



UNIVERSITÀ
DI PAVIA

DEPARTMENT OF ECONOMICS AND MANAGEMENT
MASTER PROGRAMME IN FINANCE

***Measuring Portfolio Diversification: A Comparative Analysis
of Twelve Diversification Measures Across Alternative
Portfolio Construction Strategies***

Supervisor

Prof. Riccardo Brignone

Student

Emiljana Nezha

Academic Year 2025/2026

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Riccardo Brignone, for his valuable guidance and support throughout the preparation of this thesis.

I am especially thankful for his patience, availability and encouragement, which helped me navigate the different stages of the research and bring this work to completion.

Abstract

Diversification is widely regarded as one of the central principles of portfolio management, yet more than seventy years after Markowitz's foundational work, no single, universally accepted quantitative definition of the concept exists. This thesis examines twelve diversification measures spanning weight based and risk based approaches, the latter taking several more specific forms depending on whether the focus is factor exposure, downside risk, or realized rebalancing return and asks whether they provide a consistent assessment of portfolio diversification when applied to portfolios constructed under different allocation objectives.

The empirical analysis uses weekly return data for eleven exchange-traded funds covering the major global asset classes, from January 2010 to January 2026. Six portfolio construction methods Equal Weight, GMV Capped, Tangency, Mean CVaR, MV Frontier, and Maximum Diversification are built using a 104-week rolling estimation window and evaluated out of sample over 167 monthly rebalancing dates between February 2012 and December 2025. All twelve diversification measures are computed for each portfolio at every rebalancing date, and their behaviour is assessed through measures of temporal stability, pairwise correlation, overall agreement, post-hoc pairwise testing and predictive regressions against future realized risk.

The results show that the twelve measures agree with each other only moderately and that this disagreement is statistically systematic rather than sampling noise. Two broad families of measures emerge consistently across the analyses: a weight based family that rewards evenly spread capital allocation, and a risk based family that rewards efficient allocation of risk regardless of the number of assets held. Measures within each family agree closely with one another; measures across families frequently do not. Notably, the two Effective Number of Bets measures, built on the same

underlying logic of decomposing portfolio risk into independent factors end up in different families once tested empirically, with one tracking the risk based group and the other tracking the weight based group, underscoring that a measure's theoretical construction does not always predict how it behaves in practice. The Effective Number of Bets based on Minimum Torsion and the Diversification Ratio prove the most temporally stable and the most predictive of future portfolio risk, while several weight-based measures, despite describing current portfolio structure well, carry little forward-looking information. At the portfolio level, the best performing strategy out of sample (MV Frontier) does not score highest on any diversification measure, underscoring that portfolio optimisation and diversification measurement answer genuinely different questions.

These findings suggest that diversification is not a single, well defined portfolio property, and that relying on any one measure risks capturing only part of a genuinely two dimensional concept. The thesis concludes with recommendations for combining complementary measures in both research and practice.

Abstract

La diversificazione è ampiamente considerata uno dei principi cardine della gestione del portafoglio, eppure, a più di settant'anni dal lavoro pionieristico di Markowitz, non esiste ancora una definizione quantitativa del concetto universalmente accettata. Questa tesi esamina dodici indicatori di diversificazione, che spaziano dagli approcci basati sui pesi a quelli basati sul rischio, questi ultimi con diverse forme più specifiche a seconda che l'attenzione sia focalizzata sull'esposizione ai fattori, sul rischio di ribasso o sul rendimento di ribilanciamento realizzato, e si interroga sulla loro capacità di fornire una valutazione coerente della diversificazione del portafoglio quando applicati a portafogli costruiti con diversi obiettivi di allocazione.

L'analisi empirica utilizza i dati di rendimento settimanali di undici ETF che coprono le principali classi di attività globali, da gennaio 2010 a gennaio 2026. Sei metodi di costruzione del portafoglio (Equal Weight, GMV Capped, Tangency, Mean CVaR, MV Frontier e Maximum Diversification) sono stati costruiti utilizzando una finestra di stima mobile di 104 settimane e valutati out-of-sample su 167 date di ribilanciamento mensili tra febbraio 2012 e dicembre 2025. Tutte e dodici le misure di diversificazione sono state calcolate per ciascun portafoglio ad ogni data di ribilanciamento e il loro comportamento è stato valutato attraverso misure di stabilità temporale, correlazione a coppie, concordanza complessiva, test a coppie post-hoc e regressioni predittive rispetto al rischio realizzato futuro.

I risultati mostrano che le dodici misure concordano tra loro solo moderatamente e che questa discrepanza è statisticamente sistematica piuttosto che dovuta al rumore di campionamento. Due ampie famiglie di indicatori emergono costantemente dalle analisi: una famiglia basata sui pesi che premia un'allocazione del capitale uniformemente distribuita e una famiglia basata sul rischio che premia un'allocazione efficiente del rischio indipendentemente dal numero di attività detenute.

Gli indicatori all'interno di ciascuna famiglia concordano strettamente tra loro; gli indicatori appartenenti a famiglie diverse, invece, spesso non lo fanno. In particolare, i due indicatori del Numero Effettivo di Scommesse, costruiti sulla stessa logica di base di scomposizione del rischio di portafoglio in fattori indipendenti, risultano appartenere a famiglie diverse una volta testati empiricamente, con uno che segue il gruppo basato sul rischio e l'altro quello basato sui pesi, sottolineando come la costruzione teorica di un indicatore non sempre preveda il suo comportamento nella pratica. Il Numero Effettivo di Scommesse basato sulla Torsione Minima e il Rapporto di Diversificazione si dimostrano i più stabili nel tempo e i più predittivi del rischio futuro del portafoglio, mentre diversi indicatori basati sui pesi, pur descrivendo bene l'attuale struttura del portafoglio, forniscono poche informazioni prospettiche. A livello di portafoglio, la strategia con le migliori performance tra quelle analizzate (MV Frontier) non ottiene il punteggio più alto in nessuna delle misure di diversificazione, a dimostrazione del fatto che l'ottimizzazione del portafoglio e la misurazione della diversificazione rispondono a quesiti realmente diversi.

Questi risultati suggeriscono che la diversificazione non è una singola proprietà ben definita del portafoglio e che affidarsi a una sola misura rischia di cogliere solo una parte di un concetto in realtà bidimensionale. La tesi si conclude con raccomandazioni per la combinazione di misure complementari sia nella ricerca che nella pratica.

CONTENTS

CHAPTER 1	11
Introduction	11
1.1 Context of the research	11
1.2 Research Questions and Motivation	13
1.3 Structure of the Thesis	15
CHAPTER 2	16
Defining and Measuring Diversification	16
2.1 The Concept of Portfolio Diversification	16
2.1.1 Historical Foundations of Diversification Theory	16
2.2 Defining Diversification: A Multidimensional Concept	19
2.2.1 Axiomatic Foundations of Diversification	21
2.2.2 Properties of Diversification Measures	23
2.2.3 Diversification as a Multidimensional Concept	24
Chapter 3	27
Research Methodology	27
3.1 Research Design	27
3.2 Data Collection, Sample Construction and Data Pre-processing	29
3.2.1 Data Collection and Investment Universe	29
3.2.2 Sample Construction	30
3.2.3 Data Pre-processing	31
3.3 Portfolio Construction Methods	32
3.3.1 Equal Weight Portfolio	32
3.3.2 Tangency Portfolio	33
3.3.3 Mean CVaR Portfolio	35
3.3.4 Mean Variance Efficient Frontier Portfolio	36
3.3.5 Global Minimum Variance Portfolio with a Maximum Weight Constraint (GMV Capped)	37
3.3.6 Maximum Diversification Portfolio	38

3.4 Portfolio Diversification Measures	40
3.4.1 Entropy (exp)	40
3.4.2 Effective Number of Constituents (ENC).....	42
3.4.3 Diversification Ratio (DR).....	42
3.4.4 Effective Number of Bets Based on Principal Component Analysis (ENB PCA)	44
3.4.5 Effective Number of Bets Based on Minimum Torsion (ENB MT).....	45
3.4.6 Risk Contribution Entropy (RC Entropy).....	46
3.4.7 Portfolio Diversification Index (PDI weighted)	47
3.4.8 Rao Neff Diversification Measure	48
3.4.9 Diversification Return.....	49
3.4.10 Diversification Delta (DD).....	50
3.4.11 CVaR Diversification Ratio (CvaR DR)	51
3.4.12 Diversification Quality Based on Expected Shortfall (DQ ES).....	52
CHAPTER 4	53
Empirical Results and Analysis	53
4.1 Descriptive Statistics of the Selected ETFs	53
4.2 Correlation Structure	55
4.2.1 Principal Component Structure.....	57
4.3 Portfolio Composition	59
4.4 Portfolio Performance	61
4.5 Diversification Measures	63
4.6 Stability Analysis.....	65
4.7 Measure Correlation.....	67
4.8 Measure Agreement	69
4.8.1 Post-Hoc Analysis: Which Measures Differ Significantly? (Nemenyi Test).....	70
4.8.2 Predictive Ability of Diversification Measures.....	72
4.9 Discussion	74
CHAPTER 5	75
Conclusions and Recommendations	76
5.1 Conclusions	76

5.2 Limitations of the Study.....	78
5.3 Recommendations for Future Research	79
Bibliography	80

List of Tables

Table 3.1 Investment Universe	10
Table 4.1 Descriptive Statistics of the Eleven ETFs	53
Table 4.2 Correlation Matrix of Weekly ETF Returns	56
Table 4.2.1 Principal Component Loadings and Variance Explained	58
Table 4.3 Average Portfolio Weights	59
Table 4.4 Out of Sample Portfolio Performance	61
Table 4.5 Average Diversification Scores	63
Table 4.6 Temporal Stability of Diversification Measures	65
Table 4.7 Spearman Rank Correlation Matrix	67
Table 4.8 Summary of Agreement Statistics	69
Table 4.8.1 Nemenyi Post-Hoc Pairwise P-Values	71
Table 4.8.2 Predictive Regressions of Future Risk	72

CHAPTER 1

Introduction

1.1 Context of the research

Diversification is one of the oldest ideas in finance, and also one of the most misunderstood. It gets called the only free lunch in investing (Fernholz, 2002), meaning it's the rare case where you can reduce risk without necessarily giving up return. The basic logic behind it goes back much further than modern finance. Bernoulli was writing about it in 1738, long before anyone had formalized portfolio theory, arguing that it made sense to split goods exposed to danger into several portions rather than risk them all at once. That's diversification in a sentence, really, and it took over two hundred years for the idea to get a proper mathematical treatment.

That treatment came from Markowitz in 1952, with what became known as Modern Portfolio Theory. His mean variance framework builds diversification into portfolio construction through the covariance between asset returns; the whole point of the model is that combining assets that don't move perfectly together reduces total portfolio risk below what you will get from any of them individually. It is a genuinely elegant result, and it is still the backbone of how most people think about portfolio construction today. But there is a gap in it that does not get talked about enough: Markowitz's model never actually defines diversification as its own concept. It shows you how to build a portfolio that achieves a certain risk reduction, but it does not give you a way to measure, after the fact, how diversified that portfolio actually is.

That gap did not go unnoticed. Over the following decades a fairly large number of alternative measures got proposed to fill it some based on how weights are spread across a portfolio, some on how risk itself is distributed, some on the underlying factor structure driving returns. And yet,

despite all this work, the concept remains oddly hard to pin down. Ask most people what a diversified portfolio looks like and they'll give you a reasonable answer without much hesitation: it's one where a problem with a single holding doesn't drag the whole thing down. Simple enough as an intuition. Turning that intuition into one agreed upon number, though, has proven a lot harder than it sounds. Meucci (2009) makes this point directly there is still no broadly accepted, precise, quantitative definition of diversification in the literature, more than seventy years after Markowitz. This is not just a theoretical loose end. It has real consequences for anyone trying to build a diversified portfolio, and 2008 made that painfully clear. A lot of portfolios that looked well diversified going into the crisis, spread across many assets, many sectors, sometimes many strategies, turned out to share far more risk than their construction suggested, the moment correlations moved together under stress. That experience is a big part of why both academics and practitioners have kept pushing for diversification measures that go beyond simply counting how many assets a portfolio holds.

1.2 Research Questions and Motivation

There is no shortage of diversification measures in literature at this point, each one built on a different theoretical foundation, each with its own strengths and its own blind spots. This thesis works through twelve of them, chosen deliberately to cover several genuinely different ways of thinking about diversification rather than twelve variations on the same idea. Roughly, they split into two broad families, each built around a different underlying logic.

The first family is weight based: Entropy (exp), the Effective Number of Constituents, Risk Contribution Entropy, and the Portfolio Diversification Index. These look at diversification through how evenly capital, or risk, ends up spread across a portfolio's holdings, the more balanced the allocation, the higher the score.

The second family is risk based: the Diversification Ratio, Rao Neff, and the CVaR Diversification Ratio. These care less about how many assets are held and more about how much of the portfolio's risk is genuinely independent once the correlations between assets and, for CVaR DR specifically, the shape of the downside are taken into account. Two further measures, Diversification Return and DQ ES, sit closer to the downside risk and realized return side of this divide: rather than scoring diversification directly the way the other measures do, they capture something closer to its consequences, the realised benefit of rebalancing across imperfectly correlated assets, in the case of Diversification Return, and the quality of downside protection specifically, in the case of DQ ES.

The two Effective Number of Bets measures deserve a separate note, because they don't sit neatly inside either family despite sharing a name and a common origin. Both decompose portfolio risk into a set of uncorrelated factors and count how many of those factors are actually doing work,

using Principal Component Analysis, the other using Minimum Torsion. In principle, that makes both of them risk-based by construction. Interestingly, the two measures, ENB PCA and ENB MT end up in different families despite sharing a name and a conceptual origin, which is itself one of the findings this thesis returns to in Chapter 4: ENB PCA tracks the risk-based family closely, while ENB MT tracks the weight-based family instead.

Given all of this, the central question this thesis is built around is fairly direct: among these twelve measures, do they actually agree with each other about how diversified a given portfolio is, once that portfolio has been built under a different allocation objective and if they do not agree, which ones hold up better over time, which ones are more mutually consistent, and which ones actually tell you something useful about future risk rather than just describing the present moment?

Part of what motivates this question is a gap I noticed in how most existing research is framed. A lot of studies stop at evaluating the realised performance of a portfolio that was built by optimising some particular diversification measure, how much return it made, how much risk it took to get there. Much less attention goes to whether different measures actually agree with each other when pointed at the same portfolios, or to whether some measures are more reliable than others once market conditions change. That gap matters more than it might first appear. A portfolio that looks highly diversified under one measure can look considerably less so under a different, equally reasonable measure, and anyone who leans on just one indicator risks reaching a conclusion that wouldn't survive being checked against another. This thesis is built around examining that disagreement head on across six portfolio construction methods, over a rolling out of sample period spanning more than a decade, using weekly return data for eleven exchange traded funds covering the major global asset classes.

1.3 Structure of the Thesis

Below is a presentation of how the rest of the thesis is organised.

Chapter 2 goes through the underlying concepts of risk and diversification. It lays out the main definitions of diversification found in the literature, discusses what properties a good diversification measure is generally expected to have, and works through a question that turns out to be less obvious than it sounds: whether a portfolio that minimises risk is, by that fact alone, actually diversified.

Chapter 3 sets out the research methodology, the data and the eleven ETF investment universe, the rolling out of sample back testing framework, the six portfolio construction methods examined (Equal Weight, GMV Capped, Tangency, Mean-CVaR, MV Frontier, and Maximum Diversification), and the twelve diversification measures introduced above, along with how each one is actually calculated.

Chapter 4 covers the empirical results: the statistical and correlation structure of the investment universe, how the six portfolios are composed and how they performed out of sample, the average diversification scores each measure assigns to them, and a set of further analyses, how stable each measure is over time, how the measures correlate with each other, how much they agree overall, which pairs differ significantly, and whether any of them predict future risk, that together get at how consistent and how useful these twelve measures really are.

Chapter 5 concludes the thesis. It pulls together the main findings, discusses where the study falls short, and points toward directions future research could usefully take.

CHAPTER 2

Defining and Measuring Diversification

2.1 The Concept of Portfolio Diversification

2.1.1 Historical Foundations of Diversification Theory

Portfolio diversification is a key idea in today's finance world and is seen as one of the most important rules in how investments are managed. It has proved to be very difficult to define even if it is perhaps the most important of investment principles. The basic ideas behind diversification have changed a lot over many years, starting from simple observations about uncertainty and sharing risks, to advanced methods used today for building investment portfolios.

The idea of diversification started with the early work on probability theory and how people make decisions when they aren't sure about the outcome. One of the first talks about this idea came from Bernoulli in 1738. He said that when people are unsure about the results of their choices, they should not put all their money into just one chance.

For nearly two centuries, diversification remained largely an intuitive investment principle without a formal mathematical foundation. Investors generally accepted that spreading investments across different opportunities reduced exposure to adverse outcomes, yet no rigorous framework existed for determining how diversification should be implemented or evaluated. Diversity became the focus of methodical scientific study just in the middle of the twentieth century.

The modern theory of diversification originates with the contribution of Markowitz (1952), whose article "Portfolio Selection" fundamentally revolutionized investment management. Markowitz challenged the existing practice of analysing investments individually and proposed that rational

investors should evaluate securities within the context of a portfolio. The key insight he had was that the risk of a portfolio depends not only on the characteristics of its underlying assets but also on the relationships among them. Specifically, portfolio variance is influenced by the covariance structure of asset returns, which suggests that combining asset with poor correlation can lower the total risk of the portfolio.

The significance of Markowitz's contribution goes well beyond the introduction of mathematical portfolio optimisation. By demonstrating that diversification derives its value from covariance relationships, Markowitz fundamentally altered the way investors think about risk. Prior to Modern Portfolio Theory, risk was largely viewed as an attribute of individual securities. Markowitz showed instead that risk is a property of the portfolio. Therefore, a risky asset may not add much risk to a diversified portfolio, while an apparently safe asset may increase portfolio risk if it exhibits strong positive correlation with existing holdings.

A significant outcome of Modern Portfolio Theory is the differentiation between overall risk and risk that can be mitigated through diversification. According to Markowitz (1952), investors can substantially reduce portfolio variance by combining assets with dissimilar return behaviours. This reduction in risk occurs without necessarily lowering expected returns and therefore creates an economic benefit that can be achieved at relatively low cost. The resulting relationship between risk and return is represented by the Efficient Frontier, which identifies portfolios offering the highest expected return for a given level of risk or the lowest level of risk for a given expected return.

Further theoretical advances confirmed Markowitz's framework's impact. Tobin (1958) developed Modern Portfolio Theory using the Separation Theorem, demonstrating that portfolio selection could be divided into two independent decisions, such as the selection of an optimal risky portfolio

and the allocation of wealth between the risky portfolio and a risk-free asset. This extension simplified the investment decision-making process and further strengthened the theoretical importance of diversified portfolios.

The development of the Capital Asset Pricing Model (CAPM) by Sharpe (1964), Lintner (1965), and Mossin (1966) provided additional support for diversification theory. Within the CAPM framework, investors are rewarded only for bearing systematic risk because unsystematic risk can be eliminated through diversification. This outcome established diversification as a central implication of equilibrium asset pricing theory. If unsystematic risk can be removed without reducing expected return, rational investors should hold diversified portfolios rather than concentrated positions.

One of the most significant results of diversification theory is still the difference between systematic and unsystematic risk. Systematic risk represents exposure to broad economic forces affecting the whole market, including changes in interest rates, inflation, economic growth, geopolitical events, and financial crises. Because these risks affect most assets simultaneously, diversification cannot remove them. Unsystematic risk, by contrast, arises from firm-specific events such as management decisions, operational failures, litigation, and developments in product markets. Since these risks are mostly unrelated across firms, correct portfolio building can significantly minimize them.

Diversification theory's risk reduction advantages have mostly been supported by empirical data. Evans and Archer (1968) showed that maintaining comparatively small portfolios of stocks helps to reduce a sizable amount of unorganized risk. Fisher and Lorie (1970) and Statman (1987) later refined their results, demonstrating that the level of diversification attained relies not just on the number of assets owned but also on the correlation structure among such assets. These studies

confirmed a crucial lesson from Modern Portfolio Theory: diversification is essentially a matter of reliance rather than only of portfolio size.

Modern Portfolio Theory has had a huge impact, but academics have progressively come to see that diversification is more complicated than they first thought. The expansion of financial markets, the rise of alternative asset classes, developments in factor investing, and the growing significance of downside risk management have all helped to expand the understanding of diversification. In addition to securities, contemporary investors spread their portfolios across countries, industries, asset classes, investment styles, and systematic risk variables. Consequently, diversification has developed from a somewhat simple portfolio building guideline into a multidimensional notion including many distinct risk management components.

Although most people agree that diversification is good, different theoretical viewpoints highlight different components of diversification, which helps to explain why so many indices based on portfolio weights, covariance patterns, risk contributions, exposure to factors, and losses of risk have grown developed.

2.2 Defining Diversification: A Multidimensional Concept

Even though diversification is an important aspect in finance, there is not yet a clear and widely accepted mathematical definition for it. The fundamental concept is that investors should avoid over concentration by spreading their money among several investing possibilities, but it is difficult to convert this notion into a particular portfolio feature.

Existing research presents this lack of clarity. Meucci (2009) notes that although diversification is crucial for portfolio management, does not exist a universally accepted quantitative definition for it. Similarly, Fragkiskos (2014) highlights the existence of several different approaches such as

diversification based on the number of assets, correlations between returns, tail risk, contributions to risk, principles from information theory, and foundational portfolios.

The fact that there are different ways to do it shows that the debate about diversification goes beyond just technical measurement problems. Rather, these several techniques could point to different perspectives on what diversification fundamentally present.

Diversification can be viewed as spreading money across numerous investment options instead of concentrating on a limited number of assets. A portfolio with a wide range of weightings is typically regarded as more varied than one with a small number of important assets from this angle. However, merely having a significant portfolio is insufficient. Markowitz (1952) showed that the covariance correlations between asset returns mainly depend on the benefits of diversifying your investments. Including assets that resemble ones you already own could increase the total number of holdings without significantly enhancing actual diversification.

Thus, a second viewpoint considers diversity as the distribution and interaction of portfolio risk. Reilly and Brown (2012) pointed that integrating assets with returns not exactly positively correlated reduces unnecessary risk through diversification.

But it is important to understand that we should not mistake diversification just for risk reduction. A minimum variance portfolio, for instance, would be mostly invested in a small number of low risk assets and have a low overall risk. Though connected concepts, low portfolio risk and strong diversification are not always related .

This is the reason why this disparity indicates a more structured study of diversification.

2.2.1 Axiomatic Foundations of Diversification

When we talk about diversification, there is a theoretical way of understanding it that does not rely on any statistical formula, like variance, but instead on the logic of preferences of a person making decisions under uncertainty.

Let X denote a set of uncertain alternatives and let $\succeq$ represent the preference relation of a decision maker, with $\sim$ denoting indifference. Consider N alternatives $x_1, \dots, x_N \in X$ and non-negative weights $\alpha_1, \dots, \alpha_N$, satisfying

$$\alpha_i \geq 0, \quad \sum_{i=1}^N \alpha_i = 1$$

A decision maker exhibits preference for diversification when

$$x_1 \sim x_2 \sim \dots \sim x_N \Rightarrow \sum_{i=1}^N \alpha_i x_i \succeq x_j,$$
$$j = 1, \dots, N$$

The economic interpretation is that when several uncertain alternatives are regarded as equally desirable, a convex combination of those alternatives should be at least as desirable as concentration in any single one of them. So we can say that diversification is therefore represented as a preference for combination over concentration.

This concept is closely related to the convexity of preferences. For two alternatives $x, y \in X$, convexity implies that

$$x \succcurlyeq y \Rightarrow \alpha x + (1 - \alpha)y \succcurlyeq y, \quad 0 \leq \alpha \leq 1$$

The importance of this formulation is that diversification does not depend on a specific statistical definition of risk. A diversified combination need not minimize variance or another particular risk measure in order that it to be preferred.

Chateauneuf and Tallon (2002) extend this reasoning through more specific notions of diversification, including sure diversification and comonotonic diversification. Under sure diversification, it is supposed that a decision maker is indifferent among several uncertain alternatives,

$$x_1 \sim \dots \sim x_N,$$

and that their convex combination generates a certain outcome c :

$$\sum_{i=1}^N \alpha_i x_i = c$$

Preference for sure diversification implies

$$c \succcurlyeq x_i, \quad i = 1, \dots, N$$

The idea behind it is that if combining equally desirable uncertain alternatives eliminates uncertainty completely, the resulting certain position should be at least as desirable as the original uncertain positions.

The concept of comonotonicity further emphasizes the role of dependence. Two uncertain alternatives x and y are comonotonic when

$$[x(\omega) - x(\omega')][y(\omega) - y(\omega')] \geq 0, \quad \forall \omega, \omega' \in \Omega$$

Because comonotonic alternatives always move in the same direction relative to one another, there's little chance for one position to offset losses in another. This confirms something portfolio theory has long emphasized, the fact that holding more positions does not necessary guarantee real diversification if those positions tend to rise and fall together.

This axiomatic perspective matters for the present study because it shows that diversification is a broader idea than just minimizing risk. At its core, it reflects a preference for spreading exposure rather than concentrating it, but how much benefit that spreading delivers depends mostly on how the underlying risks relate to one another.

2.2.2 Properties of Diversification Measures

Axiomatic literature addresses what it means to prefer diversification. A different aspect is how a quantitative measure of diversification should behave. Carmichael et al. (2015) identify several desirable properties that provide a useful conceptual framework for evaluating diversification measures.

First, a portfolio holding only one asset should present the lowest degree of diversification. This shows degeneracy with respect to portfolio size and follows directly from the intuitive requirement that diversification involves exposure to more than one investment opportunity.

Second, a portfolio composed of perfectly similar assets should also score minimal diversification. Increasing the number of economically identical positions should not create real diversification. This represents a boundary case with respect to asset dissimilarity.

Third, a meaningful diversification measure should satisfy duplication invariance. Artificially duplicating an existing asset should not increase measured diversification simply because the same economic exposure is represented multiple times.

Fourth, diversification should generally be non-decreasing with portfolio size, provided that the additional assets represent genuinely distinct investment opportunities.

Finally, diversification should be non-decreasing with asset dissimilarity. Other things being equal, a portfolio composed of more dissimilar assets should provide greater diversification than one composed of highly similar assets.

These qualities help to distinguish between the idea of diversification and its quantification. While these qualities pertain to how an indicator meant to assess diversification should act, the axiomatic framework addresses what diversification means.

This is especially important when contrasting several quantitative methods. Two indicators developed from separate theoretical underpinnings could react differently to variations in portfolio mix without always being incorrectly specified.

2.2.3 Diversification as a Multidimensional Concept

Taken together, the theoretical literature indicates that diversification cannot be reduced to a single portfolio trait. Rather, it reflects several distinct but related underlying logics.

For the purposes of this study, these can be grouped into two broad dimensions.

The first is allocation diversification — how capital is distributed among portfolio assets. Its primary concern is whether portfolio exposure is concentrated in a small number of holdings or spread more evenly across them, independent of how those assets actually behave.

The second is risk diversification, how portfolio risk itself, rather than capital, is distributed and how the underlying assets interact. It rests on the observation that equal capital allocations do not necessarily produce equal risk contributions, so the diversification benefit of a portfolio ultimately depends on the relationship structure among asset returns, not on the allocation alone.

Risk diversification is not a single, uniform idea, however; it takes several more specific forms depending on which aspect of risk is being examined. Factor diversification looks at exposure to independent underlying sources of systematic risk rather than to individual securities as such from this angle, a set of nominally different stocks may offer little real diversification if they are all driven by the same economic factor, a distinction emphasised by Meucci (2009). Downside-risk diversification examines whether diversification benefits hold up specifically under adverse market conditions, reflecting the fact that dependence structures between assets can shift under financial stress, so that a portfolio which appears well diversified in normal conditions may offer considerably less protection during extreme downturns.

These categories should not be read as a rigid taxonomy or as mutually exclusive definitions, and the boundary between them, particularly between the general notion of risk diversification and its more specific factor and downside focused forms is, in practice, less clean than the conceptual distinction suggests. They are best understood as a conceptual map of the perspectives most relevant to this study, rather than a strict classification that every diversification measure must fall neatly into.

What matters most for the empirical work that follows is that a portfolio can score very differently depending on which of these dimensions is being measured. An evenly weighted portfolio may show strong allocation diversification while remaining heavily exposed to a single systematic factor. Equally, a portfolio that looks diversified under normal return correlations may show

significant concentration in downside risk once financial stress changes how assets move together. Consequently, disagreement between quantitative measures built on different dimensions of diversification should not automatically be treated as evidence that one measure is misspecified. Instead, as Chapter 4 shows directly, different measures often capture genuinely different, and only partially overlapping, aspects of portfolio structure with some measures, as the Effective Number of Bets variants illustrate, not sitting as cleanly within one dimension as their construction might suggest.

Chapter 3

Research Methodology

In this chapter we will present the research methodology adopted to evaluate portfolio diversification measures. It describes the data collection process, sample construction, and data preprocessing procedures used in the empirical analysis. Then, it outlines the portfolio construction methods used to generate different portfolio allocations, followed by the diversification measures used to assess the degree of diversification achieved by each portfolio. Finally, the rolling out of sample back testing framework is presented, explaining how portfolio weights and diversification measures are estimated and evaluated over time. Together, these methodological components provide the foundation for the empirical analysis presented in Chapter 4.

3.1 Research Design

Using historical financial market data, this study explores and compare several portfolio diversification metrics through a quantitative empirical research design. The main goal is not to compare portfolio optimization techniques themselves, but to evaluate how various diversification measures behave when applied to portfolios built in line with several allocation and risk management criteria. Therefore, rather than the main subject of the study, portfolio development is a methodical approach for producing portfolios with varied features.

Diversification metrics across portfolios developed using six portfolio construction techniques, such as Equal Weight (EW), Global Minimum Variance with a Maximum Weight Constraint (GMV-Capped), Tangency, Mean-CVaR, Mean-Variance Efficient Frontier (MV Frontier), and Maximum Diversification Portfolio (MDP) are used evaluated using a comparative research

framework. Among these strategies are naïve allocation, constrained variance reduction, risk adjusted return optimisation, downside risk optimisation, mean variance allocation, and direct diversification optimisation, each of which reflects a unique way to build a portfolio. Their inclusion creates portfolios with varied allocation and risk characteristics, offering in this way a shared empirical environment in which different diversification metrics could be methodically evaluated.

Weekly returns from eleven exchange-traded funds (ETFs), representing the main asset classes around the world are used in the empirical study. A 104-week rolling estimate window is used to determine portfolio characteristics, and portfolio allocations are changed at monthly rebalancing periods. Rebalancing dates just use data acquired up to then for parameter estimation and portfolio creation. This method that rolls out of sample, stops look ahead bias and lets portfolio mix and diversification properties change naturally over time.

Following portfolio construction, the same group of twelve diversification measures is determined for every portfolio on every evaluation date. Using a common investment universe, estimation framework, and set of portfolio strategies provides a consistent basis for comparison while letting portfolio characteristics differ depending on the construction approach.

The statistical techniques detailed later in this chapter are used to compare the resulting diversification evaluations across portfolio constructions and over time. This structure lets the study investigate whether alternative diversification metrics offer coherent evaluations of portfolio diversification or seize separate and perhaps complementary aspects of the idea.

3.2 Data Collection, Sample Construction and Data Pre-processing

3.2.1 Data Collection and Investment Universe

The data used in our study consists of weekly return data for eleven exchange traded funds (ETFs) representing major global asset classes. The data were obtained from Yahoo Finance. These ETFs are chosen because they offer broad market exposure, high liquidity, and transparent pricing, making them appropriate for diversified investment portfolios. ETFs for U.S. equities, developed international equities, emerging market equities, U.S. Treasury bonds, investment grade and high yield corporate bonds, real estate, gold, and broad commodities are used. These asset classes taken together provide exposure to markets with different return characteristics and risk profiles, creating this way an appropriate setting for evaluating portfolio diversification measures.

Table 3.1 summarises the ETFs included in the empirical analysis and the asset class represented by each fund.

Table 3.1 Investment Universe

ETF	Asset Class
SPY	U.S. Equities
QQQ	U.S. Technology Equities
EFA	International Developed Equities
EEM	Emerging Market Equities
IEF	Intermediate-Term U.S. Treasury Bonds
TLT	Long-Term U.S. Treasury Bonds
LQD	Investment-Grade Corporate Bonds
HYG	High-Yield Corporate Bonds

VNQ	Real Estate Investment Trusts (REITs)
GLD	Gold
DBC	Broad Commodities

Historical adjusted closing prices were collected for all ETFs and then converted into weekly return series. These weekly returns, which provide the input data for portfolio development, diversification measurement, and portfolio evaluation throughout the study are subjected to empirical analysis

3.2.2 Sample Construction

Weekly return observations for eleven ETFs from 15 January 2010 to 2 January 2026 comprise the empirical sample. The whole dataset includes 834 weekly return observations, which is enough historical data for portfolio estimation and out of sample assessment.

The study uses a 104 week rolling estimate window to provide a realistic investment environment. The most recent 104 weekly observations are used on every portfolio creation date to project the covariance matrix and expected returns needed for portfolio construction. Portfolio weights are then applied to the following out of sample period once they are found, after that the estimation window is rolled forward and the process starts again.

The empirical study concentrates on the out of sample period from February 2012 to December 2025, during this time portfolios are adjusted monthly. We create a series of rolling portfolio allocations that represent evolving market conditions and guarantees us that every investing decision is only based on data accessible at the relevant rebalancing date. Therefore, the study

prevents look ahead bias and gives an objective evaluation of portfolio diversification metrics under actual investment circumstances.

3.2.3 Data Pre-processing

Before portfolio construction, the gathered market information were pre processed in order to guarantee consistency and applicability for the study. Historical adjusted closing values were converted into weekly return series, the main input for every next portfolio optimization analysis.

The weekly return data set was set up as a structured return matrix whereby every row is a weekly observation and every column reflects one of the eleven ETFs in the investment universe. This matrix served as the basis for projecting the statistical inputs needed for portfolio development inside every rolling estimation window.

To guarantee uniformity all through the analysis, all portfolios were assessed with the same return dataset, the same 104-week rolling estimation window, and the same monthly rebalancing plan. Only historical observations were utilized at every rebalancing date to project portfolio inputs, therefore assuring that no future knowledge found its way into the optimization process.

The resulting return matrix gave the sample covariance matrix, portfolio weights generated by each portfolio construction method, and input data for estimating expected returns. The portfolio weights so determined were then utilized to evaluate the portfolio diversity indices examined in this study.

3.3 Portfolio Construction Methods

The portfolio construction methods employed in this study are used to generate portfolios with heterogeneous allocation and risk characteristics, providing the empirical basis for the comparative evaluation of diversification measures. Portfolio optimisation is therefore not treated as the primary object of the analysis; rather, the different construction methods serve as a means of producing portfolio structures that vary systematically in terms of capital allocation, risk exposure, and concentration. This variation allows the selected diversification measures to be evaluated across portfolios with fundamentally different underlying characteristics. Six portfolio construction methods are considered: the Equal Weight (EW) portfolio, the Tangency portfolio, the Mean CVaR portfolio, the Mean Variance Efficient Frontier (MV Frontier) portfolio, the Global Minimum Variance portfolio with a Maximum Weight Constraint (GMV Capped), and the Maximum Diversification Portfolio (MDP). Together, these methods span a broad range of construction principles, from a simple, non-optimised allocation rule to optimisation strategies based on variance minimisation under a concentration constraint, risk-adjusted return maximisation, downside risk minimisation, mean variance efficiency, and explicit diversification maximisation.

3.3.1 Equal Weight Portfolio

The Equal Weight (EW) portfolio is the simplest allocation strategy used in this study and serves as the benchmark portfolio throughout the empirical analysis. Unlike the optimisation-based methods described in the following subsections, the EW portfolio allocates the same proportion of capital to each asset, without taking into account expected returns, risk, or the correlations among assets. As a result, every ETF contributes equally to the portfolio.

The portfolio weights are determined as:

$$w_i = \frac{1}{N}, i = 1, 2, \dots, N$$

where:

- w_i is the weight allocated to asset i ;
- N is the total number of assets included in the portfolio.

In our case , since the investment universe includes eleven ETFs, each asset receives a weight of 1/11 (9.09%) at every rebalancing date.

The EW portfolio is included as a benchmark because it does not rely on any statistical estimation or optimisation procedure. It therefore provides a simple, naive form of diversification, and offers a useful point of comparison for evaluating whether the more sophisticated portfolio construction methods used in this study lead to different diversification outcomes when assessed with the selected diversification measures.

3.3.2 Tangency Portfolio

The Tangency Portfolio seeks to maximise the portfolio's risk-adjusted return, by finding the combination of assets that gives the highest expected excess return per unit of risk. Unlike the GMV-Capped portfolio, which focuses only on minimising portfolio variance, the Tangency Portfolio also takes into account expected returns and the risk-free rate, together with the covariance structure of asset returns. As a result, it corresponds to the portfolio with the highest Sharpe ratio among all feasible portfolios under the given constraints.

The optimisation problem is defined as:

$$\max_w \frac{w^\top \mu - r_f}{\sqrt{w^\top \Sigma w}}$$

subject to

$$\sum_{i=1}^N w_i = 1, w_i \geq 0, i = 1, \dots, N$$

where:

- μ is the vector of expected weekly returns, estimated from the 104-week rolling estimation window;
- r_f is the risk-free rate, estimated using the Ken French weekly risk-free rate series for 2010–2024, supplemented by the 3-month U.S. Treasury Bill rate (converted to a bond-equivalent yield) for 2025, when the Ken French series was not yet available;
- Σ is the sample covariance matrix of weekly asset returns;
- w is the vector of portfolio weights.

The optimisation is re-run at each monthly rebalancing date, using updated estimates of expected returns, the covariance matrix, and the average risk-free rate over the corresponding 104-week window. The resulting portfolio maximises expected excess return relative to volatility, while satisfying the full-investment and long-only constraints.

This formulation follows the classical Tangency Portfolio of Sharpe (1964), developed within the Capital Asset Pricing Model (CAPM), and is consistent with Tobin's (1958) separation theorem.

In this study, the Tangency Portfolio serves as the return-oriented benchmark, allowing its

diversification characteristics and performance to be compared with those of the Equal Weight, GMV-Capped, Mean-CVaR, and MV Frontier portfolios.

3.3.3 Mean CVaR Portfolio

The Mean CVaR Portfolio is developed by minimising the portfolio's Conditional Value at Risk (CVaR) at the 95% confidence level, focusing on downside risk rather than on total return variability. The optimization follows the linear programming formulation of Rockafellar and Uryasev (2000), which estimates CVaR directly from the historical distribution of portfolio returns.

The optimisation problem is defined as:

$$\min_{w, \alpha, u} \alpha + \frac{1}{(1 - \beta)T} \sum_{t=1}^T u_t$$

subject to

$$u_t \geq 0, t = 1, \dots, T,$$

$$u_t \geq -r_t^T \mathbf{w} - \alpha, t = 1, \dots, T,$$

$$\sum_{i=1}^N w_i = 1,$$

$$w_i \geq 0, i = 1, \dots, N.$$

where:

- α is the portfolio Value at Risk;
- u_t are the auxiliary variables used to measure losses beyond VaR;
- $\beta = 95\%$ is the confidence level;

- T is the number of weekly observations in the rolling estimation window;
- w is the vector of portfolio weights.

The optimisation is repeated at each monthly rebalancing date, using the historical weekly returns within the corresponding estimation window. By minimising expected tail losses directly, the Mean CVaR portfolio offers a risk based allocation that complements the variance based GMV Capped model used in this study.

3.3.4 Mean Variance Efficient Frontier Portfolio

The Mean Variance Efficient Frontier (MV Frontier) portfolio is built by maximising expected portfolio return, subject to a fixed annual volatility target. Unlike the Tangency portfolio, which maximises the Sharpe ratio, this optimisation looks for the highest expected return that can be reached without exceeding the given risk limit.

The optimisation problem is formulated as:

$$\max_w w^T \mu$$

subject to

$$w^T \Sigma w \leq (10\% \text{ annually})^2,$$

$$\sum_{i=1}^N w_i = 1,$$

$$w_i \geq 0, i = 1, \dots, N.$$

where:

- μ is the vector of expected weekly returns;
- Σ is the sample covariance matrix estimated over the rolling estimation window;
- w is the vector of portfolio weights.

The optimisation is re runned at each monthly rebalancing date, with updated estimates of expected returns and portfolio risk. The resulting portfolio is the one with the highest attainable expected return that still satisfies the 10% target volatility constraint. In this study, the MV Frontier portfolio adds a further allocation strategy for testing the robustness and consistency of the selected diversification measures across different portfolio construction approaches.

3.3.5 Global Minimum Variance Portfolio with a Maximum Weight Constraint (GMV Capped)

The unconstrained Global Minimum Variance (GMV) portfolio minimises portfolio variance without placing any upper limit on individual asset weights. While this gives the lowest possible portfolio variance, it can also produce highly concentrated portfolios, since low volatility assets tend to dominate optimization. A preliminary implementation of this unconstrained version confirmed the pattern: it allocated a very large share of capital, often above 60%, to a small number of low-volatility fixed-income assets. This allocation is variance-optimal in an unconstrained sense but is not very appealing in practice given the concentration risk involved, and it produces a diversification profile that largely overlaps with the more conservative variant used in this study.

To keep the minimum-variance objective while reducing concentration risk, this study uses a constrained version of the GMV portfolio (GMV-Capped), where an upper bound is placed on the weight of any single asset. The optimisation problem is formulated as:

$$\min_w w^\top \Sigma w$$

subject to

$$\sum_{i=1}^N w_i = 1, 0 \leq w_i \leq c, i = 1, \dots, N$$

where:

- w is the portfolio weight vector;
- Σ is the covariance matrix of asset returns, estimated over the 104-week rolling estimation window;
- c is the maximum allocation allowed for any single asset.

In this study, the upper bound is set at $c = 0.25$, so no single asset can receive more than 25% of the portfolio. This keeps the minimum-variance objective while preventing the portfolio from concentrating too heavily in a small number of assets, and is re-estimated at each monthly rebalancing date using the most recent 104 weekly observations.

The GMV Capped portfolio is included because it represents a risk oriented allocation strategy, based on the relationships between asset returns rather than on expected performance, while being closer to a strategy that could realistically be implemented than its unconstrained counterpart. It therefore offers a useful point of comparison for testing whether portfolio diversification measures give consistent results across portfolios built on very different allocation principles.

3.3.6 Maximum Diversification Portfolio

The Maximum Diversification Portfolio (MDP), proposed by Choueifaty and Coignard (2008), builds portfolio allocations by directly maximising the Diversification Ratio (DR). Unlike the risk-

minimisation approaches used in this study, such as the GMV Capped and Mean CVaR portfolios, the MDP explicitly looks for the portfolio that achieves the highest level of diversification, by balancing individual asset volatilities against overall portfolio risk.

The optimisation problem is formulated as:

$$\max_w DR(w) = \frac{w^T \sigma}{\sqrt{w^T \Sigma w}}$$

subject to

$$\sum_{i=1}^N w_i = 1, w_i \geq 0, i = 1, \dots, N$$

where:

- w is the vector of portfolio weights;
- σ is the vector of individual asset volatilities;
- Σ is the covariance matrix of asset returns.

The objective function maximises the ratio between the weighted average volatility of the individual assets and the overall portfolio volatility. Higher values of the Diversification Ratio mean that portfolio risk has been reduced more effectively through imperfect correlations among the assets held.

Since the objective function is nonlinear, the optimization problem is solved numerically, under the same long only and full investment constraints used throughout this study.

3.4 Portfolio Diversification Measures

This study evaluates portfolio diversification using twelve measures, each capturing a different aspect of what "diversification" means. All of them try to quantify the same general idea how well a portfolio spreads its risk but they start from different theoretical grounds, so in practice they end up looking at the same portfolio from quite different angles. Some are based purely on how capital is allocated across assets, others on the covariance structure of returns, others on factor exposures derived from that covariance structure, and others still on downside risk specifically, rather than risk in general. As will become clear throughout this section, a portfolio that scores well on one of these dimensions does not necessarily score well on the others, which is precisely why relying on a single measure can be misleading.

For each portfolio, all twelve measures are calculated at every monthly rebalancing date, using the weights obtained from the rolling construction process described in Section 3.3. Applying the same set of measures across the different portfolio construction methods makes it possible to compare, in a consistent way, how each measure behaves, how sensitive it is to changes in allocation, and how useful it actually turns out to be once weights shift from one rebalancing period to the next. The following subsections go through each measure in turn: where it comes from, how it is calculated, and perhaps most importantly what it actually means when its value goes up or down.

3.4.1 Entropy (exp)

Entropy comes originally from information theory (Shannon, 1948), where it measures how much uncertainty, or "information," is contained in a random variable. Applied to a portfolio, the idea is to treat the weights themselves as if they were a probability distribution, an asset's weight is treated as its "probability" of being drawn. The more evenly this distribution is spread across the available

assets, the more uncertainty it contains, and under this framework the more diversified the portfolio is considered to be. A portfolio concentrated in one or two positions has very little uncertainty about where the risk is coming from; a portfolio spread evenly across many assets has a great deal.

The measure used in this study is the exponential form of Shannon entropy:

$$\text{Entropy}(\text{exp}) = \exp \left(- \sum_{i=1}^N w_i \ln(w_i) \right)$$

Where w_i is the weight allocated to asset i and N is the total number of assets in the portfolio.

The two extreme cases make the intuition behind the formula easy to understand. If the entire portfolio sits in a single asset, all the "probability mass" is concentrated in one place, and entropy collapses to its minimum value of 1, there is nothing left to be uncertain about. At the other extreme, if the weights are perfectly equal across all N assets, entropy reaches its maximum possible value of exactly N . This maximum case corresponds precisely to the Equal Weight portfolio described in Section 3.3.1, which is why that portfolio is expected, almost by construction, to dominate this particular measure.

It is worth being clear about what Entropy does and does not capture. It says nothing at all about the volatility of the individual assets, how they correlate with one another, or how attractive their expected returns are, it will only look at how the capital itself is spread out. A portfolio can therefore score very highly on Entropy while still being poorly diversified from a risk standpoint, if, for example, several of its equally weighted holdings happen to be highly correlated with one

another. This limitation is exactly what motivates several of the risk-based and factor-based measures introduced later in this section.

3.4.2 Effective Number of Constituents (ENC)

ENC asks essentially the same underlying question as entropy, on how concentrated the portfolio's capital allocation is, but expresses the answer in a way that is much easier to interpret in practice. Rather than producing an entropy score on an abstract scale, it converts the degree of concentration directly into a number of "equivalent" equally weighted assets.

$$ENC = \frac{1}{\sum_{i=1}^N w_i^2}$$

This formula is simply the inverse of the Herfindahl–Hirschman index, a measure originally developed in industrial economics to assess how concentrated a market is among a small number of firms, applied here to a portfolio instead of a market. The logic transfers directly: a portfolio invested entirely in one asset produces $ENC = 1$ while a portfolio split perfectly evenly across N assets produces $ENC = N$

The practical advantage of ENC over Entropy is interpretability. Because ENC and Entropy are built on very similar mathematical logic, the two measures tend to move together closely across portfolios and over time, a relationship that is confirmed empirically later in this study, but ENC is generally the easier of the two to communicate to a non-technical audience.

3.4.3 Diversification Ratio (DR)

Entropy and ENC share an important limitation: both look only at portfolio weights and ignore everything about how the underlying assets actually behave. Neither can distinguish between a portfolio spread evenly across eleven genuinely different asset classes and one spread evenly

across eleven assets that all move together almost perfectly. The Diversification Ratio, introduced by Choueifaty and Coignard (2008), was designed specifically to close this gap by bringing risk, rather than just capital, into the picture.

The intuition behind DR is to compare two different notions of portfolio volatility. The first is the volatility an investor would expect if the assets in the portfolio simply added up their risks independently, with no diversification benefit at all — this is the weighted average of the individual asset volatilities. The second is the volatility the portfolio actually experiences once the real correlation structure between the assets is taken into account. The ratio between these two quantities is:

$$DR(w) = \frac{\sum_{i=1}^N w_i \sigma_i}{\sigma_p}, \quad \sigma_p = \sqrt{w^\top \Sigma w}$$

Because the actual, correlation-adjusted volatility of a portfolio is always lower than or equal to the simple weighted sum of its constituents' volatilities, DR can never fall below 1. A DR of exactly 1 corresponds to the case where all the assets in the portfolio are perfectly correlated with one another, combining them provides no risk-reduction benefit whatsoever, since they all move identically. As the assets held become progressively less correlated or, at the limit, negatively correlated the gap between the two volatility measures widens, and DR rises above 1 accordingly. In this sense, DR captures something closer to what most investors intuitively mean when they talk about diversification actually "working", not just holding many things, but holding things that do not all fall at once.

3.4.4 Effective Number of Bets Based on Principal Component Analysis (ENB PCA)

A portfolio can appear well diversified according to every weight based measure discussed so far, and still be driven almost entirely by a single underlying source of risk. A simple example illustrates the problem well, a portfolio split evenly across ten different stocks looks perfectly diversified by weight, yet if all ten stocks are simply different ways of expressing exposure to the same broad equity market, the portfolio's risk is, in an economic sense, coming from just one place. Meucci (2009) proposed addressing this by decomposing the covariance matrix into a set of orthogonal, mutually uncorrelated components, principal components and then asking how the portfolio's risk, rather than its capital, is distributed across those components.

The share of total portfolio risk explained by principal component k is calculated as:

$$p_k = \frac{(E^T w)_k^2 \lambda_k}{w^T \Sigma w}$$

where E is the matrix of eigenvectors of the covariance matrix and λ_k is the eigenvalue associated with component k . The effective number of bets then applies exactly the same exponential entropy logic used for portfolio weights in Section 3.4.1, but this time to the distribution of factor risk rather than the distribution of capital:

$$ENB_{PCA} = \exp \left(- \sum_{k=1}^N p_k \ln(p_k) \right)$$

The scale of interpretation is identical to before, a value of 1 means all portfolio risk is explained by a single principal component, and a value approaching N means risk is spread evenly across all of them, but what is being measured underneath is fundamentally different from anything discussed so far. Because of this, it is entirely possible, and in fact common, for a portfolio to score

close to its maximum on Entropy while scoring close to its minimum on ENB PCA at the same time. Far from being a contradiction, this gap is itself informative: it tells the analyst that a portfolio may be diversified in name, in the sense of holding many differently weighted assets, without being diversified in substance, in the sense of drawing its risk from genuinely independent sources.

3.4.5 Effective Number of Bets Based on Minimum Torsion (ENB MT)

Principal components solve the correlation problem highlighted above, but they introduce a new one of their own: they are purely mathematical constructs, built to maximise explained variance, with no guarantee of corresponding to anything an investor would recognise. Telling someone that "forty percent of the portfolio's risk comes from principal component two" is technically correct but practically meaningless, since component two has no obvious economic identity. Meucci, Santangelo, and Deguest (2015) proposed the Minimum Torsion transformation as a response to exactly this problem: it also produces a set of mutually uncorrelated factors, satisfying the same mathematical requirement as PCA, but it is constructed so as to distort the original structure of the assets as little as possible in the process, so that the resulting factors remain reasonably close to the assets an investor actually holds, rather than being an abstract statistical rotation of them.

Under this transformation, portfolio exposures to the transformed factors are first obtained as:

$$b = t^{-\top} w$$

where t denotes the Minimum Torsion transformation matrix. The proportion of total portfolio risk explained by factor i is then:

$$p_i = \frac{b_i^2 (t \Sigma t^\top)_{ii}}{w^\top \Sigma w}$$

and, exactly as with ENB PCA, the Effective Number of Bets is obtained by applying the entropy transformation to this risk distribution:

$$ENB_{MT} = \exp\left(-\sum_{i=1}^N p_i \ln(p_i)\right)$$

The scale runs from 1 to N in the same way as before. What makes ENB MT genuinely useful alongside ENB PCA, rather than redundant, is that the two measures can and often do produce quite different assessments of the same portfolio, precisely because they decompose risk along different sets of factors. A portfolio judged poorly diversified under the PCA based decomposition may be judged considerably better diversified once the Minimum Torsion decomposition is applied instead, since the latter is less prone to collapsing several genuinely different assets into a single dominant statistical component. This disagreement, examined empirically later in this study, is itself an important finding, it shows that even among measures built on the same underlying "effective number of bets" concept, the specific choice of orthogonalisation method can materially change the conclusion reached about how diversified a portfolio really is.

3.4.6 Risk Contribution Entropy (RC Entropy)

Weight based measures such as Entropy and ENC implicitly assume that allocating capital equally is the same thing as diversifying risk equally. This assumption does not generally hold. If one asset in an otherwise equally weighted portfolio happens to be considerably more volatile than the rest, it will end up contributing a disproportionate share of total portfolio risk, even though its capital weight looks no different from anyone else's. Maillard, Roncalli, and Teiletche (2010) working

from the same underlying logic that motivates risk-parity portfolio construction proposed shifting the focus away from how capital is allocated and toward how risk itself is allocated instead.

The first step is to calculate how much each asset actually contributes to total portfolio variance.

The normalised risk contribution of asset i is defined as:

$$RC_i = \frac{w_i(\Sigma w)_i}{w^\top \Sigma w}$$

Once these risk contributions are obtained, the same entropy logic used throughout this section is applied to their distribution rather than to the raw weights:

$$RC_Entropy = \exp\left(-\sum_{i=1}^N RC_i \ln(RC_i)\right)$$

As with the other entropy based measures, the scale runs from 1, where a single asset accounts for essentially all portfolio risk, to N , where risk is shared evenly across every holding. Because RC Entropy incorporates both the portfolio weights and the full covariance structure of the assets, it sits conceptually between the purely weight based measures introduced earlier and the purely risk based measures such as DR, it keeps the intuitive entropy framework, but grounds it in something economically more relevant than capital allocation alone.

3.4.7 Portfolio Diversification Index (PDI weighted)

Where the ENB measures ask how many independent bets a portfolio effectively holds, the Portfolio Diversification Index asks a closely related but distinct question: how much of total portfolio variance is spread across how many of the underlying principal components. Following Rudin and Morgan (2006), the relative importance of principal component i is first calculated as:

$$RS_i = \frac{\lambda_i}{\sum \lambda}$$

where λ_i is the eigenvalue associated with component i , and the index itself is then computed as:

$$PDI_{weighted} = 2 \sum (i \cdot RS_i) - 1$$

The construction of this formula is designed so that it explicitly rewards a portfolio whose variance is explained by later, smaller components rather than concentrated entirely in the first one. The two extreme cases are again a useful way to build intuition: if a single component explains essentially all of the portfolio's variance, then *PDI* collapses toward its minimum value of 1, representing a portfolio dominated by one source of systematic risk and effectively undiversified in factor terms. If, instead, variance is distributed evenly across all N components, *PDI* rises toward its maximum value of N , representing full diversification across every available factor. Because *PDI* is calibrated directly against the number of assets in the investment universe, in much the same way *ENC* is calibrated against Entropy, it offers a more immediately interpretable companion to *ENB PCA* both are built on the same principal component decomposition, but *PDI* is somewhat more sensitive to how variance is spread across the *later*, smaller components specifically, rather than treating all components symmetrically.

3.4.8 Rao Neff Diversification Measure

Rather than decomposing the covariance matrix into orthogonal factors at all, the Rao–Neff measure works directly with the correlation matrix, and asks a much more direct question: given the portfolio's own weights, what is the average correlation between the assets it actually holds?

$$Rao_Neff = \frac{1}{w^T Corr w}$$

The intuition here is perhaps the most straightforward of any measure in this study. If a portfolio is invested entirely in a single asset, its implied average correlation with itself is, by definition, exactly 1, and the measure falls to its minimum value of 1 as well. As the assets actually held in the portfolio become progressively less correlated with one another whether because they represent different asset classes, different geographies, or simply different risk exposures the weighted average correlation term in the denominator falls, and Rao Neff rises correspondingly. Because the measure has no fixed upper bound and depends entirely on how negative the achievable correlations are within the specific investment universe under consideration, it should be read as a relative rather than an absolute gauge of diversification: it is most informative when comparing portfolios built from the same underlying set of assets, exactly as is done throughout this study.

3.4.9 Diversification Return

Booth and Fama (1992) made an observation that, while well known among practitioners, is easy to overlook in a purely static analysis of portfolio weights, simply rebalancing a portfolio back to fixed target weights, in the presence of assets whose prices do not move in lockstep, generates a return premium on its own, entirely separate from whatever expected return any individual asset might offer. This happens because rebalancing systematically sells assets that have risen in price and buys assets that have fallen, which for imperfectly correlated assets tends to add a small but persistent boost to compounded returns over time. Diversification Return is designed to capture exactly this rebalancing premium:

$$DivReturn = 0.5(w^\top \sigma^2 - w^\top \Sigma w)$$

subsequently annualised. This measure differs from every other one presented in this section in an important way: it is not a diversification score bounded between 1 and N , but a realised return effect, expressed as a percentage, that grows larger as the gap widens between the weighted average variance of the individual assets and the actual variance experienced by the combined portfolio. Because it captures a performance outcome rather than a structural property of the portfolio's construction, Diversification Return should be interpreted somewhat differently from the other eleven measures in this study – a point that becomes especially clear in the empirical analysis of Chapter 4, where this measure behaves quite distinctly from the rest

3.4.10 Diversification Delta (DD)

Diversification Delta, proposed by Salazar Flores et al. (2017), takes a slightly different angle again: rather than measuring the entropy of the portfolio in isolation, it compares the entropy of the portfolio's individual constituents, taken separately, with the entropy of the combined portfolio itself.

$$DD^* = \frac{\sum w_i \exp(H_i) - \exp(H_p)}{\sum w_i \exp(H_i)}$$

The numerator captures the difference between how much uncertainty the individual assets contain on their own, weighted by their portfolio weights, and how much uncertainty the portfolio contains once they are actually combined together. Because the measure is normalised by the weighted individual-asset entropy in the denominator, its values sit close to the 0–1 range in practice: a value near 0 indicates that combining the assets into a portfolio has added almost nothing beyond what each asset already offered on its own, while a higher value indicates that the act of combining them

generated meaningful diversification benefits that would not have been available by holding any of the assets individually. In this sense, DD* is specifically designed to isolate the contribution of *portfolio construction itself* to diversification, separating it from whatever diversification properties the individual assets already happened to possess before being combined.

3.4.11 CVaR Diversification Ratio (CvaR DR)

The Diversification Ratio presented in Section 3.4.3 is built entirely on standard deviation, which implicitly treats gains and losses as symmetric and equally likely. As the descriptive statistics in Chapter 4 make clear, this assumption does not hold well for the assets used in this study: most exhibit negative skewness and substantial excess kurtosis, meaning that large negative returns occur more often, and are more extreme, than a symmetric distribution would suggest. A diversification measure built purely on volatility can therefore understate how much risk a portfolio is really exposed to during periods of market stress. Cesarone et al. (2025) proposed addressing this directly by replacing volatility with Expected Shortfall (ES), a downside risk measure that focuses specifically on the average size of losses beyond a given threshold within the same ratio structure used for DR:

$$CVaR_DR = \frac{\sum w_i ES_\alpha(X_i)}{ES_\alpha(\text{portfolio})}$$

The interpretation closely mirrors that of the standard Diversification Ratio but is applied to the tail of the return distribution rather than to its centre. A ratio above 1 means that the portfolio's realised Expected Shortfall is smaller than the weighted average of its constituents' individual Expected Shortfalls, indicating that diversification is still providing a meaningful benefit specifically during the kind of extreme, tail-risk events that a purely volatility based measure like DR might not fully capture.

3.4.12 Diversification Quality Based on Expected Shortfall (DQ ES)

DQ ES, proposed by Han, Lin, and Wang (2023), goes a step further than CVaRDR by evaluating not just the *magnitude* of tail risk diversification, but its overall *quality* – a distinction the authors draw explicitly. Rather than simply comparing a weighted average of individual tail losses to the portfolio's actual tail loss through a ratio, the measure is derived through an optimisation procedure:

$$DQ_{ES} = \frac{1}{\alpha} \min_r E \left[\left(r \left(L - \sum E S_{\alpha}(L_i) \right) + 1 \right)_+ \right]$$

Here, the procedure searches over the parameter r to find the value that best characterises how effectively diversification reduces the portfolio's losses within the tail of the distribution, given the joint behaviour of all the assets involved during periods of extreme stress, rather than treating each asset's contribution to tail risk in isolation. Because it is built specifically around this joint, optimisation based perspective, DQ ES is intended to complement, rather than replace, CvaR DR: where CvaR DR offers a direct and easily interpretable ratio, DQ ES offers a more comprehensive, if less immediately intuitive, assessment of how well-designed a portfolio's diversification really is once extreme, correlated losses are taken into account.

CHAPTER 4

Empirical Results and Analysis

4.1 Descriptive Statistics of the Selected ETFs

Table 4.1 reports the descriptive statistics of the eleven ETFs used in this study, calculated over the full 834-week sample period. These figures give a first sense of how differently the assets behave before any portfolio construction takes place, and they underpin everything that follows: the optimisation models in Section 3.3 and the diversification measures in Section 3.4 are both built directly from these return, volatility, and risk characteristics.

Table 4.1: Descriptive statistics of the eleven ETFs

Asset	Ann. return	Ann. vol.	Ann. Variance (%)	Sharpe	VaR 5%	VaR 1%	CVaR 5%	Skew	Ex. kurt.
SPY	14.26	16.24	2.64	0.79	3.35	5.88	5.28	-0.52	5.18
QQQ	18.82	19.12	3.66	0.91	3.83	7.18	5.80	-0.25	1.59
EFA	7.73	17.35	3.01	0.36	3.59	6.43	5.64	-0.42	3.98
EEM	5.54	19.63	3.85	0.21	4.22	6.82	5.99	-0.24	1.64
IEF	2.90	6.38	0.41	0.23	1.40	2.32	1.97	-0.18	0.90
TLT	3.71	14.13	2.00	0.16	3.21	4.73	4.30	-0.10	1.07
LQD	4.29	8.67	0.75	0.33	1.56	2.48	2.53	0.39	49.48
HYG	5.50	9.16	0.84	0.44	1.68	3.50	2.85	0.11	28.30
VNQ	10.36	20.75	4.30	0.43	4.11	6.98	6.48	-0.34	13.11
GLD	9.08	15.23	2.32	0.50	3.30	5.50	4.61	-0.20	1.38
DBC	1.72	17.11	2.93	0.02	3.88	6.11	5.32	-0.08	2.80

QQQ produced the highest return in the sample, 18.82% a year, with SPY not far behind at 14.26%, both benefiting from the broader strength of U.S. equities over the sample period. At the other end, DBC returned only 1.72% annually, and the two government bond funds, IEF and TLT, were similarly modest at 2.90% and 3.71% unsurprising given their defensive role in the universe. Bonds like these are not held to make money, they are held to hold steady when everything else does not.

Volatility and variance are really the same story told on two different scales In volatility terms, VNQ was the most volatile ETF in the sample (20.75%), followed by EEM (19.63%) and QQQ (19.12%), with IEF by far the calmest at 6.38%. Switch to variance and the ranking does not change, but the distance between assets does, VNQ's 4.30 against IEF's 0.41 is a tenfold gap, not just a wide one. QQQ 3.66 and EEM 3.85 sit close to each other on both scales, unsurprising given they share much of the same growth and emerging market exposure. DBC is worth a closer look for the opposite reason, it has the weakest return in the table (1.72%) but still carries a sizeable variance (2.93). In other words, whoever held DBC took on real risk and got almost nothing for it. That gap between risk and reward is a good reminder that a low return does not mean low risk, the two have to be checked separately, not assumed to move together.

That same pairing shows up directly in the Sharpe ratios. QQQ comes out on top (0.91), SPY close behind (0.79), both rewarded investors properly for the risk carried. DBC trails by a wide margin (0.02), which is really just the same return and variance story restated as a single number, plenty of risk, almost nothing to show for it.

The downside risk columns point in the same direction as volatility: equity and real estate exposures carry noticeably larger weekly VaR and CVaR than the bond funds, with IEF again the most contained. This is one of the reasons the study includes CvaR based diversification measures alongside the more conventional variance based ones in Section 3.4, a portfolio that looks

diversified under variance can still be exposed to outsized losses if tail risk is not accounted for separately.

Skewness and kurtosis add a further layer. Most of the ETFs are negatively skewed, meaning sharp down weeks are somewhat more common than sharp up weeks, and every asset shows positive excess kurtosis. LQD (49.48) and HYG (28.30) stand out here, a reminder that even ostensibly conservative credit instruments can carry fat tailed risk that a normal distribution assumption would miss entirely. VNQ (13.11) is not far behind. None of this is unusual for weekly financial return data, but it does mean that variance alone is an incomplete description of risk for this universe, which again motivates the inclusion of tail-risk measures later in the analysis.

Taken as a whole, these eleven assets differ enough in return, volatility, and distributional shape to give the portfolio construction methods in Section 3.3 something meaningful to work with, there would be little point discussing diversification if the underlying assets all behaved alike.

4.2 Correlation Structure

What actually drives diversification is not the individual numbers in Table 4.1 but how these assets move relative to each other. Two assets can each look perfectly reasonable on their own and still add nothing to a portfolio if they rise and fall together. Table 4.2 gives the correlation matrix for the eleven ETFs

Table 4.2: Correlation matrix of weekly ETF returns

	SPY	QQQ	EFA	EEM	IEF	TLT	LQD	HYG	VNQ	GLD	DBC
SPY	1.00	0.92	0.86	0.74	-0.20	-0.23	0.35	0.76	0.74	0.13	0.43
QQQ	0.92	1.00	0.76	0.69	-0.15	-0.17	0.29	0.65	0.60	0.11	0.34
EFA	0.86	0.76	1.00	0.83	-0.18	-0.22	0.34	0.73	0.68	0.20	0.45
EEM	0.74	0.69	0.83	1.00	-0.15	-0.19	0.28	0.63	0.56	0.26	0.49
IEF	-0.20	-0.15	-0.18	-0.15	1.00	0.91	0.59	0.02	0.10	0.27	-0.20
TLT	-0.23	-0.17	-0.22	-0.19	0.91	1.00	0.59	-0.03	0.05	0.21	-0.26
LQD	0.35	0.29	0.34	0.28	0.59	0.59	1.00	0.65	0.56	0.29	0.11
HYG	0.76	0.65	0.73	0.63	0.02	-0.03	0.65	1.00	0.76	0.21	0.40
VNQ	0.74	0.60	0.68	0.56	0.10	0.05	0.56	0.76	1.00	0.23	0.32
GLD	0.13	0.11	0.20	0.26	0.27	0.21	0.29	0.21	0.23	1.00	0.34
DBC	0.43	0.34	0.45	0.49	-0.20	-0.26	0.11	0.40	0.32	0.34	1.00

The strongest pair in the table is SPY and QQQ (0.92), with IEF and TLT close behind (0.91).

Both make sense as the two highest, one pair is essentially two versions of the same U.S. equity exposure, the other two maturities of the same government bond, so neither buys much diversification on its own, holding both is closer to holding one asset twice than to holding two genuinely different things. At the other end, the most negative entry is TLT and DBC (-0.26), followed by SPY and TLT (-0.23), IEF and DBC (-0.20). Long duration treasuries are the one holding in this universe that consistently moves against both equities and commodities, exactly the property that makes an asset useful for diversification, rather than merely correlated in name.

Some pairings behave differently from what the asset labels imply. HYG correlates with SPY at 0.76, identical to its correlation with VNQ, despite being nominally a bond fund, while its correlation with IEF (0.02) and TLT (-0.03) is essentially zero. Whatever HYG is doing, it is not behaving like a government bond, credit spread risk, not interest rate risk, is clearly what drives it. LQD sits in a more genuinely intermediate spot, 0.59 with both treasuries, but also 0.65 with HYG

and 0.28–0.35 with the equity block. It is the one asset here without an obvious home on either the 'bond' or 'equity' side of the table.

GLD's row is the flattest of any asset, nothing above 0.34 and even its highest correlation, with DBC, is not with an equity or bond position at all. That is about as clean an empirical confirmation as one could want of gold's usual reputation as a self-contained, hard to replace diversifier. DBC cuts the other way, positively correlated with every equity ETF (0.34 to 0.49, strongest with EEM, which fits how closely emerging market growth tracks commodity demand), while at the same time holding the two most negative correlations in the whole table, against IEF and TLT. Pulled in two directions at once, it's part of why DBC behaved the way it did back in Table 4.1, real variance, poor return.

Put together, the table points to three fairly tight clusters: equities, long term treasuries, and, more loosely, commodities. LQD and HYG sit between these clusters rather than comfortably inside one, and GLD stands apart from all of them

4.2.1 Principal Component Structure

To look at this heterogeneity more formally, a principal component analysis was run on the covariance matrix, averaged across all 167 rolling estimation windows. The first three components together explain 81.2% of total variance, so despite holding eleven assets, the effective number of independent risk sources in this universe is considerably smaller than eleven.

Table 4.2.1: Principal component loadings and variance explained, averaged across all rolling windows

Asset	F1	F2	F3	F4	F5	F6
SPY	39.45	-8.84	-17.24	2.10	26.74	9.04
QQQ	42.31	-11.69	-26.96	-28.14	63.33	10.30
EFA	40.74	-8.87	-2.49	-13.02	-24.51	-0.99
EEM	42.99	-14.00	17.18	-49.27	-51.77	-23.48
IEF	-1.34	28.84	0.43	-10.30	6.09	-15.32
TLT	-5.10	65.68	-3.63	-27.42	17.39	-44.11
LQD	11.11	33.29	-2.91	1.86	1.08	-19.24
HYG	19.86	8.84	-3.45	11.61	-3.27	-6.88
VNQ	44.19	37.68	-21.47	61.29	-25.73	14.85
GLD	10.46	37.73	57.95	-21.37	6.37	67.64
DBC	23.18	-18.33	69.44	37.38	30.64	-43.20
Var. explained (%)	56.59	13.81	10.81	6.02	4.39	4.25
Cumulative (%)	56.59	70.40	81.21	87.23	91.62	95.87

The first factor alone explains 56.6% of variance and loads heavily and positively on every equity and real-estate ETF (SPY, QQQ, EFA, EEM, VNQ, each between 39% and 44%), while the government bonds load close to zero. This is best read as a general equity/growth risk factor. The second factor (13.8%) is dominated by TLT (65.7%), with positive loadings on LQD, VNQ and GLD and negative loadings on EEM and DBC, an interest-rate or duration factor, in other words. The third (10.8%) loads on DBC (69.4%) and GLD (57.9%) and is best interpreted as a commodity/inflation factor. Factors four through six, together explaining under 15% of variance, are progressively harder to interpret economically and mix residual REIT specific risk with a U.S. versus international equity tilt and idiosyncratic gold movements. This factor structure matters

directly for how the ENB_PCA and PDI weighted measures in Section 4.5 should be read. Because four of the eleven assets in the Equal Weight portfolio load heavily on the same dominant factor, that portfolio ends up with comparatively low factor-based diversification despite being maximally diversified by weight. Portfolios that instead lean toward government bonds, corporate credit, and commodities GMV Capped and Mean CVaR in particular spread their risk more evenly across F1, F2 and F3, and score correspondingly higher on both measures. Parmentier (2018) reports a broadly similar pattern in a five-asset-class universe, where a single equity/real-estate factor also accounts for roughly half of total variance.

4.3 Portfolio Composition

Table 4.3: Average portfolio weights across all monthly rebalancing periods, 2012–2025

Asset	Equal Weight	GMV capped	Tangency	Mean CVaR	MV Frontier	Max DR
SPY	9.1	7.7	13.3	7.0	10.4	11.7
QQQ	9.1	0.0	24.9	2.9	33.6	7.3
EFA	9.1	0.6	0.0	0.1	0.0	3.2
EEM	9.1	0.4	0.0	0.5	0.6	2.6
IEF	9.1	25.0	22.2	58.5	7.9	28.7
TLT	9.1	7.0	9.1	0.5	16.8	21.5
LQD	9.1	21.2	5.0	1.6	0.5	0.0
HYG	9.1	21.8	1.6	19.8	8.1	0.7
VNQ	9.1	0.0	0.9	0.0	2.7	0.6
GLD	9.1	6.4	10.5	4.3	13.4	7.4
DBC	9.1	9.9	12.6	4.8	5.9	16.3

Each construction method produces a clearly distinct allocation, which is really the point of including several of them, they are meant to stress-test the diversification measures against genuinely different portfolio structures, not just different numbers attached to the same underlying idea.

The Equal Weight portfolio is, by construction, the simplest case 9.1% in every asset, every month, with no estimation involved at all.

GMV Capped, which caps any single position at 25%, ends up concentrated mainly in fixed income: 25.0% in IEF, 21.8% in HYG, and 21.2% in LQD, with smaller positions in DBC, SPY, TLT and GLD. This is the minimum-variance logic playing out under a constraint, capital still flows toward the lowest-volatility assets, but the ceiling forces the excess into the next-cheapest options rather than letting it pile into a single bond position.

Tangency looks quite different: it leans into QQQ (24.9%) and IEF (22.2%), with meaningful stakes in SPY (13.3%), DBC (12.6%) and GLD (10.5%). This is what maximising the Sharpe ratio tends to produce here, a mix of the highest-return equity exposure with enough low-volatility ballast to keep the ratio favourable.

Mean CVaR ends up looking a lot like GMV Capped in spirit, if not in exact composition, 58.5% in IEF and 19.8% in HYG, with equities pushed to the margins. It arrives at a conservative, bond-heavy portfolio from a downside-risk angle rather than a variance angle, but without an explicit ceiling on how far it can concentrate.

MV Frontier spreads its risk budget more deliberately, QQQ still takes the largest single slice (33.6%), but TLT (16.8%), GLD (13.4%), SPY (10.4%) and HYG (8.1%) all get meaningful

allocations too, reflecting the explicit trade-off this method is built to strike between expected return and a fixed volatility target.

Max DR, finally, produces the most evenly spread allocation of any of the optimised portfolios IEF (28.7%), TLT (21.5%), DBC (16.3%), SPY (11.7%), GLD (7.4%) and QQQ (7.3%) all carry weight. That is a direct consequence of what this method is built to do: rather than chasing the lowest volatility assets, it searches for the combination that benefits most from imperfect correlations, which naturally pulls in exposure across several different asset classes at once.

Stepping back, the risk-minimising methods (GMV Capped and Mean CVaR) both gravitate toward concentrated fixed income positions, while Tangency, MV Frontier and Max DR spread capital more broadly across asset classes, each chasing a different objective, but none of them arriving at anything close to the Equal Weight allocation.

4.4 Portfolio Performance

Table 4.4: Out-of-sample performance, February 2012 – December 2025

Portfolio	Total return	CAGR	Ann. vol.	Max drawdown	% pos. weeks	Sharpe
MV Frontier	277.1	10.0	11.1	-17.0	58.5	0.77
Equal Weight	151.4	6.8	9.5	-20.8	57.4	0.57
Tangency	188.9	7.9	11.3	-22.6	56.9	0.59
Mean-CVaR	65.8	3.7	5.6	-16.8	58.8	0.39
GMV-Capped	82.7	4.4	6.7	-16.1	58.5	0.44
Max-DR	88.4	4.6	7.3	-23.7	57.2	0.44

The performance gap between the return-seeking and the risk minimising portfolios is large. MV Frontier comes out on top by a clear margin, the highest total return (277.1%), the highest CAGR (10.0%), and the highest Sharpe ratio (0.77) in the sample. Its volatility (11.1%) is among the higher end, but its maximum drawdown (−17.0%) is no worse than the more conservative portfolios, which suggests that targeting an efficient risk-return trade-off paid off here without adding proportionally more downside pain.

Tangency also performed well, a 188.9% cumulative return and 7.9% CAGR, with the second best Sharpe ratio (0.59). The cost was a considerably deeper drawdown (−22.6%), reflecting its heavier tilt toward equities. Equal Weight, meanwhile, achieved a Sharpe ratio of 0.57, barely behind Tangency, without estimating a single expected return or covariance. That is a useful reminder that naive 1/N allocation remains a genuinely competitive benchmark, not just a theoretical curiosity.

GMV Capped and Mean CVaR sit at the other end of the spectrum. Both concentrate in low risk fixed income and, as a result, post the lowest volatilities in the table (6.7% and 5.6%) and comparatively shallow drawdowns (−16.1% and −16.8%). The price is return: CAGR ranges only from 3.7% to 4.4%, and Sharpe ratios trail well behind the return-oriented portfolios. It is worth noting that an earlier, unconstrained implementation of the minimum-variance model before the 25% cap was applied produced an even lower CAGR of 2.8%; capping single-asset exposure at 25% nearly doubled the resulting Sharpe ratio, from 0.24 to 0.44, while leaving the defensive character of the strategy largely intact. That improvement is the reason the constrained version was kept as the sole minimum-variance representative in this study rather than the uncapped original.

Max DR falls somewhere in between. It posts a CAGR of 4.6%, ahead of both GMV Capped and Mean CVaR with moderate volatility (7.3%), but its maximum drawdown (−23.7%) is actually the

worst in the sample, despite the portfolio being built specifically to maximise the Diversification Ratio. That is a useful finding in its own right: maximising a diversification measure clearly does not guarantee better downside protection, and the two objectives should not be treated as interchangeable.

Put together, these results point to the optimisation objective as the main driver of realised outcomes here. Return-seeking objectives MV Frontier and Tangency in particular delivered the strongest riskadjusted performance, while pure risk minimisation bought stability at a real cost in growth. GMV Capped and Max DR each sit somewhere in between, showing that constraints and alternative objectives can soften, without eliminating, that basic tradeoff.

4.5 Diversification Measures

Table 4.5: Average diversification scores across all six portfolios.

Portfolio	Entropy	ENC	DR	ENB PCA	ENB MT	RC Ent.	PDI	Rao Neff	DivRet %	DivDelta	CVaR DR	DQ ES
Equal Weight	11.00	11.00	1.56	1.55	9.30	9.03	3.19	2.69	0.72	0.33	1.54	0.84
GMV Capped	5.03	4.65	1.74	3.98	8.56	5.19	2.45	2.49	0.39	0.36	1.79	0.88
Tangency	2.45	2.21	1.60	3.51	5.52	2.16	1.44	2.32	0.52	0.29	1.71	0.79
Mean CVaR	2.62	2.20	1.69	4.51	6.81	2.70	1.70	2.18	0.28	0.33	1.91	0.88
MV Frontier	2.57	2.36	1.42	3.08	5.06	1.93	1.42	2.17	0.62	0.26	1.48	0.81
Max DR	4.16	3.47	2.03	3.99	9.13	4.66	2.22	3.70	0.75	0.45	2.10	0.93

What stands out first in Table 4.5 is how much the ranking of portfolios changes depending on which measure is used, which is really the central point of this whole study. Equal Weight tops every purely weight based measure, as it should: Entropy (11.00), ENC (11.00), ENB MT (9.30),

RC Entropy (9.03), PDI (3.19) and Rao Neff (2.69) are all highest for this portfolio, simply because spreading capital evenly across eleven assets is what these measures are designed to reward.

Switch to a risk based view, though, and the picture changes substantially. Max DR, which never appears among the weight based leaders, takes the top spot on the Diversification Ratio (2.03), CVaR DR (2.10), Diversification Return (0.75), Diversification Delta (0.45), and DQ ES (0.93). That it leads on five conceptually different measures, having directly optimised only one of them, is a fairly strong signal that these risk-based measures are picking up on something real and consistent about how this portfolio spreads its risk, not just an artefact of one particular formula.

GMV Capped and Mean CVaR, the two riskminimising portfolios in the sample, tell a more nuanced story. Because GMV-Capped cannot put more than 25% into any single asset, it scores considerably higher on the weight-based measures, Entropy of 5.03 versus 2.62 for Mean CVaR, ENC of 4.65 versus 2.20, while the two portfolios land in a similar range on the covariance and tail risk based measures (ENB PCA, CVaR DR, DQ ES). In other words, the weight cap buys capital allocation diversification without changing much about the underlying risk profile the two strategies share.

Tangency and MV Frontier occupy the middle ground on almost everything, lower than Equal Weight on the weight based side, lower than Max DR on the riskbased side, yet these are exactly the two portfolios that delivered the strongest returns in Table 4.4. MV Frontier's Diversification Ratio (1.42) is in fact the lowest in the whole sample, which makes the point plainly, strong investment performance clearly does not require topping any particular diversification measure.

The overall pattern in this table is hard to miss. Weight based, risk based, factor based and downside measures repeatedly disagree, sometimes sharply, about which portfolio is 'most

diversified. That disagreement is not noise, it reflects the fact that these measures are genuinely capturing different things, and no single number in this table can stand in for the full picture on its own.

4.6 Stability Analysis

It is one thing for a diversification measure to give a sensible reading at a single point in time and it is another for that reading to stay reasonably consistent as markets move. A measure that swings wildly from month to month is of limited use for monitoring or comparison purposes, however sensible its underlying logic. Table 4.6 looks at this using two statistics computed across all 167 rolling windows and the six retained portfolios: the coefficient of variation (CV), and the first order autocorrelation, which captures how persistent a measure's value is from one rebalancing date to the next.

Table 4.6: Temporal stability of the twelve diversification measures, six retained portfolios.

Measure	CV	Autocorr. (lag 1)
ENB MT	0.175	0.923
DR	0.180	0.956
Entropy	0.190	0.801
PDI	0.192	0.859
ENC	0.205	0.795
DQ ES	0.223	0.969
RC Entropy	0.225	0.850
CVaR DR	0.233	0.926
ENB PCA	0.240	0.885
DivDelta	0.265	0.940

Measure	CV	Autocorr. (lag 1)
Rao Neff	0.335	0.935
DivReturn %	0.469	0.927

ENB MT comes out as the most stable measure by a reasonable margin ($CV = 0.175$), with DR (0.180) close behind and Entropy and PDI not far off (0.190, 0.192). All four keep a fairly steady reading regardless of what the market is doing in a given month, which makes them dependable for tracking diversification over time rather than just at a single snapshot.

A second tier, ENC (0.205), DQ ES (0.223), RC Entropy (0.225), CVaR DR (0.233) and ENB PCA (0.240), is somewhat noisier, but still reasonably well behaved.

DivDelta (0.265), Rao Neff (0.335) and especially Diversification Return (0.469) move around considerably more from window to window, DivReturn's CV is roughly two and a half times that of ENB MT. That extra sensitivity could be useful for spotting a genuine regime change quickly, but it also means these three measures are less reliable as a stable, ongoing gauge of diversification.

Autocorrelation adds a complementary angle: every one of the twelve measures shows fairly high persistence, from 0.795 (ENC) up to 0.969 (DQ ES), meaning diversification tends to drift gradually rather than jump around abruptly between rebalancing dates even the least persistent measure here still carries a fair amount of memory from month to month.

Together, ENB MT, DR, Entropy and PDI look like the more dependable choice for tracking diversification over an extended period, while Diversification Return and Rao Neff, despite their higher noise, might still earn their place where the goal is to catch structural shifts quickly rather than to monitor a stable baseline.

4.7 Measure Correlation

The stability analysis looked at each measure on its own, over time. This section asks a different question: how much do the twelve measures agree with each other, at any given point, about which portfolios are more or less diversified?

A Spearman rank correlation across all rolling observations answers this directly, from which high positive correlation means two measures broadly agree on portfolio rankings, while low or negative correlation means they are picking up on genuinely different things. To isolate this from each portfolio's own average level on a given measure, each measure's values are first demeaned within each portfolio before pooling.

Table 4.7: Spearman rank correlation matrix of the twelve diversification measures, six retained portfolios.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Entropy	1.00	0.96	0.24	-0.03	0.25	0.73	0.49	0.37	0.10	0.23	0.19	0.11
(2) ENC	0.96	1.00	0.21	-0.05	0.25	0.67	0.50	0.38	0.14	0.20	0.15	0.11
(3) DR	0.24	0.21	1.00	0.61	0.04	0.19	0.09	0.85	0.21	0.98	0.89	-0.12
(4) ENB PCA	-0.03	-0.05	0.61	1.00	-0.09	0.10	0.01	0.45	0.12	0.63	0.55	-0.20
(5) ENB MT	0.25	0.25	0.04	-0.09	1.00	0.51	0.53	-0.06	-0.27	0.06	0.10	0.25
(6) RC Entropy	0.73	0.67	0.19	0.10	0.51	1.00	0.65	0.09	-0.13	0.19	0.22	0.17
(7) PDI	0.49	0.50	0.09	0.01	0.53	0.65	1.00	0.02	-0.33	0.12	0.07	0.41
(8) Rao Neff	0.37	0.38	0.85	0.45	-0.06	0.09	0.02	1.00	0.32	0.83	0.72	-0.08
(9) DivRet %	0.10	0.14	0.21	0.12	-0.27	-0.13	-0.33	0.32	1.00	0.22	0.20	-0.76
(10) DivDelta	0.23	0.20	0.98	0.63	0.06	0.19	0.12	0.83	0.22	1.00	0.87	-0.12
(11) CVaR DR	0.19	0.15	0.89	0.55	0.10	0.22	0.07	0.72	0.20	0.87	1.00	-0.12
(12) DQ ES	0.11	0.11	-0.12	-0.20	0.25	0.17	0.41	-0.08	-0.76	-0.12	-0.12	1.00

Two fairly distinct clusters emerge from this matrix, consistent with the theoretical grouping in Chapter 2. The first is the allocation based cluster: Entropy and ENC move almost in lockstep ($\rho = 0.96$), which makes sense given how similarly they are built, and RC Entropy joins them closely too ($\rho = 0.73$ with Entropy, 0.67 with ENC). PDI is a somewhat looser member of the same group ($\rho = 0.49\text{--}0.65$ with the others).

The second cluster is risk based, and here the relationships are even tighter than in the weightbased group. DR and DivDelta are the standout pair ($\rho = 0.98$), with DR and CVaR DR close behind (0.89) and DR and Rao Neff also strong (0.85). These four DR, DivDelta, CVaR DR and Rao Neff are, in practice, measuring something very close to the same underlying concept despite arriving at it through quite different routes.

The two ENB variants remain worth dwelling on, because they disagree more than their shared name would suggest. ENB PCA correlates moderately with the risk based cluster (0.61 with DR, 0.55 with CVaR DR), while ENB MT sits apart from both clusters, correlating instead with the weight based group (0.51 with RC Entropy, 0.53 with PDI) and is negatively correlated with ENB PCA itself ($\rho = -0.09$). Both measures are trying to count effective independent bets, but the choice between principal components and Minimum Torsion clearly changes which family the resulting measure ends up resembling.

Diversification Return stands apart from almost everything, its strongest relationship being a sharply negative one with DQ ES ($\rho = -0.76$) rather than with any of the other ten measures (all under 0.32 in absolute value). That fits the interpretation given in Section 3.4.9: DivReturn is measuring a realised rebalancing benefit, not a structural property of the portfolio, so there is no strong reason to expect it to track the other measures closely. DQ ES itself is comparatively muted against everything except DivReturn and, to a lesser extent, PDI (0.41), evidence that it is

contributing something the existing measures do not already cover, though its exact behaviour should be read with some caution given the formula uncertainty noted in Section 3.4.12.

The overall message is that these twelve measures are not twelve versions of the same idea. Weight-based measures agree closely with each other, risk-based measures agree closely with each other, but the two groups have little to say to one another.

4.8 Measure Agreement

A high pairwise correlation between two measures does not automatically mean they rank all six portfolios the same way once every measure is considered together. To check this more directly, Kendall's coefficient of concordance (W) and the Friedman test are used to assess how much consensus exists across all twelve measures simultaneously, rather than just two at a time.

Table 4.8: Summary of agreement statistics across the twelve diversification measures, six retained portfolios

Statistic	Result	Interpretation
Kendall's W (rolling)	mean 0.434 / median 0.458	0 = no agreement, 1 = perfect agreement across the twelve measures, $n = 6$ portfolios.
Kendall's W (avg. scorecard)	0.507	Moderate agreement when portfolios are ranked on the full-sample scorecard.
Friedman test	$p < 0.05$ in 100% of windows	Rankings across measures differ systematically, not by chance.
Hierarchical clustering	2 stable clusters	A weight-based and a risk-based family of measures, consistently.

The headline number is Kendall's W of 0.434 on average across rolling windows (median 0.458), meaningful agreement, but well short of the 1.0 that would indicate the twelve measures are all telling the same story. Using the full sample average scorecard instead of the rolling estimates pushes this up to 0.507, but the conclusion does not change, agreement is moderate at best..

The Friedman test backs this up formally. The hypothesis that all twelve measures rank the portfolios identically is rejected at the 5% level in every single rolling window in the sample. That is about as clear a signal as this kind of test can give that the disagreement documented throughout this chapter is systematic, not sampling noise.

It confirms that a weight-based family Entropy, ENC, ENB MT, RC Entropy, PDI, and, somewhat more loosely, DivReturn % and DQ ES, tends to favour portfolios with more evenly spread capital, and a risk-based family DR, ENB PCA, Rao Neff, DivDelta and CVaR DR, that instead rewards efficient risk allocation, with Max DR consistently coming out on top under this family.

None of this should be read as the measures being unreliable moderate, statistically significant disagreement is itself informative. It is evidence that weight-based and risk-based diversification are genuinely different constructs, and that a study relying on only one family would be answering a narrower question than it might realise.

4.8.1 Post-Hoc Analysis: Which Measures Differ Significantly? (Nemenyi Test)

The Friedman test establishes that disagreement exists; it does not say between which specific measures. A Nemenyi post-hoc test fills that gap. Blocks here are defined as (window, portfolio) pairs 167 windows x 6 portfolios = 1,002 blocks and within each block, every measure's value is converted into the rank (1–6) it assigns to that portfolio relative to the other five. The 66 possible pairs among the twelve measures are then compared on this common rank scale.

Table 4.8.1: Nemenyi post-hoc pairwise p-values, six retained portfolios

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Entropy	—	1.000	0.593	0.017	1.000	1.000	1.000	0.999	1.000	0.593	0.897	0.701
(2) ENC	1.000	—	0.813	0.052	1.000	0.993	1.000	1.000	0.999	0.813	0.979	0.462
(3) DR	0.593	0.813	—	0.954	0.669	0.125	0.278	0.987	0.203	1.000	1.000	0.002
(4) ENB PCA	0.017	0.052	0.954	—	0.025	0.001	0.003	0.240	0.002	0.954	0.724	0.000
(5) ENB MT	1.000	1.000	0.669	0.025	—	0.999	1.000	1.000	1.000	0.669	0.933	0.627
(6) RC Entropy	1.000	0.993	0.125	0.001	0.999	—	1.000	0.860	1.000	0.125	0.380	0.990
(7) PDI	1.000	1.000	0.278	0.003	1.000	1.000	—	0.968	1.000	0.278	0.627	0.933
(8) Rao Neff	0.999	1.000	0.987	0.240	1.000	0.860	0.968	—	0.935	0.987	1.000	0.140
(9) DivRet %	1.000	0.999	0.203	0.002	1.000	1.000	1.000	0.935	—	0.203	0.520	0.967
(10) DivDelta	0.593	0.813	1.000	0.954	0.669	0.125	0.278	0.987	0.203	—	1.000	0.002
(11) CVaR DR	0.897	0.979	1.000	0.724	0.933	0.380	0.627	1.000	0.520	1.000	—	0.015
(12) DQ ES	0.701	0.462	0.002	0.000	0.627	0.990	0.933	0.140	0.967	0.002	0.015	—

The differences that turn out significant are fairly concentrated. ENB PCA stands out the most, it differs from Entropy ($p = 0.017$), RC Entropy ($p = 0.001$), PDI ($p = 0.003$), ENB MT ($p = 0.025$) and DQ ES ($p < 0.001$). That's a mix spanning both the weight based group and its own ENB partner, which says something: principal components and Minimum Torsion clearly decompose risk in quite different ways.

DQ ES is the other measure worth noting. It differs from DR ($p = 0.002$), DivDelta ($p = 0.002$), CVaR DR ($p = 0.015$) and ENB PCA ($p < 0.001$) mostly the risk based cluster. That fits what Table 4.7 already showed: DQ ES sits closer to the weight based side here.

Diversification Return, by contrast, differs significantly only from ENB PCA ($p = 0.002$). That's a small number given how different its correlations in Table 4.7 looked from the other eleven measures. Most likely this comes down to statistical power: with just six portfolios to rank, the Nemenyi test doesn't have much room to separate measures whose behavior points in different directions but doesn't consistently produce different ranks across such a small set. In other words, the weight based / risk based split still holds, though exactly which pairs cross the 0.05 line depends partly on how few items, six portfolios there are to rank in the first place. Tests like this are known to be sensitive to that.

4.8.2 Predictive Ability of Diversification Measures

Everything so far has looked at how well the measures describe diversification at the moment it is calculated. A separate, arguably more practical question is whether a measure also carries information about risk one month ahead. Each of the twelve measures was regressed against realised portfolio volatility and the worst weekly return over the following month (using the six portfolios' actual subsequent out-of-sample returns), to check exactly that.

Table 4.8.2: Predictive regressions of future risk on current diversification measures, six retained portfolios.

Measure	β (fut. vol.)	t	p	R ²	β (fut. worst wk)	t	p	R ²
ENB MT	-0.0013	-10.98	0.000	0.108	0.0014	8.68	0.000	0.070
DivDelta	-0.0218	-10.48	0.000	0.099	0.0193	6.74	0.000	0.043
DR	-0.0054	-7.97	0.000	0.060	0.0043	4.68	0.000	0.022
CVaR DR	-0.0040	-7.72	0.000	0.056	0.0031	4.44	0.000	0.019
ENB PCA	-0.0014	-7.31	0.000	0.051	0.0013	4.78	0.000	0.022
Rao Neff	-0.0009	-4.08	0.000	0.016	0.0007	2.10	0.036	0.004

Measure	β (fut. vol.)	t	p	R ²	β (fut. worst wk)	t	p	R ²
PDI	-0.0012	-3.49	0.001	0.012	0.0011	2.42	0.016	0.006
RC Entropy	-0.0003	-2.84	0.005	0.008	0.0003	2.21	0.027	0.005
Entropy	-0.0001	-0.87	0.387	0.001	0.0001	0.89	0.374	0.001
ENC	-0.0000	-0.37	0.715	0.000	0.0001	0.55	0.585	0.000
DQ ES	0.0035	1.68	0.094	0.003	-0.0045	-1.60	0.110	0.003
DivReturn %	0.0019	2.20	0.028	0.005	-0.0022	-1.95	0.052	0.004

Eight of the twelve measures come out significant, with the theoretically expected sign, in both regressions. ENB MT is comfortably the strongest predictor of future volatility ($t = -10.98$), followed closely by DivDelta ($t = -10.48$) and between them accounting for around 10% of the variation in next month's realised volatility on their own. DR, CVaR DR and ENB PCA form a clear second tier (t between -7.3 and -8.0), and Rao Neff, PDI and RC Entropy remain significant but with more modest explanatory power.

Predicting the worst week ahead is harder across the board, as expected, but the same six measures ENB MT, DivDelta, DR, CVaR DR, ENB PCA, Rao Neff and PDI all stay significant at the 5% level, with ENB MT again the strongest ($t = 8.68$).

At the weak end, Entropy and ENC show essentially no predictive power in either direction, despite describing the current state of diversification perfectly well describing the present and predicting the future are evidently not the same skill. DivReturn % is marginally significant for future volatility ($p = 0.028$) but not for the worst week, and with the 'wrong' sign relative to the other measures (positive rather than negative), consistent with it capturing a different phenomenon (realised rebalancing return) rather than risk reduction as such. DQ ES is not significant in either

regression at the 5% level, though, as with the correlation results in Section 4.7, this measure's numbers should be read with some caution given the formula uncertainty flagged in Section 3.4.12.

The pattern here reinforces what the correlation and agreement analyses already suggested that the risk based measures, ENB MT, DivDelta, DR, CVaR DR and ENB PCA foremost among them, carry real forward looking value, while the classic weight based measures (Entropy and ENC in particular) are better suited to describing a portfolio's current structure than to anticipating what comes next.

4.9 Discussion

Bringing the chapter together, the results make one thing hard to avoid, diversification is not a single number, and the six portfolios examined here made that difficult to miss. Different construction methods produced genuinely different diversification profiles, and, more importantly, which method looked 'best' flipped depending entirely on which of the twelve measures was doing the judging. Equal Weight led on every weight based measure and lagged on several risk based ones; GMV Capped and Mean CVaR traded places depending on whether the weight cap or the underlying risk exposure mattered more for a given measure; MV Frontier and Tangency delivered the strongest returns while sitting in the middle of the diversification rankings on almost everything.

The stability, correlation, and agreement analyses add weight to the same conclusion from a different angle. ENB MT and DR held up well over time; the twelve measures, taken together, agreed only moderately with each other, and both the Friedman and Nemenyi tests confirmed that this disagreement is real rather than statistical noise. The predictive regressions then showed that

some of these measures, mostly the risk based ones, carry genuine information about future risk, while others are better read as descriptions of the present moment only.

None of these points toward one measure being 'correct' and the rest being redundant. It points instead toward a genuinely multidimensional concept: how capital is allocated and how risk is allocated are related but distinct ideas, and neither one, on its own, stands in for the whole. The two Effective Number of Bets measures made this especially clear, built on the same underlying logic, yet one tracking allocation diversification and the other tracking risk diversification once the data was actually examined, which is as good a reminder as any that a measure's theoretical construction does not always predict how it will behave empirically. Treating any single dimension as a stand in for the whole would have missed most of what this chapter actually found.

CHAPTER 5

Conclusions and Recommendations

5.1 Conclusions

This study set out to see how different diversification measures judge the same set of portfolios, using six construction methods built from the same eleven ETF universe under an identical rolling out-of-sample framework, and twelve measures spanning weight-based and risk-based approaches, the latter taking several more specific forms factor based, downside risk, and realized return, depending on which aspect of risk a given measure targets. The central finding is straightforward to state but has real implications: diversification is not a single, well defined quantity, and no one measure in this study can stand in for the whole picture.

Much of this comes down to what each measure is actually built to reward. Weight-based measures favoured portfolios that spread capital evenly, which is more or less by definition; risk-based measures instead rewarded efficient allocation of risk, regardless of how many assets that risk happened to be spread across. The two Effective Number of Bets measures illustrated just how far this split can go even within a single conceptual origin: both decompose portfolio risk into uncorrelated factors, yet ENB PCA tracked the risk-based family closely while ENB MT tracked the weight-based family instead a genuine empirical finding, not a modelling inconsistency, and one that says something about how sensitive factor-based diversification is to the specific decomposition method used. The GMV Capped portfolio illustrated a related point from the allocation side: capping any single position at 25% raised its diversification scores under several weight-based measures relative to an uncapped minimum-variance benchmark, without changing much about its risk-based profile, showing that these are genuinely separable dimensions, not two

views of the same underlying fact. Max DR, for its part, scored consistently high across most of the risk-based measures without producing the best realised returns, which is its own kind of evidence that a well-diversified portfolio and a well-performing one are not automatically the same thing.

The measures are also not interchangeable in a more technical sense. ENB MT, the Diversification Ratio, and PDI held up as the most stable over time, while Diversification Return and Rao Neff moved around considerably more from one rebalancing date to the next. Correlation analysis turned up two fairly clean groupings, strong agreement within the weight-based family, strong agreement within the risk-based family, weak agreement across the two and Kendall's W confirmed that overall consensus across all twelve measures was only moderate. The Friedman and Nemenyi tests then showed, formally, that the disagreement behind that moderate consensus figure is systematic rather than a product of sampling noise.

The predictive regressions added a further, practically useful distinction: several measures, ENB MT, DivDelta, ENB PCA, DR, CVaR DR and Rao Neff among them, carried statistically significant information about future volatility and, to a lesser extent, future downside risk, while PDI and Diversification Return showed essentially no forward-looking power despite performing reasonably well as contemporaneous descriptions. Describing the current state of a portfolio and anticipating its future risk turned out to draw on somewhat different properties of a measure.

Performance results round out the picture. MV Frontier delivered the strongest risk-adjusted returns in the sample, yet it did not lead on any diversification measure its Diversification Ratio was in fact the lowest of the six portfolios. Max-DR, despite the opposite pattern strong diversification scores, unremarkable performance, did not offer the best downside protection either. Together these results make a point that is easy to state but easy to underweight in practice:

portfolio optimisation and diversification measurement are answering different questions. Optimisation produces an allocation given some chosen objective; diversification measures then assess a separate consequence of that allocation, and there is no guarantee the two will point in the same direction.

Taken as a whole, the findings argue against relying on any single diversification indicator. Investors, portfolio managers and researchers are better served by a small set of complementary measures spanning how capital is allocated and how risk is allocated, in its various forms than by any one number, however well established that number might be in the literature. That, more than any individual result in this chapter, is the main contribution this study is trying to make.

5.2 Limitations of the Study

A few limitations are worth stating plainly rather than glossing over.

The investment universe covers eleven ETFs across major asset classes, which is broad enough to be informative but still a specific choice. Results might look somewhat different with a wider set of assets individual securities, cryptocurrencies, or additional international markets, for instance.

The portfolio models are all estimated from historical return data, which carries the usual caveat: they assume that past relationships among assets remain informative about the future. Structural breaks in markets could change how well any of these strategies perform going forward.

This study also examined a specific set of six portfolio construction methods and twelve diversification measures. These were chosen to be representative of both traditional and more recent approaches, but the literature contains many alternatives that were not tested here and could plausibly produce a somewhat different picture.

Finally, none of the results account for transaction costs, taxes, liquidity constraints, or other frictions that would matter in an actual implementation. These are not small omissions in practice, and anyone applying these findings directly should keep that in mind.

5.3 Recommendations for Future Research

A few directions seem worth pursuing from here.

Extending the investment universe to cryptocurrencies, private equity, currencies, individual equities or small-cap stocks, and factor-based ETFs would test whether the relationships documented in this study hold up outside a fairly conventional multi asset class setting.

It would also be worth examining how these portfolios and measures behave under more stressed market conditions specifically financial crises, high inflation periods, or recessions rather than only across a broad historical average that necessarily blends calm and turbulent periods together.

Bringing in practical frictions, transaction costs, taxes, turnover limits, rebalancing costs, would make the results considerably more directly useful to practitioners, at the cost of a more complex modelling exercise.

There is also room to bring in more modern portfolio construction techniques, machine learning, robust optimization, Bayesian methods and to see how they compare with the more classical approaches used here.

Bibliography

- Ang, A., Hodrick, R.J., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1), pp. 259–299.
- Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), pp. 203–228.
- Bawa, V.S., Brown, S.J. and Klein, R.W. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3), pp. 215–231.
- Bera, A.K. and Park, S.Y. (2008). Optimal portfolio diversification using the maximum entropy principle. *Econometric Reviews*, 27(4–6), pp. 484–512.
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, pp. 175–192.
- Best, M.J. and Grauer, R.R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2), pp. 315–342.
- Bloomfield, T., Leftwich, R. and Long, J.B. (1977). Portfolio strategies and performance. *Journal of Financial Economics*, 5(2), pp. 201–218.
- Booth, D.G. and Fama, E.F. (1992). Diversification returns and asset contributions. *Financial Analysts Journal*, 48(3), pp. 26–32.
- Campbell, L.L. (1966). Exponential entropy as a measure of extent of a distribution. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 5, pp. 217–225.
- Carmichael, B., Koumou, G.B. and Moran, K. (2015). *Unifying portfolio diversification measures using Rao's quadratic entropy*. CIRANO Working Paper No. 2015s-16.
- Chateauneuf, A. and Lakhnati, G. (2007). From sure to strong diversification. *Economic Theory*, 32(3), pp. 511–522.
- Chateauneuf, A. and Tallon, J.-M. (2002). Diversification, convex preferences and non-empty core in the Choquet expected utility model. *Economic Theory*, 19(3), pp. 509–523.
- Chopra, V.K. (1991). *Mean-variance revisited: Near-optimal portfolios and sensitivity to input variations*. Russell Research Commentary.
- Chopra, V.K. and Ziemba, W.T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2), pp. 6–11.

- Chopra, V.K., Hensel, C.R. and Turner, A.L. (1993). Massaging mean-variance inputs: Returns from alternative global investment strategies in the 1980s. *Management Science*, 39(7), pp. 845–855.
- Choueifat, Y. and Coignard, Y. (2008). Toward maximum diversification. *The Journal of Portfolio Management*, 35(1), pp. 40–51.
- Choueifat, Y., Froidure, T. and Reynier, J. (2013). Properties of the most diversified portfolio. *Journal of Investment Strategies*, 2(2), pp. 49–70.
- De Giorgi, E.G. and Mahmoud, O. (2016). Diversification preferences in the theory of choice. *Decisions in Economics and Finance*, 39, pp. 143–174.
- Deguest, R., Martellini, L. and Meucci, A. (2013). *Risk parity and beyond: From asset allocation to risk allocation decisions*. SSRN Electronic Journal.
- Dekel, E. (1989). Asset demands without the independence axiom. *Econometrica*, 57(1), pp. 163–169.
- DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5), pp. 1915–1953.
- Evans, J.L. and Archer, S.H. (1968). Diversification and the reduction of dispersion: An empirical analysis. *The Journal of Finance*, 23(5), pp. 761–767.
- Fabozzi, F.J. (ed.) (2013). *Encyclopedia of Financial Models*. Hoboken, NJ: John Wiley & Sons.
- Fama, E.F. and French, K.R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), pp. 427–465.
- Fernholz, E.R. (2002). *Stochastic Portfolio Theory*. New York: Springer.
- Fisher, L. and Lorie, J.H. (1970). Some studies of variability of returns on investments in common stocks. *The Journal of Business*, 43(2), pp. 99–134.
- Föllmer, H. and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time*. 3rd ed. Berlin: De Gruyter.
- Fragiskos, A. (2014). *What is portfolio diversification?* SSRN Electronic Journal.
- Han, Y., Lin, H. and Wang, R. (2023). Diversification quotients based on Expected Shortfall. *Insurance: Mathematics and Economics*.
- Holton, G.A. (2004). Defining risk. *Financial Analysts Journal*, 60(6), pp. 19–25.
- Jobson, J.D. and Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371), pp. 544–554.

- Jobson, J.D. and Korkie, B. (1981). Putting Markowitz theory to work. *The Journal of Portfolio Management*, 7(4), pp. 70–74.
- Jorion, P. (1985). International portfolio diversification with estimation risk. *The Journal of Business*, 58(3), pp. 259–278.
- Kallberg, J.G. and Ziemba, W.T. (1984). Mis-specification in portfolio selection problems. In: Bamberg, G. and Spremann, K. (eds.), *Risk and Capital*. Berlin: Springer.
- Klemkosky, R.C. and Martin, J.D. (1975). The effect of market risk on portfolio diversification. *The Journal of Finance*, 30(1), pp. 147–154.
- Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, 37(5), pp. 519–531.
- Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4), pp. 110–119.
- 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), pp. 13–37.
- Maillard, S., Roncalli, T. and Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4), pp. 60–70.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), pp. 77–91.
- Merton, R.C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7(4), pp. 1851–1872.
- Meucci, A. (2009). Managing diversification. *Risk*, 22(5), pp. 74–79.
- Meucci, A., Santangelo, A. and Deguest, R. (2015). Risk budgeting and diversification based on optimized uncorrelated factors. *Risk*, 28(11), pp. 70–75.
- Michaud, R.O. (1989). The Markowitz optimization enigma: Is ‘optimized’ optimal? *Financial Analysts Journal*, 45(1), pp. 31–42.
- Morrison, D.F. (1990). *Multivariate Statistical Methods*. 3rd ed. New York: McGraw-Hill.
- Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4), pp. 768–783.
- Partovi, M.H. and Caputo, M. (2004). Principal portfolios: Recasting the efficient frontier. *Economics Bulletin*, 7(3), pp. 1–10.
- Philippatos, G.C. and Wilson, C.J. (1972). Entropy, market risk, and the selection of efficient portfolios. *Applied Economics*, 4(3), pp. 209–220.
- Reilly, F.K. and Brown, K.C. (2012). *Investment Analysis and Portfolio Management*. 10th ed. Mason, OH: South-Western Cengage Learning.

- Rockafellar, R.T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *The Journal of Risk*, 2(3), pp. 21–41.
- Rudin, A.M. and Morgan, J.S. (2006). A portfolio diversification index. *The Journal of Portfolio Management*, 32(2), pp. 81–89.
- Salazar Flores, Y., Bhatt, A. and Ivanov, K. (2017). *Diversification Delta: A different perspective*.
- Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), pp. 379–423.
- Sharpe, W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), pp. 425–442.
- Statman, M. (1987). How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis*, 22(3), pp. 353–363.
- Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25(2), pp. 65–86.