing their performance, risk profiles, and deviations from a benchmark in terms of annualized returns, volatility, and drawdowns. It serves as a quantitative analysis tool to compare the risk-adjusted returns and downside risks of these strategies against the market benchmark S&P500.

The "C" style exhibits relatively low annualized excess volatility, suggesting that its risk profile doesn't deviate significantly from the benchmark. Despite this, the notable negative Max Excess Drawdown, especially over a two-year holding period, implies that the strategy experienced considerably larger drops from its peak compared to the benchmark, signaling substantial downside risk during that period.

Conversely, the "G" style stands out with strong positive annualized returns and comparatively modest Max Excess Drawdowns, particularly noticeable within the one-year frame. The higher annualized excess volatility indicates that the strategy is riskier, yet this seems to have been rewarded with higher performance.

The "V" style is characterized by high excess volatility and the most significant Max Excess Drawdowns across all periods reviewed. This data suggests that the strategy is assuming a much higher risk, with the portfolio experiencing substantially larger dips than the benchmark, which may concern risk-averse investors.

Meanwhile, the "VCG" style presents a more balanced profile, showing moderate performance and better control over Max Excess Drawdowns than the "V" style. This indicates a strategy that, despite embracing additional volatility, has demonstrated resilience during market downturns.

Lastly, the "VG" style displays varying performance across different holding periods and high excess volatility. The less severe but still substantial Max Excess Drawdowns, particularly in the one-year period, suggest that while the strategy may take on additional risk, it has a slightly better capacity to manage large market swings compared to the "V" style.

Integrating the insights from the MPT table with the graphical analysis further refines our understanding of the VCG Machine Beta strategy's relative performance against the S&P 500. The VCG strategy showcases positive annualized returns (AR) across the 1Y, 2Y, and 3Y holding periods. The Alpha is consistently positive for all holding years, indicating that the strategy has generated returns above what would be expected given its market risk (Beta). The positive Alpha supports the bar graph's (Figure 6) indication of the strategy's ability to outperform the benchmark.

The information ratio (IR), which balances excess returns with the tracking error (TE), is positive for VCG across all periods. This ratio suggests that the strategy has efficiently translated its tracking error into excess performance relative to the benchmark.

The annualized excess volatility (AV) is moderate, signifying that the VCG portfolios are not taking on excessive risk compared to the S&P 500. This is corroborated by the scatter plot showing many portfolios with higher returns not being accompanied by high excess volatility.

The Max Excess Drawdown (Max Excess Dd) for VCG remains relatively lower than that of other strategies, especially over a 3Y period. This indicates the strategy's resilience and potential for recovery during market downturns, which is an attractive characteristic for long-term investments.

Combining these metrics with the graphical data, the VCG Machine Beta strategy emerges as a well-rounded approach. It demonstrates the capacity to outperform the S&P 500 with controlled risk and resilience during downturns. The consistent positive Alpha and moderate drawdowns suggest that the strategy may offer a favorable balance between risk and return, aligning with the objectives of Machine Beta investing.

Statistical Factors	Holding Yrs	AR	AV	TE	IR	Alpha	Beta	Max Excess Dd
C	1Y	-0.38	0.58	3.96	-0.03	0.02	0.99	-0.69
C	2Y	-0.48	0.83	3.66	-0.01	-0.06	1.02	-1.93
C	3Y	0.1	0.75	3.72	0.04	-0.01	1.02	-3.1
G	1Y	8.46	1.75	6.39	1.44	0.61	1.04	-1.91
G	2Y	8.79	2.96	7.07	1.27	0.57	1.12	-2.67
G	3Y	9.45	4.49	7.83	1.17	0.52	1.2	-3.82
V	1Y	4.72	8.92	13.11	0.02	-0.18	1.29	-5.73
V	2Y	-0.91	5.94	10.22	-0.15	-0.32	1.2	-5.9
V	3Y	0.3	4.63	9.83	-0.1	-0.09	1.13	-6.24
VCG	1Y	4.42	4.04	7.02	0.43	0.13	1.14	-2.1
VCG	2Y	2.37	3.56	6.03	0.38	0.01	1.15	-3.8
VCG	3Y	4.15	3.78	6.91	0.6	0.16	1.15	-4.35
VG	1Y	6.7	3.67	6.35	0.87	0.27	1.15	-2.4
VG	2Y	3.91	3.3	5.72	0.71	0.14	1.15	-3.76
VG	3Y	5.7	3.9	6.87	0.8	0.27	1.17	-4.75

Table 7 – Modern Portfolio Theory statistics for various investment strategies

V – Discussion

The E&R Machine Beta Model signifies a notable advancement in index construction methodologies, particularly in overcoming the limitations inherent in traditional approaches like the S&P 500's weight obsessiveness and Laspeyresian bias. This innovative model, while utilizing the same set of 500 stocks as the S&P 500, diverges significantly in its approach to stock selection, weighting, and rebalancing.

A key aspect of the E&R Model is its diversified strategy in stock selection and weighting. Instead of the MCAP weighting used by the S&P 500, the E&R Model employs quintiles to categorize stocks into three distinct groups: Statistical Value (V), Statistical Growth (G), and Statistical Core (C). This method ensures a more balanced representation and moves away from the fluctuating weights typical of MCAP indices, which can be overly sensitive to short-term price movements.

Moreover, the model's idealized allocation adopts a stable weight allocation, assigning fixed weights of 40% each to Value and Growth components, and 20% to the Core component. This fixed weighting scheme contrasts sharply with traditional indices, where weights change with every price fluctuation, thus maintaining a consistent portfolio composition over time.

Additionally, the E&R Model stands out for its longer rebalancing period, extending well beyond the usual 12 months, which can be further increased with an innovative portfolio construction. This slower rebalancing approach is crucial for reducing the Laspeyresian bias, which often results from rapid, price-driven weight changes in traditional indices. By focusing on a more gradual adjustment process, the model avoids the pitfalls of frequent rebalancing based on short-term market trends.

The dynamic nature of the E&R Model also allows for versatile portfolio construction. The simulations of the idealized methodology created a variety of portfolios with unique compositions and starting points, including individual portfolios for Value, Core, Growth, as well as combined VCG strategies which are comparable to the S&P500. The dynamic barbell approach (overweight V, G and underweight C) further minimizes the risk of bias and overconcentration in specific sectors or stocks.

The authors are confident that extending the current period of study against S&P 500 till 1965 would strengthen the case for Machine Beta based on statistical factors. However, the surprising large annualized excess returns, near 400 basis points, even for the near 7-year period under study is a significant number for an open methodology suggesting that Statistical Factors are an understudied subject which when regressed against conventional Fama French Factors should give further confirming insights into the research. The capability of idealized method should eventually be further enhanced by machine learning into building next generation active investing strategies.

VI - Conclusion

In conclusion, the traditional investment landscape, dominated by Market Capitalization (MCAP) weighted indices and influenced by the legacy of the Laspeyres index, is facing significant challenges due to inherent biases and limitations. The Laspeyres index, with its base year price bias and tendency to overstate value, has inadvertently passed these flaws into MCAP indices, leading to skewed market representations. This legacy issue is further highlighted by Fisher's research demonstrating the Laspeyres index's failure in the Time Reversal Test and the amplification of the bias because of inclusion of GOOG and GOGL into the S&P 500 list.

The limitations of Smart Beta strategies in addressing these biases underline the necessity for a new approach, which is where the concept of Machine Beta becomes critical. Machine Beta represents a paradigm shift, integrating advanced analytics, [3N] non-linear mechanisms, and statistical factors that can forecast and, interestingly, not forecast at the same time. This approach provides a more nuanced and dynamic understanding of market dynamics, moving beyond traditional predictive models.

The Exceptional & Rich (E&R) models are a practical application of the Machine Beta concept. They diverge from the traditional MCAP weighting approach, adopting a more diversified and balanced method for stock selection and weighting. This [3N] methodology effectively counters the concentration risk and overemphasis of MCAP indices. The E&R models, with their fixed weight allocations and extended rebalancing periods, is a new way to rebuild S&P 500 and outperform.

Bibliography

1. Malkiel, Burton G. "Asset Management Fees and the Growth of Finance." Journal of Economic Perspectives, vol. 27, no. 2, pp. 97-108, 2013.
2. Diewert, W.E. "Laspeyres, Ernst Louis Etienne (1834–1913)." In: The New Palgrave Dictionary of Economics. Palgrave Macmillan, London, DOI: 10.1057/978-1-349-95121-5_1006-2, 2008.
3. Dziechciarz, Jozef Zbigniew. "Review of Peter Von Der Lippe: Index Theory and Price Statistics." Argumenta Oeconomica, no. 1(20), pp. 157-160, 2008.
4. MSCI. "How Are Indexes Constructed." MSCI, 2023.
5. Weinhagen, Jonathan C. "Measuring the Substitution Effect in Producer Price Index Goods Data: 2002–16." Monthly Labor Review, U.S. Bureau of Labor Statistics, July, 2020.
6. Bloem, Adriaan M., Robert J. Dippelsman, and Nils O. Maehle. Quarterly National Accounts Manual—Concepts, Data Sources, and Compilation. International Monetary Fund, May 10, 2001.
7. Osband, Kent. "Index Number Biases During Price Liberalization." International Monetary Fund, January 1, 1992.
8. Armknecht, Paul, and Mick Silver. "Post-Laspeyres: The Case for a New Formula for Compiling Consumer Price Indexes," April, 2012.
9. "Methodology Matters." SPDJ Indices.
10. Von der Lippe, P. "Volume 16 of the Publication Series Spectrum of Federal Statistics," 2001.
11. Weniger, Jeff. "Market Capitalization-Weighted Indexes: An Accident of History." CFA®, Asset Allocation Strategist.
12. "Treatment of Google Stock Split in the S&P 100 and S&P 500," S&P Dow Jones Indices, February 3, 2014.
13. Fisher, I. The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability. Boston: Houghton Mifflin, 1922.
14. Busse, Jeffrey A., Amit Goyal, and Sunil Wahal. "Performance and Persistence in Institutional Investment Management." The Journal of Finance, vol. 65, no. 2, April, pp. 765-790, 2010.
15. Israel, Ronen, and Ross, Adrienne. "Measuring Factor Exposures: Uses and Abuses." Journal of Alternative Investments, Forthcoming.
16. Ball, R., & Brown, P. "An Empirical Evaluation of Accounting Income Numbers." Journal of Accounting Research, vol. 6, no. 2, pp. 159-178, 1968.
17. Harvey, C. R., & Liu, Y. "A Census of the Factor Zoo." Journal of Finance, 74(4), pp. 1637–1670, 2019.
18. López de Prado, M. M. Causal Factor Investing. Cambridge University Press, 2023.
19. Dichtl, Hubert, Drobetz, Wolfgang, and Wendt, Viktoria-Sophie. "Factor-Based Allocation: Is There a Superior Strategy?" March 25, 2019.
20. Chambers, Kazemi, Black. Alternative Investments: An Allocator's Approach. Wiley, 2021.
21. Fama, E. F., & French, K. R. "Size and Book-to-Market Factors in Earnings and Returns." Journal of Finance, 50(1), pp. 131-155, 1995.
22. Granger, C. W. J. "Forecasting stock market prices: Lessons for forecasters." International Journal of Forecasting, 8(1), pp. 3-13, 1992.
23. Liu, T., Granger, C. W. J., & Heller, W. P. "Using the Correlation Exponent to Decide whether an Economic Series is Chaotic." Journal of Applied Econometrics, 7(Suppl. 1), S25-S39, 1992.
24. Fama, E. F. "Efficient Capital Markets: A Review of Theory and Empirical Work." The Journal of Finance, 25(2), pp. 383-417, 1970.

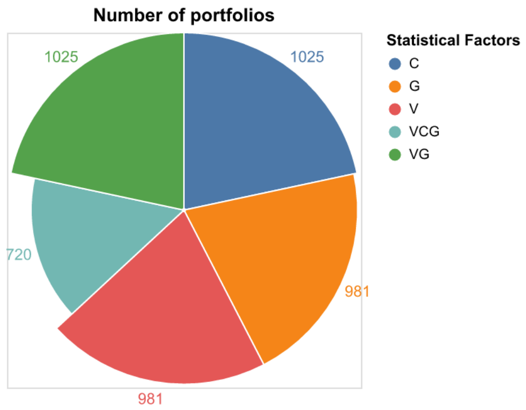
25. Konishi, T., Ramey, V. A., & Granger, C. W. J. "Stochastic Trends and Short-Run Relationships Between Financial Variables and Real Activity." NBER Working Paper No. 4275. National Bureau of Economic Research, 1993.
26. Granger, C. W. J. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." Econometrica, 37(3), pp. 424-438, 1969.
27. Cao, J., Hsu, J. C., Song, L., Xiao, Z., & Zhan, X. "Smart Beta, 'Smarter' Flows." Journal of Financial and Quantitative Analysis, 2023.
28. Huang, S., Song, Y., & Xiang, H. "The Smart Beta Mirage." Journal of Financial and Quantitative Analysis, 2020.
29. Galton, F. "Regression Towards Mediocrity in Hereditary Stature." Journal of the Royal Statistical Society, 48(3), pp. 263-292, 1886.
30. Pal, M. "Mean Reversion Framework." SSRN, 2015.
31. Pal, M. "The [3N] Model of Life." SSRN, 2021.
32. Schrödinger, E. "What Is Life? The Physical Aspect of the Living Cell." Cambridge: Cambridge University Press, 1944.
33. Pal, M. "The [3N] Method." SSRN, 2023.
34. Price, D. J. de Solla. "Networks of Scientific Papers." Science, 149(3683), pp. 510-515, 1965.
35. Pal, Mukul, "How Physics Solved Your Wealth Problem!" SSRN, October 8, 2016.
36. Simkin, M., & Roychowdhury, V. "Re-inventing Willis." Physics Reports, 2006.
37. Huygens, C. "Horologium Oscillatorium." Paris: Apud F. Muguet, 1673.
38. Czeisler, C. A., & Chiasera, E. J. "Disentrainment of the Human Circadian Pacemaker Due to Scheduled Exposures to Light." Journal of Neuroscience, 11(5), pp. 1245-1255, 1991.
39. Pal, M. "The Beta Maths." SSRN, 2016.
40. Pal, M., & Ferent-Pipas, M. “Stock Market Stationarity”. SSRN. 2015.

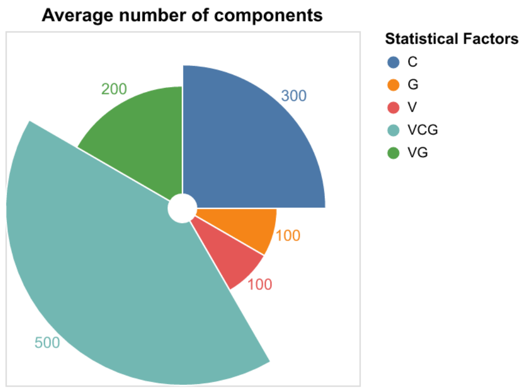
Annexure

-Input Data

Input data defines the Exceptional & Rich Model, benchmark, inception point, number of components and the number of simulations run.

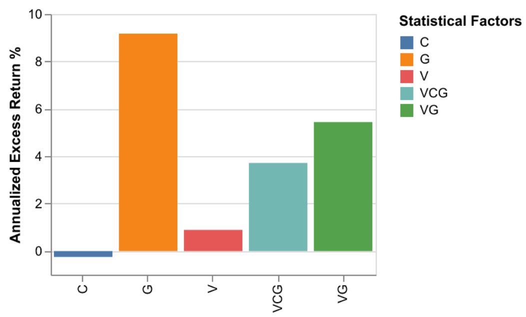
Group name	Benchmark	Total number of portfolios	Number of components	Portfolios starting year
Exceptional & Rich U.S. 500	S&P 500	4732	500	2016

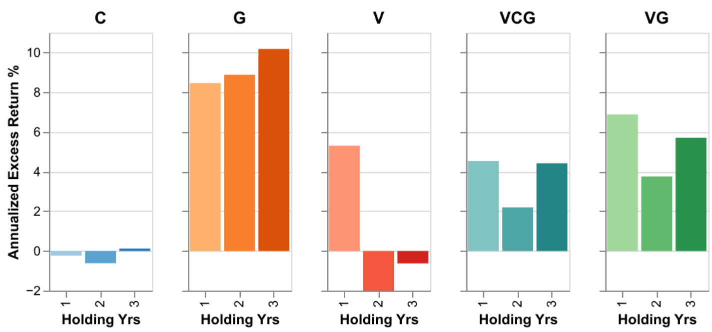


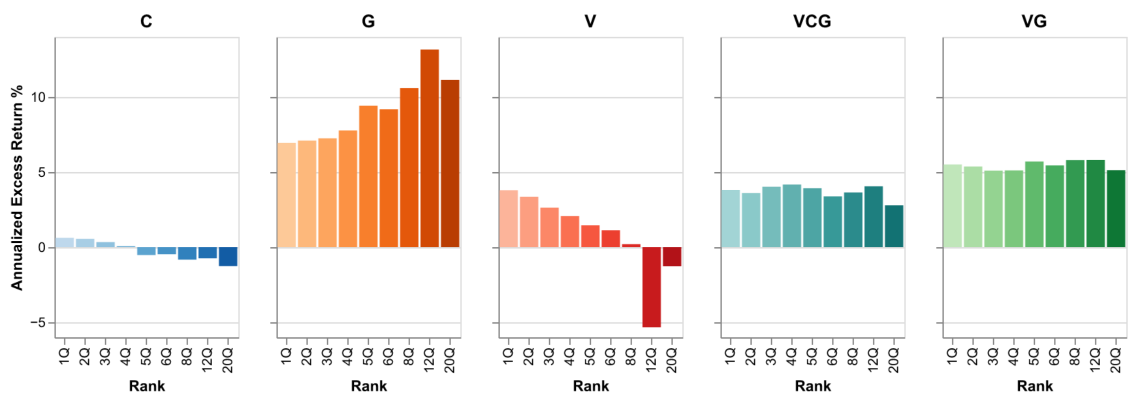


-Average annualized excess returns histograms for various factors

The histogram illustrates the average annualized excess returns for the various statistical factors, across the three different holding periods and for the respective quarterly proxy ranking periods.

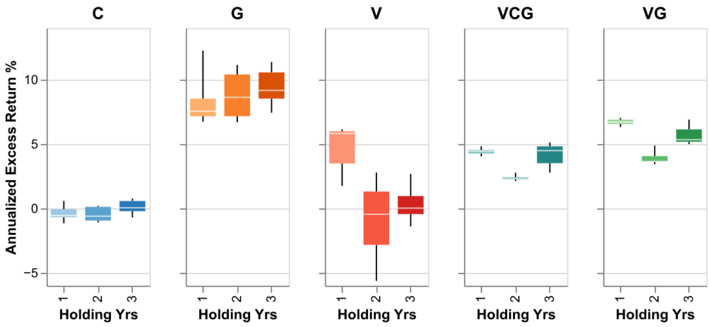


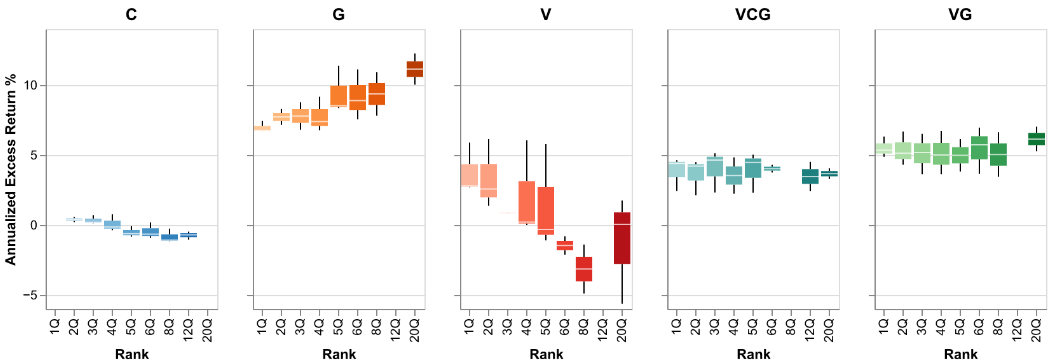




-Quartile boxplots for annualized excess returns

The boxplots show minimum, first quartile, median, third quartile, and maximum annualized Excess Return for various statistical factors, for different holding periods, and for the respective quarterly proxy rankings. The outliers were removed from the data.



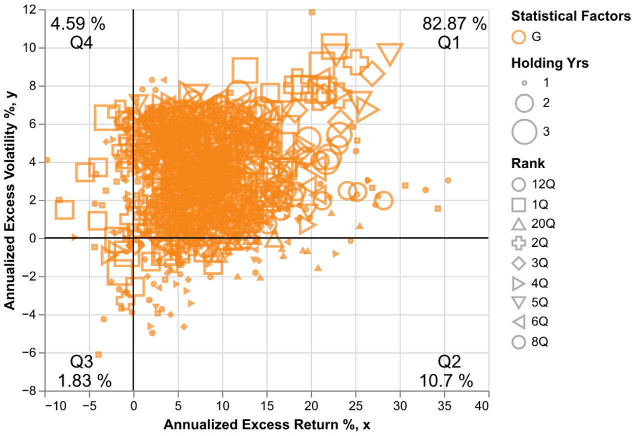


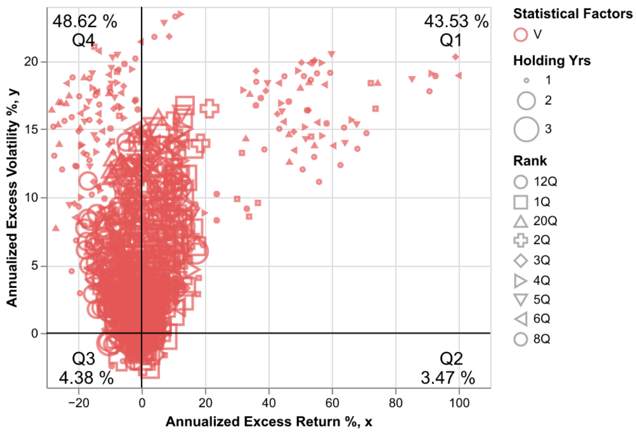
-Cartesian plots for various statistical measures

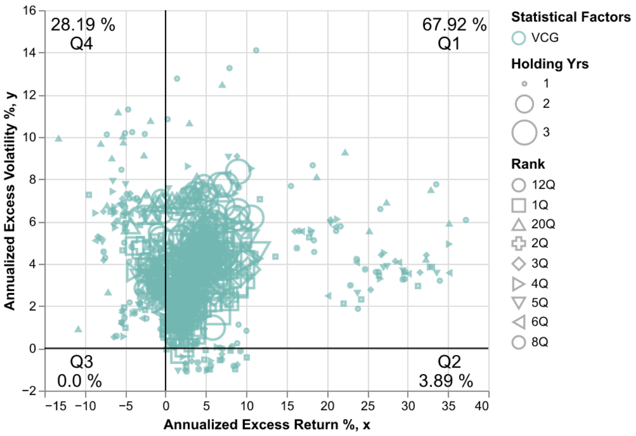
The cartesian plots below illustrate annualized excess returns vs. annualized excess volatility, information ratio vs. annualized excess returns, annualized excess volatility vs. information ratio, alpha vs. beta and tracking error vs information ratio plotted for different statistical factors, for different holding periods and for respective quarterly proxy rankings.

-Annualized excess returns vs. annualized excess volatility

The cartesian chart of annualized excess return vs. annualized excess volatility plotted for different statistical factors, for different holding periods and for respective quarterly proxy rankings.

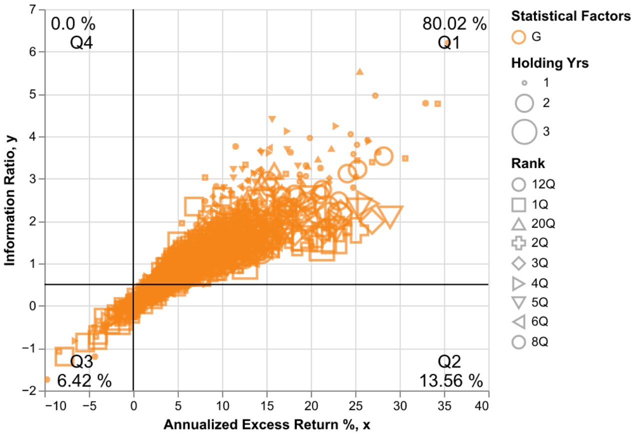


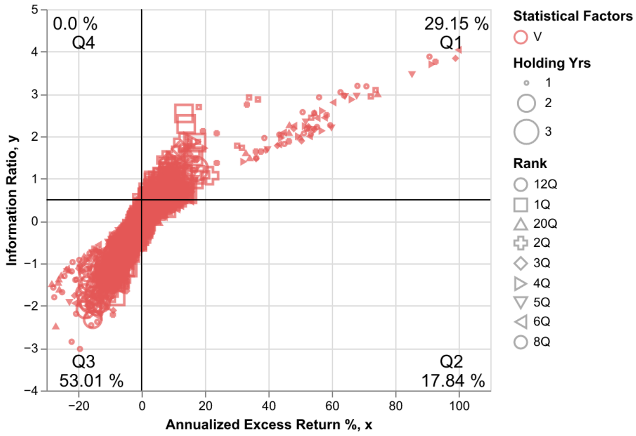


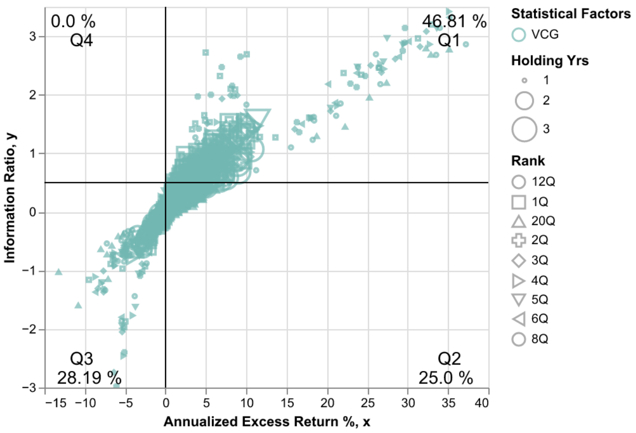


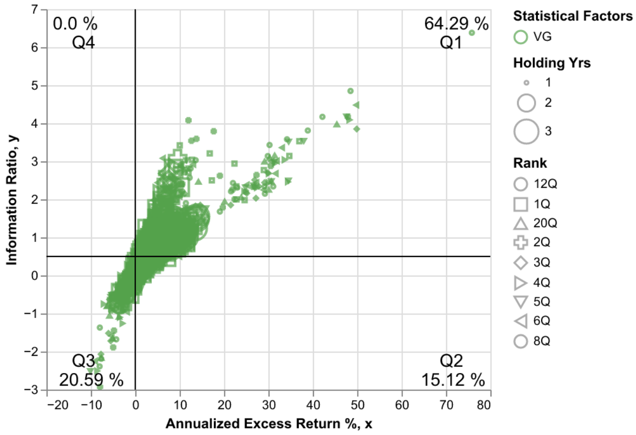
-Annualized excess returns vs. Information ratio

The cartesian chart of annualized excess return vs. information ratio plotted for different statistical factors, for different holding periods and for respective quarterly proxy rankings.



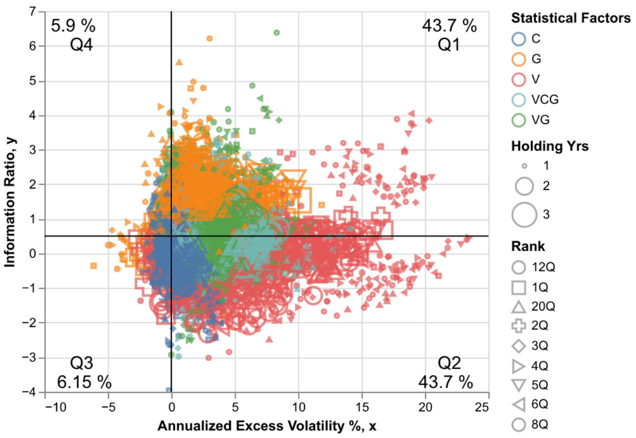


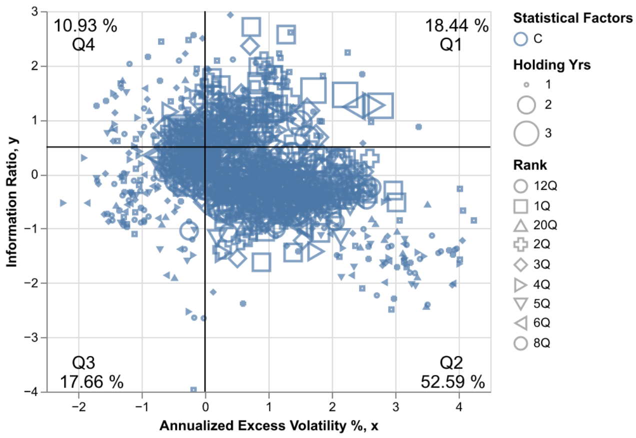


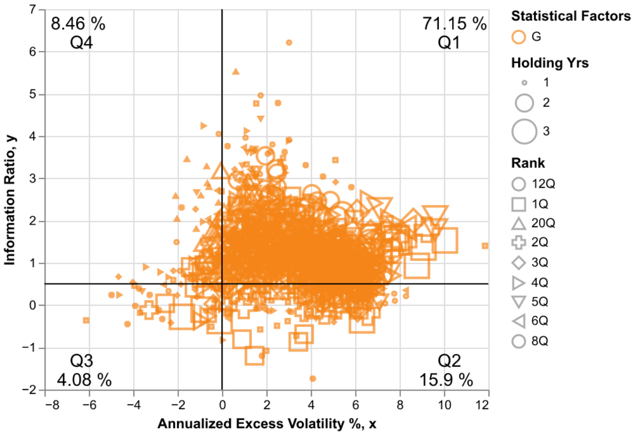


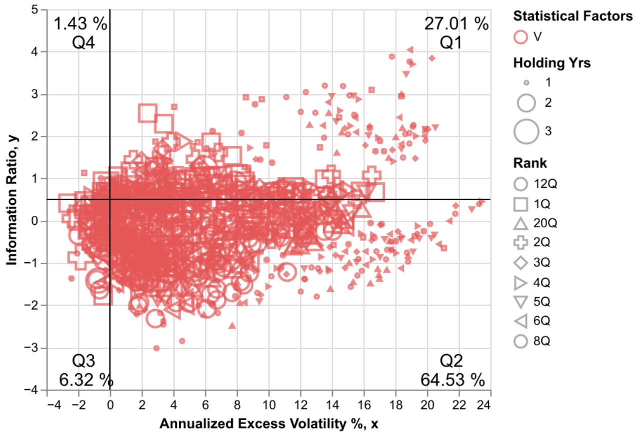
-Annualized excess volatility vs. Information ratio

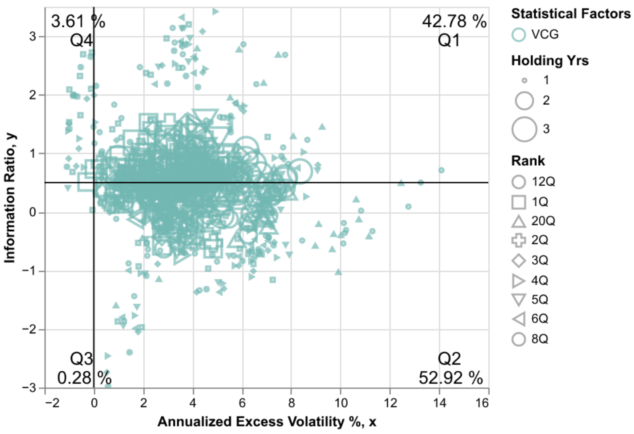
The cartesian chart of annualized excess volatility vs. information ratio plotted for different statistical factors, for different holding periods and for respective quarterly proxy rankings.

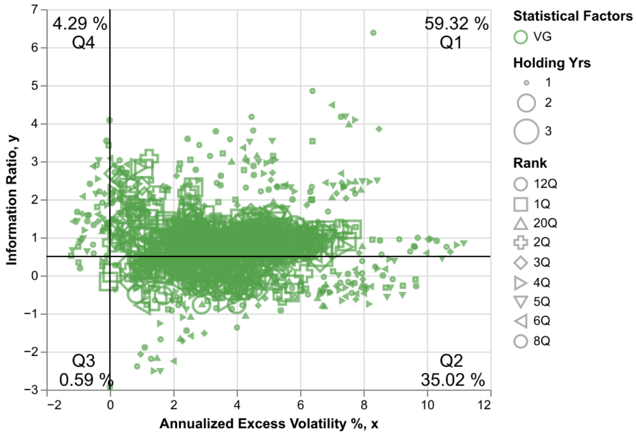






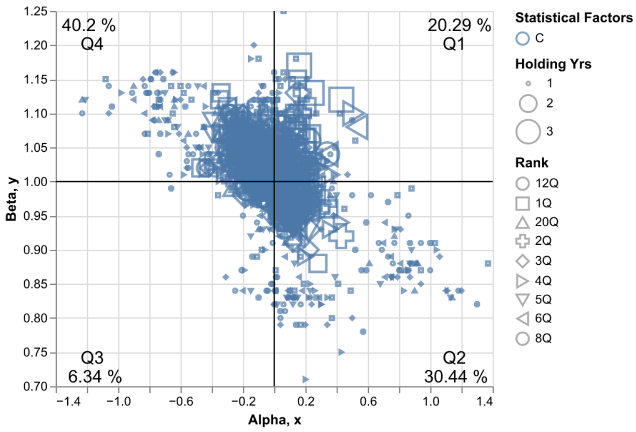


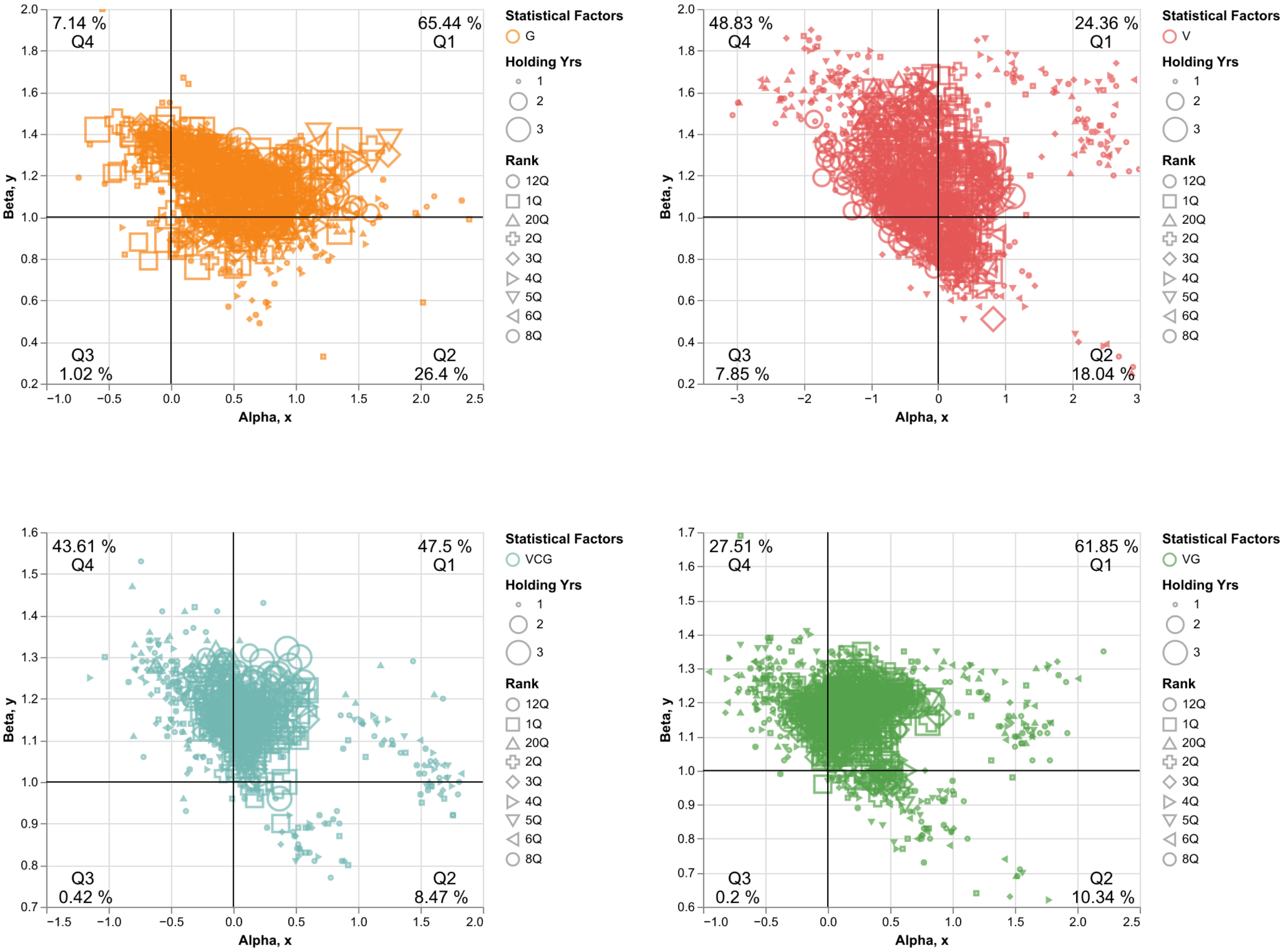




-Alpha vs. Beta

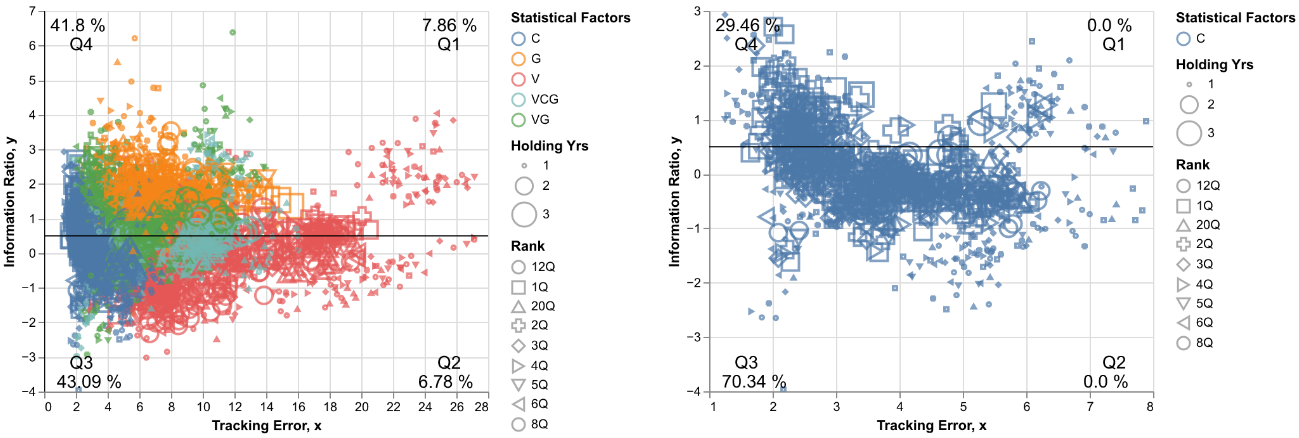
The cartesian chart of alpha vs. beta plotted for different statistical factors, for different holding periods and for respective quarterly proxy rankings.

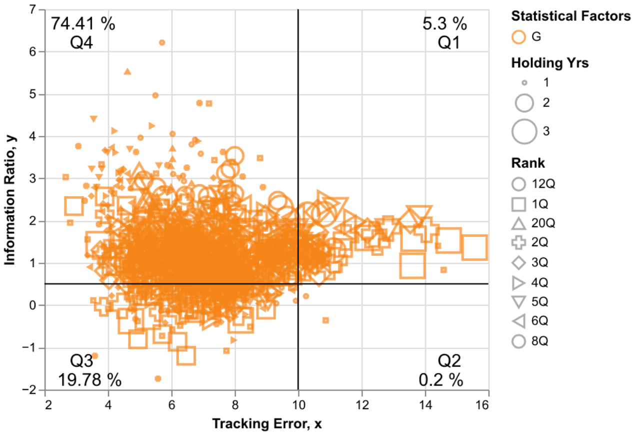


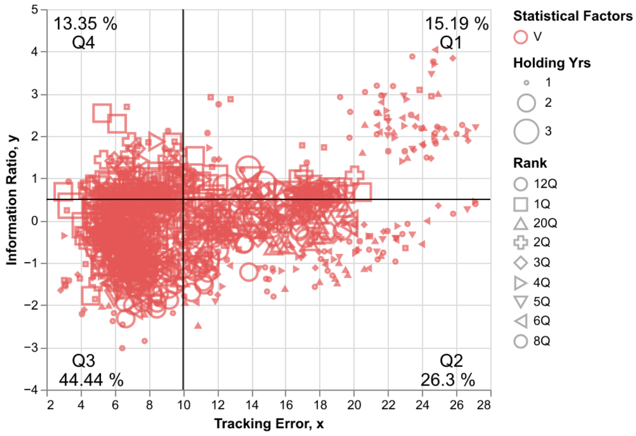


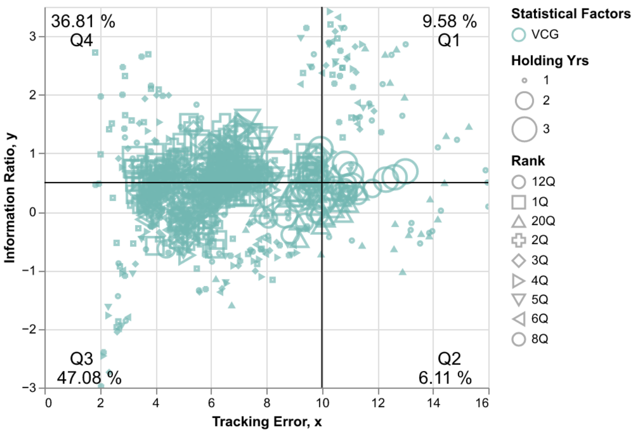
-Tracking error vs. Information ratio

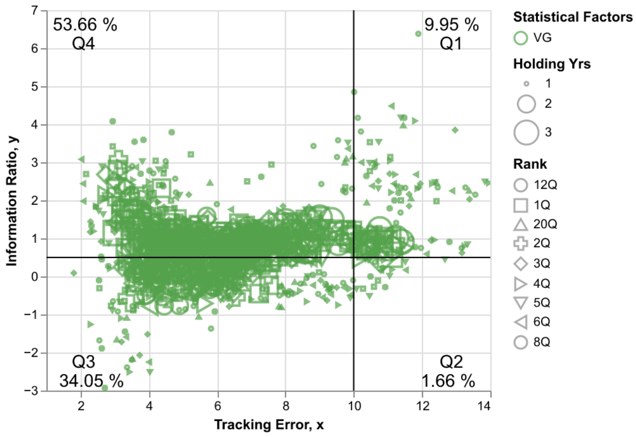
The cartesian chart of tracking error vs. information ratio plotted for different statistical factors, for different holding periods and for respective quarterly proxy rankings.











-Cartesian cluster analysis

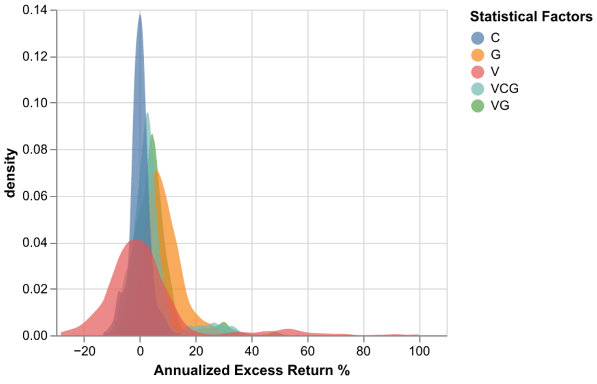
The table below carries the cluster analysis across four quadrants, highlighting positive and negative skew in the dataset, and risk-return characteristics of the various statistical factors. Overall, the VCG factor shows robustness across statistical metrics, while C and V statistical factors are with the most risk and skew.

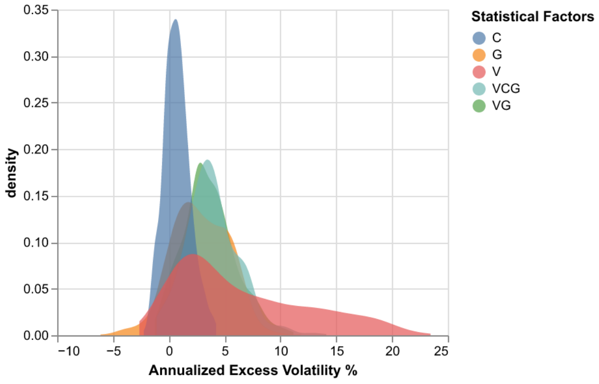
Statistics	Statistical Factors	Q1	Q2	Q3	Q4	Q1+Q2	Q2+Q3	Q3+Q4	Q4+Q1	Q1+Q3	Q2+Q4
Annualized Excess Return % vs. Annualized Excess Volatility %	C	29.85 %	19.41 %	9.27 %	41.17 %	49.26%	28.68%	50.44%	71.02%	39.12%	60.58%
	G	82.77 %	10.7 %	1.83 %	4.59 %	93.47%	12.53%	6.42%	87.36%	84.6%	15.29%
	V	43.43 %	3.47 %	4.38 %	48.62 %	46.9%	7.85%	53.0%	92.05%	47.81%	52.09%
	VCG	67.92 %	3.89 %	0.0 %	28.19 %	71.81%	3.89%	28.19%	96.11%	67.92%	32.08%
	VG	74.63 %	4.78 %	0.1 %	20.39 %	79.41%	4.88%	20.49%	95.02%	74.73%	25.17%
Annualized Excess Return % vs. Information Ratio	C	29.46 %	19.71 %	50.54 %	0.0 %	49.17%	70.25%	50.54%	29.46%	80.0%	19.71%
	G	79.71 %	13.56 %	6.42 %	0.0 %	93.27%	19.98%	6.42%	79.71%	86.13%	13.56%
	V	28.54 %	17.84 %	53.01 %	0.0 %	46.38%	70.85%	53.01%	28.54%	81.55%	17.84%
	VCG	46.39 %	25.0 %	28.19 %	0.0 %	71.39%	53.19%	28.19%	46.39%	74.58%	25.0%
	VG	63.61 %	15.12 %	20.59 %	0.0 %	78.73%	35.71%	20.59%	63.61%	84.2%	15.12%
Annualized Excess Volatility % vs. Information Ratio	C	18.44 %	52.59 %	17.66 %	10.93 %	71.03%	70.25%	28.59%	29.37%	36.1%	63.52%
	G	71.15 %	15.9 %	4.08 %	8.46 %	87.05%	19.98%	12.54%	79.61%	75.23%	24.36%
	V	27.01 %	64.53 %	6.32 %	1.43 %	91.54%	70.85%	7.75%	28.44%	33.33%	65.96%
	VCG	42.78 %	52.92 %	0.28 %	3.61 %	95.7%	53.2%	3.89%	46.39%	43.06%	56.53%
	VG	59.32 %	35.02 %	0.59 %	4.29 %	94.34%	35.61%	4.88%	63.61%	59.91%	39.31%
Alpha vs. Beta	C	17.46 %	30.44 %	6.34 %	40.2 %	47.9%	36.78%	46.54%	57.66%	23.8%	70.64%
	G	63.71 %	26.4 %	1.02 %	7.14 %	90.11%	27.42%	8.16%	70.85%	64.73%	33.54%
	V	23.14 %	18.04 %	7.85 %	48.83 %	41.18%	25.89%	56.68%	71.97%	30.99%	66.87%
	VCG	45.69 %	8.47 %	0.42 %	43.61 %	54.16%	8.89%	44.03%	89.3%	46.11%	52.08%
	VG	60.59 %	10.34 %	0.2 %	27.51 %	70.93%	10.54%	27.71%	88.1%	60.79%	37.85%
Tracking Error vs. Information Ratio	C	0.0 %	0.0 %	70.34 %	29.46 %	0.0%	70.34%	99.8%	29.46%	70.34%	29.46%
	G	5.3 %	0.2 %	19.78 %	74.41 %	5.5%	19.98%	94.19%	79.71%	25.08%	74.61%
	V	15.19 %	26.3 %	44.44 %	13.35 %	41.49%	70.74%	57.79%	28.54%	59.63%	39.65%
	VCG	9.58 %	6.11 %	47.08 %	36.81 %	15.69%	53.19%	83.89%	46.39%	56.66%	42.92%
	VG	9.95 %	1.66 %	34.05 %	53.66 %	11.61%	35.71%	87.71%	63.61%	44.0%	55.32%

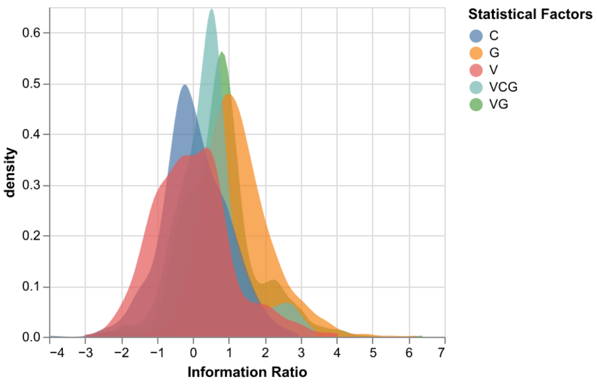
-Statistical distributions

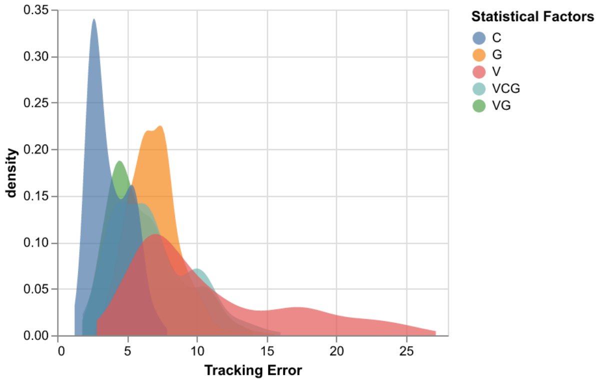
-Area charts

The area charts below showcase the distribution of annualized excess return, annualized excess volatility and information ratio across various statistical factors.





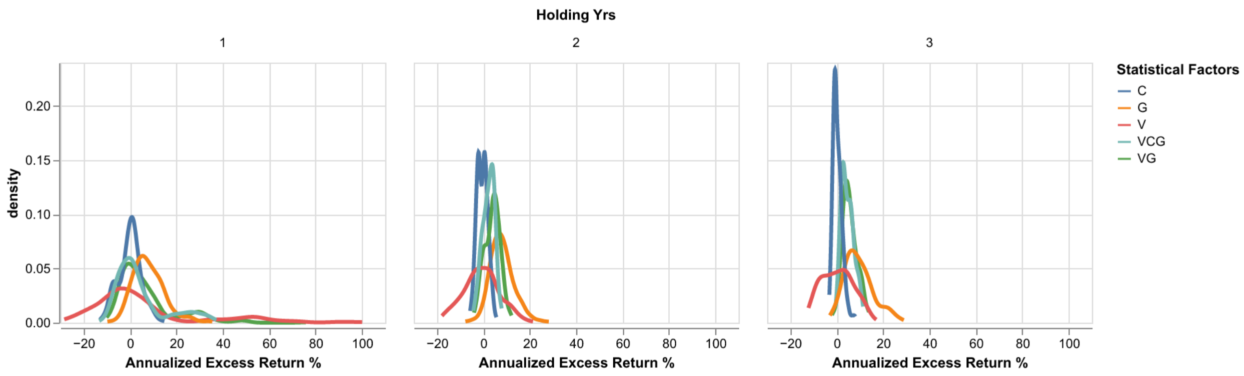


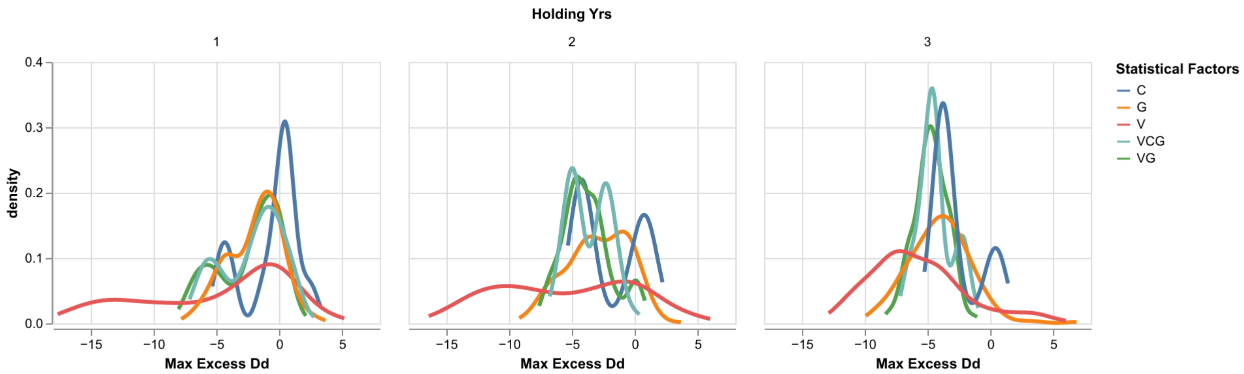


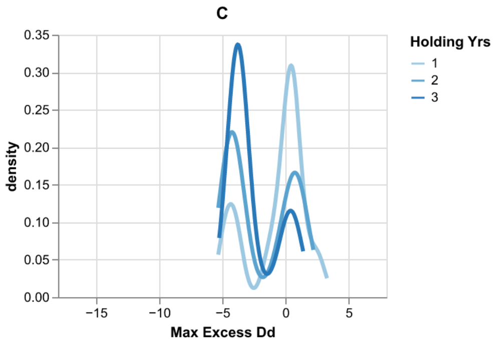
Statistical Factors	Mean	Median	Mode	Std Dev	Variance	Skewness	Kurtosis	Range
C	-0.13	-0.05	1.58	3.6	12.95	-0.01	1.59	27.85
G	8.38	7.62	12.62	6.27	39.25	0.76	1.18	45.2
V	2.32	-0.59	1.73	17.14	293.83	2.39	7.61	128.36
VCG	3.8	2.55	-2.34	7.96	63.43	1.91	4.33	50.47
VG	5.56	4.24	4.24	8.52	72.57	2.57	10.22	85.94

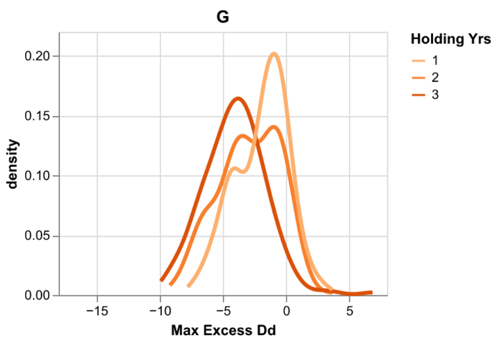
-Line charts

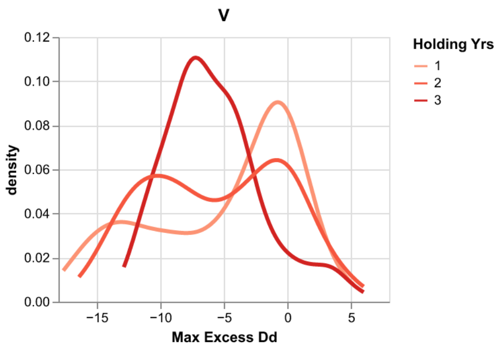
The line charts below showcase the annualized excess return distribution for each holding periods, max excess drawdown distribution for each holding periods and max excess drawdown distribution for each statistical factor. Max excess drawdown represents the difference between portfolio max drawdown and benchmark max drawdown.

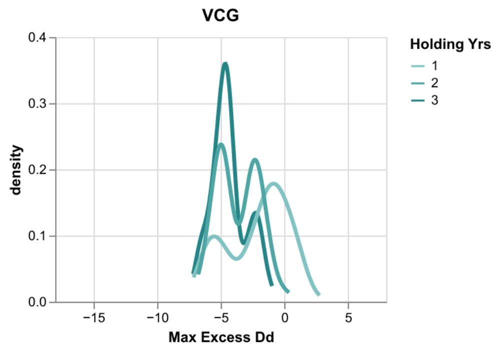


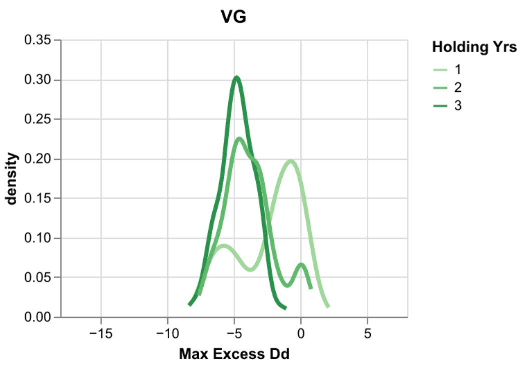












Statistical Factors	Holding Yrs	Mean	Median	Mode	Std Dev	Variance	Skewness	Kurtosis	Range
C	1Y	-0.14	0.27	1.44	4.86	23.62	-0.05	-0.02	27.85
C	2Y	-0.47	-0.29	-2.11	2.22	4.91	0.14	-0.57	11.33
C	3Y	0.36	-0.1	-0.93	1.95	3.81	1.07	1.54	11.21
G	1Y	8.02	7.04	4.84	6.88	47.38	0.84	1.3	45.2
G	2Y	8.05	7.61	6.56	5.21	27.13	0.55	0.96	35.95
G	3Y	9.57	8.71	12.62	6.27	39.37	0.7	0.2	31.88
V	1Y	5.61	-1.1	-4.72	24.02	576.85	1.58	2.19	128.36
V	2Y	-0.75	-0.76	-17.86	7.6	57.83	0.08	-0.23	39.28
V	3Y	0.37	0.03	-5.47	6.68	44.63	0.15	-0.89	29.25
VCG	1Y	4.51	1.02	-2.34	11.07	122.49	1.3	0.73	50.47
VCG	2Y	2.23	2.45	0.44	2.57	6.63	-0.28	-0.53	12.21
VCG	3Y	4.73	4.5	2.28	2.63	6.94	0.51	-0.56	11.92
VG	1Y	6.83	3.32	-0.52	12.17	148.21	1.72	3.43	85.94
VG	2Y	3.86	4.17	0.24	3.32	11.0	-0.09	-0.5	16.21
VG	3Y	5.61	5.18	4.05	3.18	10.1	0.41	-0.13	15.96

-Modern portfolio theory (MPT) statistics

The table of averaged MPT statistics for each statistical factor and holding periods without outliers.

(**AR** - Annualized Excess Returns, **AV** - Annualized Excess Volatility, **TE** - Tracking Error, **IR** - Information Ratio, **Max Excess Dd** - Maximum Excess Drawdown)

Statistical Factors	Holding Yrs	AR	AV	TE	IR	Alpha	Beta	Max Excess Dd
C	1Y	-0.38	0.58	3.96	-0.03	0.02	0.99	-0.69
C	2Y	-0.48	0.83	3.66	-0.01	-0.06	1.02	-1.93
C	3Y	0.1	0.75	3.72	0.04	-0.01	1.02	-3.1
G	1Y	8.46	1.75	6.39	1.44	0.61	1.04	-1.91
G	2Y	8.79	2.96	7.07	1.27	0.57	1.12	-2.67
G	3Y	9.45	4.49	7.83	1.17	0.52	1.2	-3.82
V	1Y	4.72	8.92	13.11	0.02	-0.18	1.29	-5.73
V	2Y	-0.91	5.94	10.22	-0.15	-0.32	1.2	-5.9
V	3Y	0.3	4.63	9.83	-0.1	-0.09	1.13	-6.24
VCG	1Y	4.42	4.04	7.02	0.43	0.13	1.14	-2.1
VCG	2Y	2.37	3.56	6.03	0.38	0.01	1.15	-3.8
VCG	3Y	4.15	3.78	6.91	0.6	0.16	1.15	-4.35
VG	1Y	6.7	3.67	6.35	0.87	0.27	1.15	-2.4
VG	2Y	3.91	3.3	5.72	0.71	0.14	1.15	-3.76
VG	3Y	5.7	3.9	6.87	0.8	0.27	1.17	-4.75