Estimating Portfolio Risk using Value-at-Risk and Expected Shortfall: A Comparative Study of Traditional and EVT Approaches

Bryan Mutua Kibui
Nicholas Johnson

WorldQuant University
Bryanmutua999@gmail.com
Tsenghsiencheng@gmail.com

Abstract

Our paper delves into the application of different VaR and ES techniques, namely Parametric, Historical Simulation, Monte Carlo Simulation, and Extreme Value Theory, and tests their reliability across a wide array of traded financial instruments such as equity and bond ETFs, commodities, forex, and cryptocurrencies, using metrics such as the Christoffersen Test and Lopez Loss functions. The research dataset spans from January 1, 2010, to December 31, 2024. The focus of this paper lies in examining the benefits and shortcomings of each VaR and ES technique in estimation, as measured by the frequency and severity of violations. We also balance academic curiosity in understanding the different models with experiments regarding the usefulness of VaR and ES models across several multi-asset portfolios. Initially, Historical Simulated VaR appears to have the best results on the Christoffersen Test and Lopez Loss functions, as returns are distributed in a non-Gaussian manner. However, when applying rolling 90-day VaR and ES, we found that Monte Carlo Simulation using parametric assumptions and Parametric VaR and ES models provide better results, since rolling windows, which introduce time-varying uncertainty and prevent look-ahead bias in VaR and ES estimation, reduce the apparent advantage of historical models and allow parametric models to perform relatively better. Applying regime classification to normal and high volatility periods, all of the VaR and ES methods yield exception rates below 5%, which suggests they are conservative but still useful in predicting VaR and ES. However, during the period of high volatility, all the models except for EVT overshoot, which suggests their reliability during high volatility periods is quite limited. As for EVT, even during the high volatility period, the number of exceptions and the exception rate remain at 0%, which suggests the model is conservative. This highlights the opportunity where EVT-based techniques can be utilised, especially when navigating risk during a high-volatility period.

Notebook accessible in GitHub: https://github.com/nicthc1608/WQU_Capstone.
Google Colab:
https://colab.research.google.com/drive/12x-oQSQdZhv17F1GmN8HsqEhfIGz_YyQ?usp=sharing

Keywords: *Value at Risk (VAR), Expected Shortfall (ES), Risk Management, Risk Estimation, Tail Risk, Monte Carlo Simulation, Risk, Volatility, Parametric method, Historical method, Extreme Value Theory (EVT), Multiasset Portfolio, Risk Modelling, Portfolio Management.*

1. Introduction

Trading or investing in financial markets is often a precarious and perplexing ordeal, as there is a consensus that this undertaking is inherently risky and the payout is uncertain. As such, many attempts and methods have been experimented with to justify and rationalise the economic decision to trade or invest. The market and its risk pattern often follow an economic boom-bust cycle. Sometimes, risks may be suppressed when the market gets overly optimistic, while there have been numerous events in the past where risk premiums are high across markets and even asset classes, this phenomenon is known as risk-on and risk-off (Haldane) and such have occurred during the Asian Financial Crisis in 1998, the Global Financial Crisis in 2008, and a black swan event like COVID-19. These highlight the need for robust risk management.

Value at Risk (VaR) and Expected Shortfall (ES) are some of the metrics often employed to quantify risk in percentage and unit currency terms. For VaR, the goal is to estimate the maximum amount of monetary pecuniary loss within a confidence interval, while ES captures the average loss should the risk surpass the VaR threshold (Chakraborty). Many techniques have been developed to expand on these, especially in dealing with tail risk, which relates to the third and fourth moments—skewness and kurtosis—of financial data. Moreover, many existing studies often focus on single asset classes or provide a static estimation of risk with disregard for regime change.

The three most common approaches to estimating VaR and ES are the parametric method, which involves assuming Gaussian normality of the data to estimate tail risk; the Monte Carlo method, which simulates the return distribution based on given parameter inputs; and the historical method, which assumes the return distribution will follow historical returns (RiskMetrics Group). However, a study has suggested flaws and inadequacies in these approaches to estimating VaR and ES, leading to understatement and simplification of the actual risk assumed (Kamronnaher et al). Therefore, more insights can be drawn from performing VaR and ES analysis across multiple asset classes and market regimes to examine how risk evolves and intersects. The diversification benefits of having different asset classes might reduce overall risk. Lastly, the VaR and ES analysis based on Extreme Value Theory (EVT) offers a unique approach by capturing extreme values through the Pareto distribution, essentially modelling risk preparedness for worst-case scenarios (Baran and Witzany).

For our Capstone Project, we have decided to pursue Topic 6 on Value-at-Risk methodologies under the commodities track. Our focus is on estimating and comparing Value-at-Risk (VaR) and Expected Shortfall (ES) across a diverse portfolio of asset classes, including SPDR ETFs, cryptocurrencies, bonds, commodities, and major forex pairs. We plan to implement and compare four methodologies: Parametric VaR, Historical VAR, Monte Carlo Simulations and EVT-based (Extreme Value Theory) VaR.

The study will be based on a diverse array of asset classes, such as:

Asset Class	Ticker	Description
Cryptos	BTC-USD	Bitcoin
	ETH-USD	Ethereum
Bonds	TLT	20+ Yr Treasuries

	SHY	1–3 Yr Treasuries
	LQD	Investment Grade Corporate Bonds
	EMB	Emerging Market Bonds
Forex	EURUSD=X	Euro
	USDJPY=X	Japanese Yen
	AUDUSD=X	Australian Dollar
	USDCAD=X	Canadian Dollar
SPDR ETFs	SPY	S&P 500
	XLF	Financials
	XLE	Energy
	XLK	Technology
	XLV	Healthcare
	XLY	Consumer Discretionary
	XLP	Consumer Staples
	XLRE	Real Estate
Commodities	GC=F	Gold Futures
	BZ=F	Brent Oil
	HG=F	Copper
	ZC=F	Corn

The research time horizon spans from January 1, 2010, to December 31, 2024 (14 years) and will use daily price data. Our dataset, gathered using the yfinance package, covers key instruments such as SPY, BTC-USD, TLT, EURUSD=X, and BZ=F. This 14-year period captures multiple market regimes—bull and bear markets, economic shocks such as COVID-19, and subsequent recoveries—which allows us to assess the robustness of each model under varying macroeconomic conditions.

As for the project's purpose and scope, we will assess the VaR and ES dynamics using the four techniques outlined above, starting with preliminary data analysis to understand the characteristics of the financial data. We will explore volatility clustering, observe patterns of heightened correlation across different assets during crisis periods, and assess whether diversification benefits persist during such times. We will study the VaR and ES of each asset individually, then within segmented timeframes reflecting differing market regimes, and finally within mixed-asset portfolios to evaluate potential improvements in the risk-return tradeoff.

The project aims to address a gap by presenting the functional application of different statistical tools in expressing VaR and ES, especially in constructing a multi-asset, time-varying risk model. This research is relevant to a broad audience, from educating retail and institutional investors on risk awareness to informing regulators and policymakers on potential systemic vulnerabilities. Ultimately, we aim to produce a toolset and empirical insight that empowers better portfolio construction, risk stress testing, and macro-financial interpretation. The entire analysis will be conducted in Python, with a focus on clear, interpretable outputs for practitioners, regulators, and researchers alike.

2. Problem Statement

Risk management has been the central theme throughout investment planning and analysis. Risk can be defined as the element of uncertainty and the possibility of adverse outcomes, which leads to losses. There are many facets of risks pertinent to trading and investing in financial assets that involve diverse aspects that are both technical and algorithmic, as well as emotions and impulses. Many ways can be used to categorise risk. Based on their origins, risks can be categorised as systematic (about the market) and idiosyncratic (unique risk). Risks in financial assets are multifaceted and may include credit risk, liquidity risk, counterparty risk, and more. Risks can be measured quantitatively, such as through returns and variances, volatility (standard deviation), factor beta, as well as Value-at-Risk and Expected Shortfall, which will be the focus of this paper.

Value-at-Risk (VaR) can be defined as the maximum loss in terms of the monetary unit that the position is exposed to with a certain degree of confidence over a determined holding period (Holton). There are many methods to assess the VaR of a portfolio, such as parametric VaR, historical simulation VaR, Monte Carlo simulated VaR, and EVT-based VaR—each with its pros and cons, which will be analysed more thoroughly in this paper. However, it is worth noting that there are periods when the comovement of assets increases, and risk becomes more consolidated and magnified, especially during high-stress periods (Noussair and Popescu). According to a study done by Loretan and English, there could be contagion effects that ripple across different markets, which heighten volatility and correlation in the market. The risk is multifaceted and includes the possibility of portfolio foreclosure, forced liquidation, or deleveraging.

This highlights the relevance of a risk measurement that not only measures the capital at risk within a certain confidence interval that can be breached, but also how much, on average is the amount lost is if a breach/violation of VaR occurs. Expected Shortfall, also known as Conditional Value-at-Risk (CVaR), is the average of the sum of the worst losses in excess of the confidence level. Mathematically, CVaR can be defined as:

$$ES^{(\alpha)}(X) = -\frac{1}{\alpha}\left(E[X1_{\{X\leq x^{(\alpha)}\}}] - x^{(\alpha)}(P[X \leq x^{(\alpha)}] - \alpha\right) \quad \text{.................. Equation 1}$$

$$ES^{(\alpha)}(X) = \frac{1}{\alpha}\int_0^\alpha VaR_u(X)du \quad \text{.. Equation 2}$$

The methods to calculate this CVaR/ES also employ the same techniques as VaR, such as parametric, historical simulation, Monte Carlo, and EVT (Acerbi and Tarsche).

To understand in greater detail the benefits of these different VaR and CVaR analysis approaches, we will simulate a multi-asset portfolio with a vast investment universe spanning SPDR ETFs, cryptocurrencies, forex, bonds, and commodities. To artificially build the portfolio, we will retrieve the daily prices of these different asset classes spanning 14 years, from 2010 to 2024. The rationale is that, throughout these 14 years, the data will have encompassed a full economic boom-bust cycle—covering bull markets, bear markets, periods of economic downturn, as well as black swan events such as COVID-19, during which risk and uncertainty were heightened. Our goal is to produce practical and interpretable results that help understand macroeconomic trends and, especially, for quantifying financial risk to support better management. We plan to do this using tools such as Python and Excel.

Ultimately, we hope that through our research on the different methodologies of estimating VaR and ES, we can help market practitioners gain a better conceptual understanding of risk and develop a more holistic framework through improved stress modelling and benchmarking. Lastly, we hope to inspire more academicians to develop a multifaceted approach to explore, understand, manage, and even exploit opportunities that may arise from risk.

3. Literature Review

Risk management has been a central part of portfolio management. A considerable number of market practitioners have been experimenting with ways to build the most optimised portfolio and maximise the risk-return tradeoff. Various approaches can be employed to manage the aforementioned risks: namely, avoidance, reduction or mitigation, transfer, and acceptance of risk, depending on the projected payoff (Mossé Cyber Security Institute). However, to make a conscionable decision regarding how to act upon risk, one must first be able to define the risks and present them in such a manner as to ensure sufficient knowledge to compare the risks and corresponding payouts. This is precisely where Value at Risk (VaR) and Expected Shortfall (ES) become useful. Markowitz (1952) and his Modern Portfolio Theory have been referred to as key contributors in this area, and the theory formed the basis of risk mitigation through diversification. Subsequently, other researchers like Rockafellar and Uryasev came up with Conditional Value-at-Risk (CVaR) to incorporate losses over VaR levels. Platanakis and Urquhart (2020) have further demonstrated the diversification advantages of crypto asset portfolio inclusions, while Chakraborty et al. (2021) highlighted how the traditional risk models can be ineffective in case of crises. Such researchers are the frontiers of quantitative risk modelling.

Value-at-Risk (VaR) has been considered as one of the principal metrics of measuring financial risk, particularly within regulation and portfolio management applications. Nevertheless, the consistent reproach of VaR failing to recognise disastrous losses has led to added examination of the Expected Shortfall (ES) in addition, also called Conditional Value-at-Risk (CVaR). This literature review assesses some of the most important milestones in the science of modelling risk and pinpoints where the current methodologies are at present deficient, situating this project within the larger academic and applied financial risk environment. The current literature review will assess what scholarly work has been done in this area, its similarities, and the gap in methods our project can fill.

To begin, it is essential to underline the importance of risk management, a principle cemented by numerous past events, such as the Global Financial Crisis of 2008. The collapse of the banking system may be attributed to a failure to manage liquidity risk, which was arguably exacerbated by mark-to-market accounting (Financial Stability Board). Under this framework, firms with significant funds tied up in illiquid assets were forced to foreclose those assets. Additionally, there was insufficient supervision by regulatory bodies and an unaligned incentive structure, which incentivised financiers to sell to clients—often retail investors unfamiliar with risk—and persuade them that the asset-backed securities, specifically subprime, were low-risk. These actors deceptively marketed tranched asset-backed securities with favourable bond ratings. We also observe the manifestation of contagion risk, wherein the failure of one financial institution triggers the collapse of others, as the perceived value of related assets diminishes. This contagion effect reverberates globally. From this anecdote, we can distil key points crucial to risk management: the necessity of understanding the dynamics and realities of risk and the ability to objectively assess it under both normal and high-volatility conditions. A frequent shortcoming of parametric, historical methods has been identified in several studies: the inability to fit heavy-tailed distributions and non-linear dependencies well. This problem can be especially observed during crises, when volatility clustering and extreme losses happen more often. This has caused the field to shift to the sturdier models like EVT and Monte Carlo simulation.

Nevertheless, these basic approaches are restricted in circumstances of fat tails and non-linear dependencies, which are commonly recorded in currency markets nowadays. Geometrically, traditional parametric methods, which assume normality, can be natural because they are simple, but they do not perform well in fat tails and volatility clustering. Historical simulation gives flexibility using the empirical data; on the other hand, Monte Carlo is based on random sampling from assumed distributions. However, neither of them can perform well enough during the crisis (Chakraborty et al., 2021).

About Value at Risk, some individuals may misinterpret the estimated VaR as the maximum amount of capital that could be lost. This is a significant misconception, as VaR estimates the potential loss under normal market conditions given a specific confidence interval. However, these expectations may prove inaccurate, thereby exposing traders or investors to an illusion of control and safety. This psychological bias can lead to overleveraging and overtrading, ultimately exacerbating the risks involved. Expected Shortfall (ES), by contrast, may offer a marginally improved—though not perfect—framework for interpreting the amount of capital at risk, as it seeks to answer: if losses exceed the VaR threshold, what is the expected average magnitude of that loss?

4. Competitor Analysis

Competitors in the academic context refer to the preceding works that have tried to address a closely related issue or commented on the related approaches. We intend to evaluate these rival studies, outline their strong and weak points, and demonstrate how our project can contribute to them.

Study	Strengths	Weaknesses	Opportunity (What We Do Differently)
Bhattacharyya & Ritolia (2006) *Conditional VaR Using EVT: Indian Equities*	Applied EVT to real-world market data. Introduced Conditional VaR into risk modelling	Focused only on Indian equities; Limited asset class diversity. No portfolio simulation	Apply EVT to a multi-asset global portfolio. Evaluate across different economic regimes
Delis & Plikas (2017) *FHS vs EVT on Currency Pairs*	Compared to advanced tail-risk techniques. Highlighted EVT's benefits in FX	Narrow scope (only currencies). Short time frame	Extend the scope to 5 asset classes. Use a longer historical window (14 years)
Chakraborty et al. (2021) *Literature Review on Extreme Risk*	Synthesised research on VaR, ES, and EVT across many studies	No empirical testing. Purely theoretical	Apply findings practically to stress-tested portfolios. Evaluate models across crises
Loretan & English (2000) *Correlation in Volatile Markets*	Identified correlation spikes during crises. Introduced regime sensitivity into financial risk	Focused on correlation, not VaR/ES directly	Combine VaR/ES modelling with correlation insights. Model regime-aware risk

Basel Committee (2019) *Regulatory Push toward ES*	Supported ES as superior to VaR for capturing tail risk. Industry relevance	Lacks implementation guidance. Assumes institutional context	Test ES empirically across assets. Translate regulatory intent into data-driven insights.

We perceive other works and research in the field of VaR and ES to be opportunities and not threats, since every research ultimately draws from the same goal, which is how to better depict and interpret the level of risk, and thus act accordingly in a way that can mitigate the risks. Moreover, there is a clear possibility to hybridise the different models in a way that better sheds light on VaR. However, we perceive a direct threat to our study to be the increasing disregard for risk, especially in financial decision-making. This is apparent in the rise of betting under the guise of investing and the trend of going all-in on concentrated assets in hopes of getting rich quickly. This phenomenon is expounded more clearly by Newall and Weiss-Cohen, where the authors explain the increasing tendency toward irrationality and risk-seeking behaviour, such as trading high-risk derivatives and taking on excessive leverage. These behaviours not only harm the individuals involved but also pose systemic risks to the financial system by increasing market volatility and exacerbating noise in the markets. We argue that, in turn, this may deceptively distort the true estimates of VaR and ES.

5. Research Methodology

5.1 Research Objectives

The main objective of this research is to evaluate the reliability of different Value-at-Risk (VaR) and Expected Shortfall (ES) estimation techniques, namely parametric, historical, Monte Carlo simulation, and Extreme Value Theory, for portfolios across a multitude of asset classes. Moreover, we aim to design a model for portfolio selection in which the profit and risk trade-off profile can be maximised.

5.2 Research Design

The research will primarily be quantitative, aiming to understand the statistical properties of different assets, uncover the return distributions of various asset classes, and examine interactions such as correlation and co-movements across them. We will observe patterns like spikes in volatility and correlation, particularly during major market events. The analysis begins by applying different models to assess risk—specifically Value-at-Risk (VaR) and Expected Shortfall (ES)—both for individual assets and combined portfolios, to identify diversification benefits. We will also evaluate these assets and portfolios using metrics such as the Sharpe ratio, Sortino ratio, and Jensen's alpha, to assess their performance individually and in combination. Finally, we will analyse how these risk measures behave over time to understand the models' temporal dynamics.

5.3 Data Collection

The financial data we will use consists of historical daily prices for assets across various classes, including cryptocurrencies, forex, bonds (via ETFs), equities (via ETFs), and commodities. This data will be retrieved from Yahoo Finance and will cover the period from January 1, 2010, to December 31, 2024. According to the data audit we conducted, some of the assets have missing dates since some of the SPDRs had not been launched on January 1, 2010, and there are rolling dates for commodities; as such some dates are missing. We take this into account and perform the comparison on the data of the available dates existing for each asset within the asset classes.

5.4 Variable Definition and Selection

- Prices are the market values of different assets across different points in time and areas, expressed in the unit currency. The prices of the assets we analyse are in USD, except for forex pairs, which can be expressed in the units of other currencies (for example, USD/JPY). The price returns are derivatives of prices and can be defined as logarithmic or arithmetical, from which mean return, volatility/standard deviation, skewness, and kurtosis are derived (first, second, third, and fourth order moments of price returns).
- Value-at-Risk is defined as the maximum expected loss over a specified time horizon, given a certain confidence level, and can be expressed either in the amount of money at risk or in percentage terms. However, the maximum expected loss here should not be understood as the maximum in absolute terms, but as relative to the confidence level and the given inputs (Chen).
- Expected shortfall can be defined as the average amount of loss over a specified time horizon, more than the confidence level used for the VaR, and likewise can be expressed in the amount of capital or in percentage terms (Cruickshank). To clarify, it does not set a cap on the maximum amount that can be lost, preserving enough signal over noise. Higher-timeframe data, such as weekly or monthly, may capture broader trends more effectively. However, assessing risk levels like VaR and ES on those timeframes can be too simplistic, as extreme events that are critical to risk estimation may be overlooked. On the other hand, shorter timeframes—such as minute or hourly data—can reflect more of the price fluctuations and risk dynamics, but they come with significantly larger data sizes and tend to be noisier. Given these trade-offs, we adopt a daily timeframe to calculate the average of losses that exceed the VaR.

5.5 Model Framework

The five models we are going to use and compare in our research to estimate VaR and ES include:

- Parametric Approach
- Historical Data Approach
- Monte-Carlo Simulation using Parametric/Gaussian Assumptions
- Monte-Carlo Resampling from Empirical Historical Distribution
- Extreme Value Theory Approach using Peak Over Threshold and Pareto Distribution

VaR models will use a 99% confidence level and ES models will use a 97,5% confidence level

5.6 Evaluation Metrics

To evaluate the performance of different estimation techniques in estimating the VAR and ES, we will be using these metrics:

- Christoffersen Conditional Coverage Test
- Lopez Loss Function

Alongside VAR and ES to evaluate the performance of the risk-optimised portfolios constructed, we aim to also use the following metrics:

- Sharpe ratio
- Sortino ratio
- Jensen's alpha with S&P 500 as the market benchmark

5.7 Tools & Software

The models and codes will be written in Python and use libraries such as Numpy, Pandas, Matplotlib, Seaborn, Scipy, and Yfinance. The Python notebook can be accessed on GitHub: https://github.com/nicthc1608/WQU_Capstone and on Google Colab: https://colab.research.google.com/drive/12x-oQSQdZhv17F1GmN8HsqEhfIGz_YyQ?usp=sharing

5.8 Limitations and Constraints

In undertaking our research, we encountered several limitations and constraints, particularly in selecting data that balances granularity while preserving enough signal over noise. Higher-timeframe data, such as weekly or monthly, may capture broader trends more effectively. However, assessing risk levels like VaR and ES on those timeframes can be too simplistic, as extreme events that are critical to risk estimation may be overlooked. On the other hand, shorter timeframes—such as minute or hourly data—can reflect more of the price fluctuations and risk dynamics, but they come with significantly larger data sizes and tend to be noisier. Given these trade-offs, we adopt a daily timeframe as a compromise. It provides enough granularity to capture meaningful risk behaviour while keeping the data manageable and relatively less noisy.

Moreover, there is also a degree of sensitivity when it comes to the outcomes generated by different VaR and ES models. The parametric model, for instance, relies on the assumption of normality, which might understate tail risk and overlook important statistical properties such as skewness, kurtosis, and non-linearity—traits that are often present in financial data. The historical method, on the other hand, is highly dependent on the chosen time window. Whether the reference period includes high or low volatility, or whether there's an underlying trend, can significantly affect the results.

Monte Carlo models present similar challenges. They rely heavily on the inputs used—such as the assumed statistical distribution and the number of simulations performed—which introduces another layer of subjectivity. Likewise, we see a similar pattern with the Extreme Value Theory (EVT)-based risk model. Since it focuses only on the tail events, it often relies on a relatively small sample size. These extreme observations may not always fit well with the assumed distribution, leading to higher variance and potential instability in the results.

Given these differing strengths and weaknesses, we intend to observe and evaluate the performance of each model across different time periods. To do so, we will apply a set of evaluation metrics—including the Christoffersen Conditional Coverage Test and the Lopez Loss Function—to better understand when and under what conditions each model performs best.

6. Exploratory Data Analysis

6.1 Descriptive Statistics

Firstly, we begin by plotting the daily closing prices and generating descriptive statistics for each asset within the selected asset classes. As shown in Figure 1, from January 1, 2010, to December 31, 2024, the majority of the assets exhibit positive mean daily log returns—except for Treasury bonds, EUR/USD, and AUD/USD, which show slightly negative means. Additionally, some assets display characteristics typically associated with their respective classes. For instance, assets considered more "risky"—such as equities, cryptocurrencies, industrial commodities like oil, and emerging market bonds—tend to have relatively higher standard deviations (i.e., volatility) and excess kurtosis, indicating fatter tails and therefore a greater probability of extreme events. On the other hand, more defensive assets—like Treasury bills—tend to show lower volatility and more platykurtic distributions. Another notable observation is that most of the assets exhibit negative skewness, except for TLT, SHY, and LQD, which represent Treasury bonds, Treasury bills, and high-quality investment-grade corporate bonds, respectively.

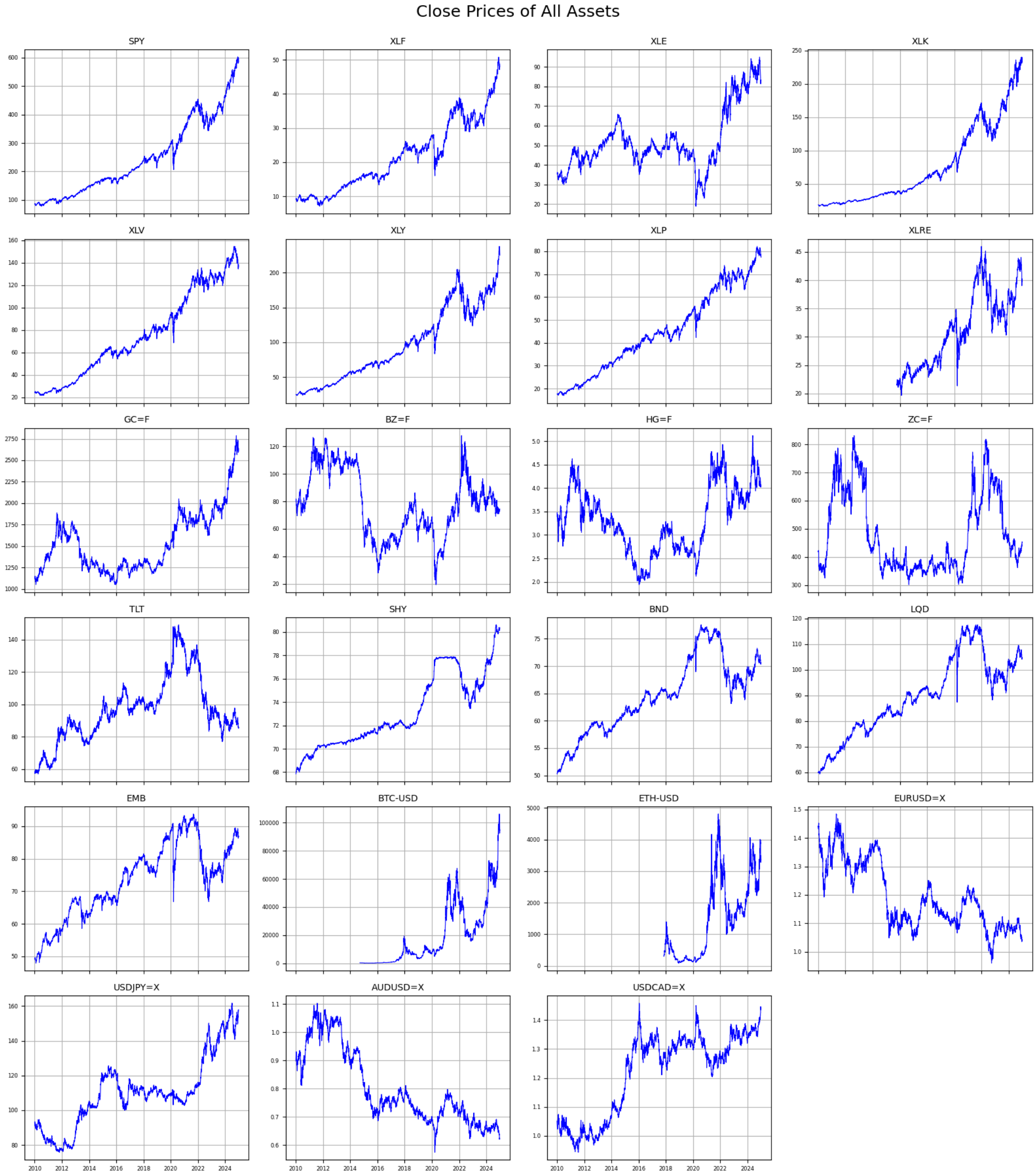
Figure 1: Descriptive Statistics (Daily Log Returns %):

Ticker	Count	Mean	Std	Skew	Kurtosis
SPY	1789	0.05	1.22	-0.82	12.68
XLF	1789	0.04	1.52	-0.59	15.02
XLE	1789	0.03	2.06	-0.93	15.34
XLK	1789	0.08	1.64	-0.44	7.89
XLV	1789	0.04	1.11	-0.47	10.79
XLY	1789	0.05	1.50	-0.77	7.76
XLP	1789	0.03	1.00	-0.56	15.93
XLRE	1789	0.02	1.42	-1.08	17.91
GC=F	1789	0.04	0.94	-0.27	4.05
BZ=F	1789	0.01	2.59	-1.37	19.63
HG=F	1789	0.02	1.42	-0.12	1.65
ZC=F	1789	0.02	1.71	-1.71	19.71
TLT	1789	-0.01	1.02	0.10	4.71
SHY	1789	0.01	0.11	0.73	8.51
LQD	1789	0.01	0.61	0.34	24.96
EMB	1789	0.01	0.70	-2.56	37.50
BTC-USD	1789	0.14	4.42	-0.70	10.25
ETH-USD	1789	0.13	5.64	-0.62	8.97
EURUSD=X	1789	-0.01	0.45	-0.13	1.98
USDJPY=X	1789	0.02	0.55	-0.39	4.02
AUDUSD=X	1789	-0.01	0.64	-0.04	1.52
USDCAD=X	1789	0.01	0.42	0.00	1.21

Note: High kurtosis in Bitcoin confirms fat-tailed behaviour, while equities and bonds remain closer to normality

Observing the price trends across different assets (Table 1) from 2010 to 2024, we can see that the majority of asset classes—such as equities, gold, cryptocurrencies, bonds, and the USD against other currencies—tend to move upward over time, although the momentum and intensity of the increase vary. In contrast, assets like crude oil, copper, and corn show more of a tendency to mean-revert over the long term.

Figure 2: Close Prices of Assets from January 1, 2010, to December 31, 2024



Looking at Figure 2, we can observe a tendency for higher correlations among assets within the same asset class. However, when comparing assets across different classes, the correlations weaken significantly, with some pairs showing little to no correlation or even negative relationships. This suggests that diversification benefits are much smaller when holding assets from the same class, and more meaningful when constructing a multi-asset-class portfolio.

Figure 3: Correlation Matrix of Daily Log Returns Across Assets

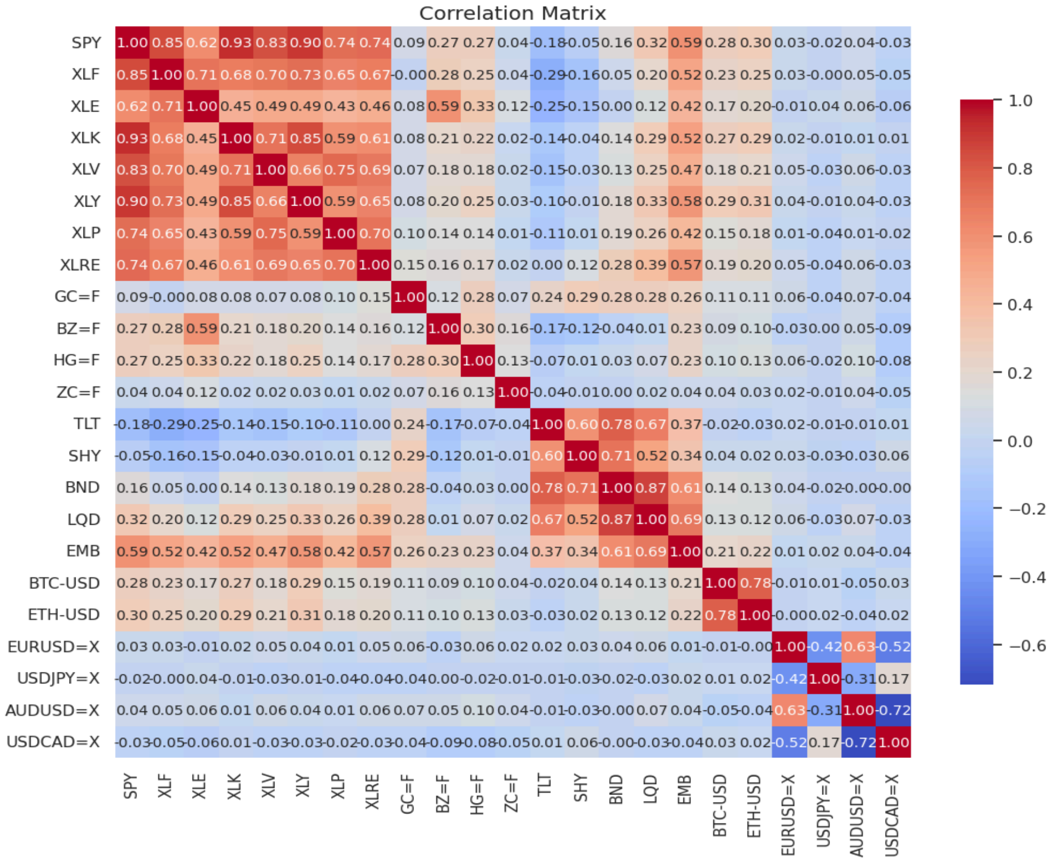


Figure 3 shows the volatility patterns across all assets. While the overall level of volatility varies by asset, we notice that it tends to fluctuate within a certain band. One striking observation is the volatility crowding phenomenon—a sudden and abrupt spike in volatility that appeared across nearly all asset classes during 2020 to 2021, corresponding to the COVID-19 shock and the crisis that followed. Such an example can be seen from

Figure 4: Rolling 1 Year Correlation of SPY and Sector ETFs

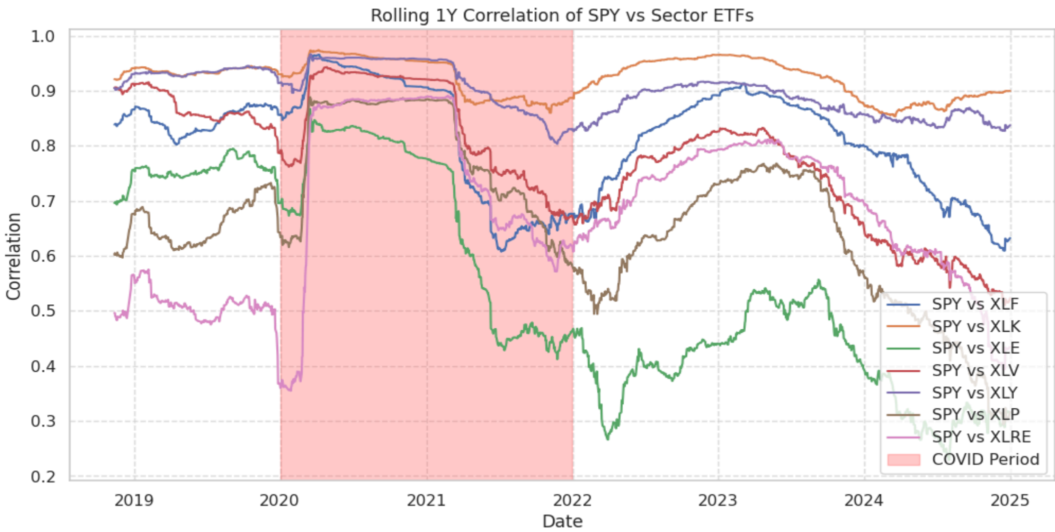
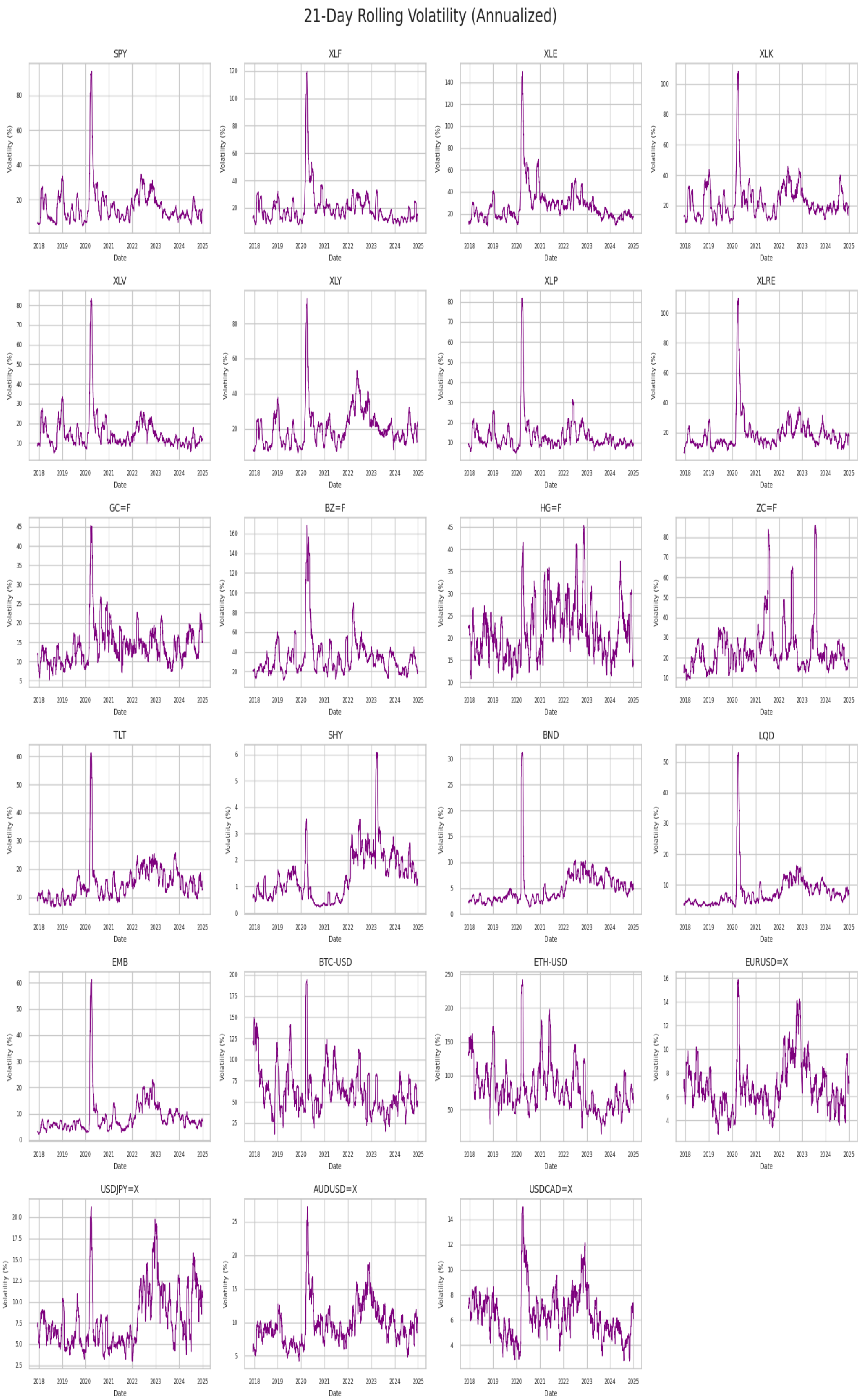


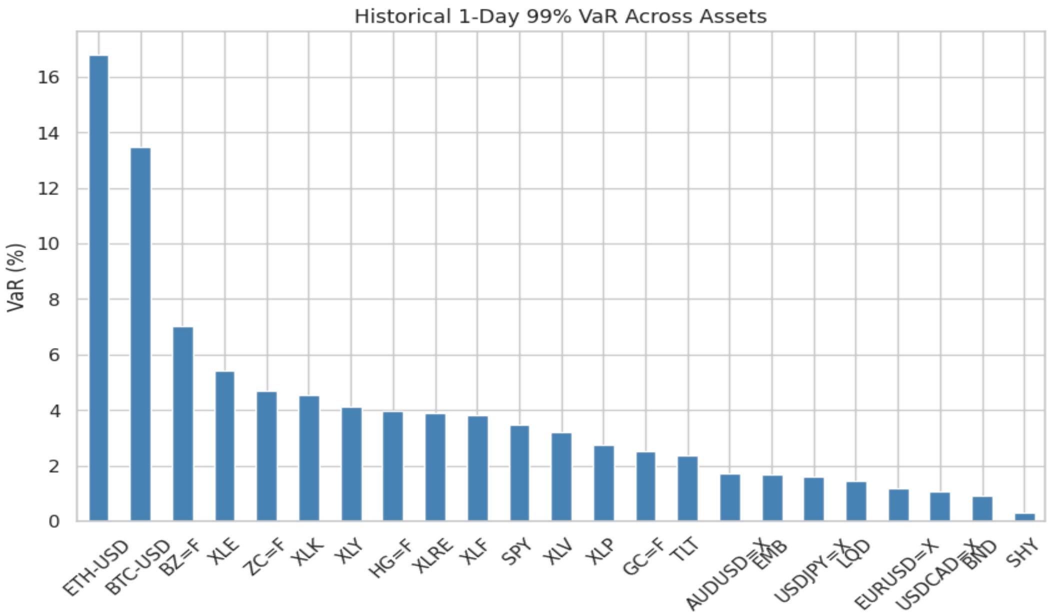
Figure 4, where the rolling 1-year correlations between SPY and each SPDR ETF are heightened from 2020 to 2022 during the COVID-19 Period, with a sudden jump in correlations beginning early 2020, after which the rolling correlation diverges again.

Figure 5: 21-Day Annualised Rolling Volatility Across Assets



7. Results and Discussions

Figure 6. Historical VaR across assets (99% level)



We applied 99% confidence interval for VaR models and 97.5% confidence interval for ES models and ran them across the following methodologies: parametric, historical, Monte Carlo with bootstrap historical distributions, Monte Carlo with parametric assumptions, and Extreme Value Theory. These were applied to all assets within the asset classes. We can observe that the results are quite close across models, given the long time horizon of our analysis (2010–2024). The values for MC Parametric and Parametric, as well as MC Bootstrap Historical and Historical models, are notably similar—this is expected, as the more observations there are for Monte Carlo simulations, the closer they approximate the original parametric or historical distributions. This pattern is consistent throughout. Moreover, our Extreme Value Theory results do not differ drastically from the remainder of the models, reinforcing the idea that with a long time horizon, there is more exceedance, and therefore the VaR and ES estimates become quite close to those from other models. As a sanity check, we can observe that the ES estimated from all models is consistently higher than the corresponding VaR, which, by definition, should always be the case.

Figure 7: Comparison of VAR Across Assets and Estimation Models

	Parametric (99%)	Historical (99%)	MC Bootstrap (99%)
ETH-USD	13.1100	16.8000	18.2500
BTC-USD	10.2800	13.4800	13.4800
BZ=F	6.0200	7.0300	7.0100
XLE	4.8000	5.4100	5.4300
XLK	3.8100	4.5400	4.5800
ZC=F	3.9700	4.6800	4.6700
XLY	3.4800	4.1300	4.1800
XLF	3.5400	3.8200	3.8200
XLRE	3.3100	3.8800	3.9300
HG=F	3.3100	3.9900	3.9800
SPY	2.8500	3.4800	3.5000
XLV	2.5700	3.2100	3.2000
XLP	2.3200	2.7400	2.7600
GC=F	2.1900	2.5400	2.5800
TLT	2.3700	2.3800	2.3800
EMB	1.6300	1.7000	1.7100
AUDUSD=X	1.5000	1.7100	1.7700
LQD	1.4200	1.4300	1.4300
USDJPY=X	1.2700	1.5900	1.5900
EURUSD=X	1.0500	1.1800	1.2200
USDCAD=X	0.9800	1.0600	1.0800
BND	0.9000	0.9300	1.0000
SHY	0.2500	0.3200	0.3200

	MC Parametric (99%)	EVT (99%)
ETH-USD	12.9700	16.4800
BTC-USD	10.2400	12.8300
BZ=F	6.0100	7.7200
XLE	4.8000	5.7600
XLK	3.7800	4.6500
ZC=F	3.9700	4.6300
XLY	3.4200	4.4300
XLF	3.4800	4.3700
XLRE	3.2800	4.0100
HG=F	3.3200	3.8000
SPY	2.8000	3.7300
XLV	2.5400	3.1700
XLP	2.3000	2.8400
GC=F	2.1500	2.7000
TLT	2.3900	2.6200
EMB	1.6300	1.9800
AUDUSD=X	1.5000	1.6700
LQD	1.4000	1.6500
USDJPY=X	1.2500	1.5700
EURUSD=X	1.0600	1.2000
USDCAD=X	0.9700	1.0900
BND	0.8900	1.0400
SHY	0.2500	0.2900

Figure 8: Comparison of ES Across Assets and Estimation Models

	Parametric (97.5%)	Historical (97.5%)	MC Bootstrap (97.5%)
ETH-USD	13.1800	17.6700	17.8300
BTC-USD	10.3300	13.5600	13.4600
BZ=F	6.0500	8.3300	8.4500
XLE	4.8300	6.1900	6.2500
XLK	3.8300	4.8200	4.8600
XLY	3.5000	4.5900	4.5400
XLF	3.5600	4.6500	4.7200
HG=F	3.3300	3.9000	3.9100
ZC=F	3.9900	5.2800	5.3400
SPY	2.8600	3.9200	3.8800
XLRE	3.3200	4.3900	4.3500
XLV	2.5900	3.4300	3.4500
GC=F	2.2000	2.7300	2.7000
TLT	2.3800	2.6500	2.6700
XLP	2.3300	3.1000	3.0800
EMB	1.6400	2.2300	2.2500
AUDUSD=X	1.5100	1.6900	1.7000
USDJPY=X	1.2800	1.6400	1.6500
LQD	1.4200	1.7500	1.8000
EURUSD=X	1.0600	1.2000	1.2000
USDCAD=X	0.9900	1.0900	1.0800
BND	0.9000	1.1100	1.1300
SHY	0.2500	0.3000	0.3000

	MC Parametric (97.5%)	EVT (97.5%)
ETH-USD	13.0200	15.0300
BTC-USD	10.2500	12.0300
BZ=F	6.0400	6.8100
XLE	4.8300	5.1600
XLK	3.7800	4.4300
XLY	3.4500	4.2400
XLF	3.5100	3.9200
HG=F	3.3300	3.7700
ZC=F	3.9800	3.7100
SPY	2.8100	3.5200

XLRE	3.3000	3.4300
XLV	2.5500	2.7600
GC=F	2.1600	2.7200
TLT	2.4000	2.5400
XLP	2.3000	2.3600
EMB	1.6400	1.6700
AUDUSD=X	1.5200	1.6600
USDJPY=X	1.2600	1.5000
LQD	1.4100	1.4900
EURUSD=X	1.0700	1.2100
USDCAD=X	0.9800	1.1700
BND	0.9000	0.9500
SHY	0.2500	0.2700

Moreover, we compare the reliability of the results from all the models across all assets, and as expected—given the relatively close values across the board—the test statistics for both the Christoffersen Conditional Coverage Test and the Lopez Loss Function are quite similar. To reiterate, we used the Christoffersen Conditional Coverage Test to check whether the proportion of violations (i.e., exceedances over the VaR threshold) is consistent with the chosen confidence level. Since the confidence level used for VaR is 99%, we would expect the frequency of exceedances to not exceed 1%. Additionally, we assess whether these exceedances occur independently and are randomly distributed over time, rather than clustering together. In addition, we used the Lopez Loss Function to evaluate and penalise the models based on the severity of the deviations in each instance of violation.

Based on the two tests above, we rank the best-performing model for each asset class. According to Figures 9 and 10, the preliminary results indicate that parametric and Monte-Carlo Parametric models perform the best, showing smaller test statistics (i.e., more consistent and independent violations) and lower loss severity compared to other models—even though the differences are relatively small. This result is interesting because our earlier exploratory data analysis revealed that the returns are distributed in a non-Gaussian manner—something the parametric model does not account for.

However, it should be noted that although the Christoffersen Conditional Coverage Test and Lopez Loss Function show parametric and MC parametric to perform better, the number of violations for parametric models is higher compared to other models such as historical, MC historical, and EVT-based models, and the difference in test statistics is only slight. A higher number of violations indicates that parametric models may have the tendency to underestimate fat-tail or extreme events, which leads to them violating the VaR threshold more often.

Figure 9: Best VAR Models according to Christoffersen Conditional Coverage Test

	Asset	Model	Violations	Test Statistic	p-Value
0	AUDUSD=X	Parametric (99%)	27	-280.0436	1.0000
1	BND	Parametric (99%)	22	-237.2567	

					1.0000
2	BTC-USD	MC Parametric (99%)	35	-344.6953	1.0000
3	BZ=F	Parametric (99%)	34	-336.8373	1.0000
4	EMB	Parametric (99%)	21	-228.4379	1.0000
5	ETH-USD	Parametric (99%)	39	-375.5600	1.0000
6	EURUSD=X	Parametric (99%)	25	-263.1759	1.0000
7	GC=F	Parametric (99%)	33	-328.9193	1.0000
8	HG=F	Parametric (99%)	30	-304.7870	1.0000
9	LQD	Parametric (99%)	21	-228.4379	1.0000
10	SHY	MC Parametric (99%)	29	-296.6102	1.0000
11	SPY	MC Parametric (99%)	39	-375.5600	1.0000
12	TLT	Parametric (99%)	19	-210.5065	1.0000
13	USDCAD=X	MC Parametric (99%)	26	-271.6488	1.0000
14	USDJPY=X	MC Parametric (99%)	39	-375.5600	1.0000
15	XLE	Parametric (99%)	26	-271.6488	1.0000
16	XLF	MC Parametric (99%)	29	-296.6102	1.0000
17	XLK	MC Parametric (99%)	41	-390.6721	1.0000
18	XLP	Parametric (99%)	26	-271.6488	1.0000
19	XLRE	MC Parametric (99%)	27	-280.0436	1.0000
20	XLV	Parametric (99%)	35	-344.6953	1.0000
21	XLY	MC Parametric (99%)	39	-375.5600	1.0000
22	ZC=F	Parametric (99%)	26	-271.6488	1.0000

Figure 10: Best VAR Models According to the Lopez Loss Function

	Asset	Model	Violations	Lopez Loss
0	AUDUSD=X	Parametric (99%)	27	0.0003
1	BND	MC Parametric (99%)	22	0.0001
2	BTC-USD	MC Parametric (99%)	35	0.0121
3	BZ=F	MC Parametric (99%)	34	0.0042
4	EMB	MC Parametric (99%)	21	0.0003
5	ETH-USD	MC Parametric (99%)	39	0.0196
6	EURUSD=X	Parametric (99%)	25	0.0001
7	GC=F	MC Parametric (99%)	33	0.0005
8	HG=F	Parametric (99%)	30	0.0013

9	LQD	MC Parametric (99%)	21	0.0002
10	SHY	MC Parametric (99%)	29	0.0000
11	SPY	MC Parametric (99%)	39	0.0009
12	TLT	Parametric (99%)	19	0.0007
13	USDCAD=X	MC Parametric (99%)	26	0.0001
14	USDJPY=X	MC Parametric (99%)	39	0.0002
15	XLE	MC Parametric (99%)	26	0.0027
16	XLF	MC Parametric (99%)	29	0.0014
17	XLK	MC Parametric (99%)	41	0.0016
18	XLP	MC Parametric (99%)	26	0.0006
19	XLRE	MC Parametric (99%)	27	0.0013
20	XLV	MC Parametric (99%)	35	0.0007
21	XLY	MC Parametric (99%)	39	0.0014
22	ZC=F	MC Parametric (99%)	26	0.0019

In conjunction with analysing the models, we also analyse the asset data we have within each asset class to check for the Sharpe ratio, Sortino ratio, and Jensen's alpha. In other words, we check how well each asset performs on a risk-adjusted basis and more than the assumed market risk (beta). We decide to use both the Sharpe ratio and the Sortino ratio to paint a picture that the risk that is more relevant to our endeavour is the downside risk; as such, the Sortino ratio, which captures downside risk only, might be more relevant. An interesting finding is that the Sortino ratio for a majority of assets is higher than the Sharpe ratio (except SHY, BND, EURUSD, AUDUSD, and USDCAD), which shows that the upside volatility is higher than the downside volatility. This is in contrast with our exploratory data analysis earlier that showed most of the asset returns are negatively skewed, which might indicate a higher probability of downside than upside and the possibility of negative jumps happening—something that can be dug deeper into. Moreover, the equity asset class tends to have a higher market beta compared with the rest. Another interesting finding here is that cryptos (BTC-USD and ETH-USD) and oil (BZ) seem to have moderate to high positive correlation with market beta—something that is worth investigating deeper into, perhaps in future studies, as cryptos are often touted as a diversification option for equities. We can see that most assets have either a slight and quite negligible positive or negative Jensen's alpha, apart from cryptos which exhibit quite high positive Jensen's alpha, which means addition of cryptos into a portfolio can potentially generate returns above what is expected given their market beta.

Figure 11: The Comparisons of Sharpe Ratio, Sortino Ratio, Market Beta and Jensen's Alpha for Each Asset

	Sharpe Ratio	Sortino Ratio	Beta	Jensen's Alpha
SPY	0.6630	0.8115	1.0003	-0.0000
XLF	0.4787	0.6192	1.1333	-0.0221
XLE	0.2371	0.3133	1.1031	-0.0596
XLK	0.7562	0.9798	1.1535	0.0306
XLV	0.6094	0.8070	0.7926	0.0085
XLY	0.6864	0.8676	1.0698	0.0167
XLP	0.5718	0.7383	0.6136	0.0090
XLRE	0.2789	0.3488	0.8436	-0.0427
GC=F	0.2468	0.3260	0.0437	0.0344

BZ=F	0.0828	0.1064	0.6306	-0.0363
HG=F	0.0331	0.0492	0.4154	-0.0398
ZC=F	0.0512	0.0663	0.1235	0.0001
TLT	0.0517	0.0790	-0.2891	0.0406
SHY	-1.4002	-1.9548	-0.0095	-0.0175
BND	-0.1255	-0.1566	0.0006	-0.0062
LQD	0.1339	0.1714	0.0814	0.0011
EMB	0.1349	0.1520	0.2765	-0.0189
BTC-USD	0.8576	1.1353	0.7843	0.4851
ETH-USD	0.6391	0.8749	1.2510	0.1531
EURUSD=X	-0.5526	-0.8074	0.0146	-0.0433
USDJPY=X	0.0916	0.1298	0.0082	0.0046
AUDUSD=X	-0.4548	-0.6698	0.0455	-0.0524
USDCAD=X	-0.0896	-0.1372	-0.0301	0.0034

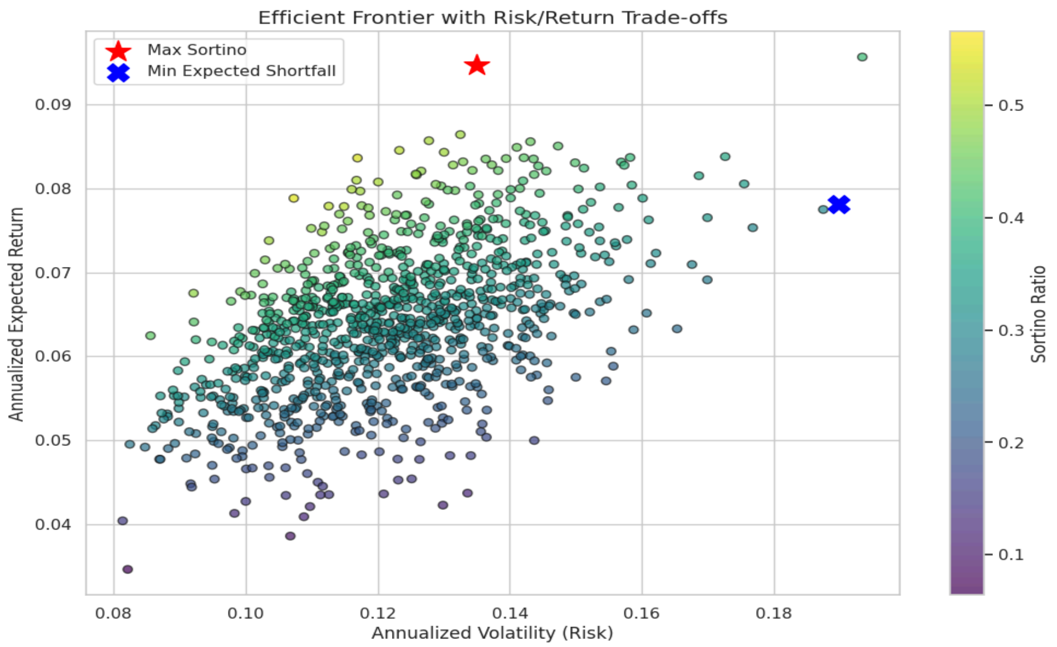
Moreover, in Figure 11, we also model the dependence in the joint distribution among every asset to better reflect the tail dependence. There is a higher correlation among assets belonging to the same asset classes for equities, cryptos, and bonds. Moreover, we can check that there is little joint distribution dependence between currencies and commodities and the assets belonging to different asset classes. This means that diversification benefits are more evident when combining assets across these classes, as the low dependence reduces the likelihood of simultaneous extreme losses.

Figure 12: Copula-based Correlation Matrix among the Assets

	SPY	XLF	XLE	XLK	XLV	XLY	XLP	XLRE	GC=F	BZ=F	HG=F	ZC=F	TLT	SHY	BND	LQD	EMB	BTC-USD	ETH-USD	EURUSD=X	USDJPY=X	AUDUSD=X	USDCAD=X
SPY	1.00	0.74	0.52	0.90	0.73	0.87	0.60	0.62	0.08	0.20	0.26	0.02	-0.12	-0.04	0.06	0.26	0.48	0.21	0.22	0.02	0.01	0.03	-0.03
XLF	0.74	1.00	0.60	0.55	0.55	0.63	0.51	0.49	0.00	0.19	0.25	0.05	-0.23	-0.11	-0.07	0.10	0.33	0.21	0.18	0.00	0.01	0.05	-0.05
XLE	0.52	0.60	1.00	0.35	0.37	0.40	0.31	0.31	0.09	0.52	0.32	0.11	-0.24	-0.11	-0.11	0.03	0.23	0.12	0.13	-0.03	0.03	0.04	-0.03
XLK	0.90	0.55	0.35	1.00	0.60	0.80	0.44	0.49	0.08	0.15	0.22	0.00	-0.08	-0.03	0.07	0.25	0.43	0.21	0.22	0.01	0.01	0.01	0.00
XLV	0.73	0.55	0.37	0.60	1.00	0.56	0.63	0.58	0.06	0.10	0.17	0.02	-0.09	-0.02	0.07	0.22	0.38	0.11	0.13	0.03	-0.01	0.05	-0.04
XLY	0.87	0.63	0.40	0.80	0.56	1.00	0.49	0.53	0.06	0.15	0.24	0.02	-0.07	-0.01	0.10	0.28	0.48	0.22	0.23	0.02	0.01	0.03	-0.03
XLP	0.60	0.51	0.31	0.44	0.63	0.49	1.00	0.62	0.08	0.09	0.13	0.00	-0.03	0.02	0.10	0.22	0.33	0.08	0.09	-0.03	0.00	0.00	-0.01
XLRE	0.62	0.49	0.31	0.49	0.58	0.53	0.62	1.00	0.12	0.08	0.14	0.00	0.07	0.10	0.23	0.35	0.45	0.14	0.13	0.04	-0.01	0.04	-0.03
GC=F	0.08	0.00	0.09	0.08	0.06	0.06	0.08	0.12	1.00	0.10	0.27	0.05	0.23	0.23	0.28	0.26	0.24	0.09	0.08	0.05	-0.04	0.04	-0.03
BZ=F	0.20	0.19	0.52	0.15	0.10	0.15	0.09	0.08	0.10	1.00	0.28	0.13	-0.18	-0.08	-0.10	-0.04	0.09	0.04	0.05	-0.03	-0.03	0.04	-0.06
HG=F	0.26	0.25	0.32	0.22	0.17	0.24	0.13	0.14	0.27	0.28	1.00	0.11	-0.01	-0.01	0.00	0.06	0.18	0.10	0.13	0.03	-0.03	0.08	-0.07
ZC=F	0.02	0.05	0.11	0.00	0.02	0.02	0.00	0.00	0.05	0.13	0.11	1.00	0.00	0.00	0.00	0.03	0.04	0.02	0.02	0.02	0.01	0.05	-0.05
TLT	-0.12	-0.23	-0.24	-0.08	-0.09	-0.07	-0.03	0.07	0.23	-0.18	-0.01	0.00	1.00	0.50	0.86	0.72	0.38	0.00	0.00	0.04	-0.03	-0.02	0.04
SHY	-0.04	-0.11	-0.11	-0.03	-0.02	-0.01	0.02	0.10	0.23	-0.08	-0.01	0.00	0.50	1.00	0.66	0.49	0.29	0.02	0.01	0.01	-0.02	-0.04	0.05
BND	0.06	-0.07	-0.11	0.07	0.07	0.10	0.10	0.23	0.28	-0.10	0.00	0.00	0.86	0.66	1.00	0.86	0.54	0.05	0.05	0.06	-0.03	0.00	0.02
LQD	0.26	0.10	0.03	0.25	0.22	0.28	0.22	0.35	0.26	-0.04	0.06	0.03	0.72	0.49	0.86	1.00	0.64	0.10	0.09	0.06	-0.02	0.03	0.01
EMB	0.48	0.33	0.23	0.43	0.38	0.48	0.33	0.45	0.24	0.09	0.18	0.04	0.38	0.29	0.54	0.64	1.00	0.15	0.15	0.02	0.02	0.02	-0.01
BTC-USD	0.21	0.21	0.12	0.21	0.11	0.22	0.08	0.14	0.09	0.04	0.10	0.02	0.00	0.02	0.05	0.10	0.15	1.00	0.71	-0.01	0.02	-0.06	0.04
ETH-USD	0.22	0.18	0.13	0.22	0.13	0.23	0.09	0.13	0.08	0.05	0.13	0.02	0.00	0.01	0.05	0.09	0.15	0.71	1.00	0.00	0.02	-0.05	0.03
EURUSD=X	0.02	0.00	-0.03	0.01	0.03	0.02	-0.03	0.04	0.05	-0.03	0.03	0.02	0.04	0.01	0.06	0.06	0.02	-0.01	0.00	1.00	-0.41	0.62	-0.51
USDJPY=X	0.01	0.01	0.03	0.01	-0.01	0.01	0.00	-0.01	-0.04	-0.03	-0.03	0.01	-0.03	-0.02	-0.03	-0.02	0.02	0.02	0.02	-0.41	1.00	-0.31	0.18
AUDUSD=X	0.03	0.05	0.04	0.01	0.05	0.03	0.00	0.04	0.04	0.04	0.08	0.05	-0.02	-0.04	0.00	0.03	0.02	-0.06	-0.05	0.62	-0.31	1.00	-0.72
USDCAD=X	-0.03	-0.05	-0.03	0.00	-0.04	-0.03	-0.01	-0.03	-0.03	-0.06	-0.07	-0.05	0.04	0.05	0.02	0.01	-0.01	0.04	0.03	-0.51	0.18	-0.72	1.00

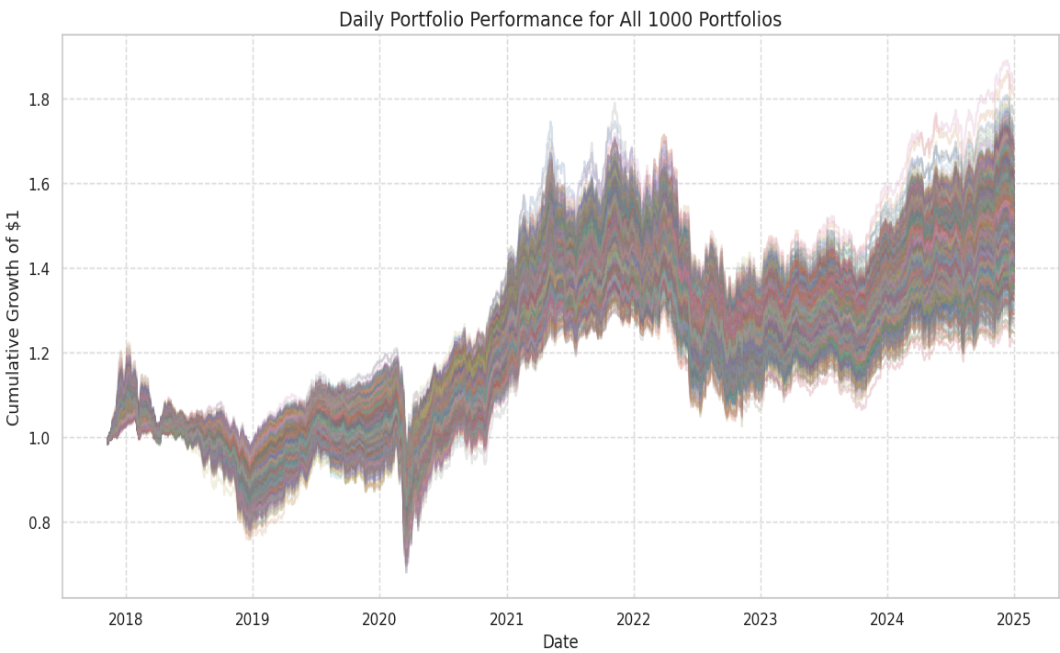
Based on the preliminary findings, we generate 1000 portfolios with random weights for each asset (multi-asset class portfolios) and rank them based on VAR, ES and Sortino ratio.

Figure 13: Efficient Frontier Based on 1000 Multi-asset Portfolios



We then continue by studying the returns of these 1000 portfolios. As some assets do not have the earliest trading price from before 2018, the portfolios are analysed for their daily cumulative returns from November 10, 2017, to December 31, 2024.

Figure 14: Cumulative Performance for 1000 Portfolios from Nov 10, 2017 to Dec 31, 2024



There are periods of downtrends throughout 2018 to 2019, followed by early 2020 when there was a sharp drop due to the COVID-19 pandemic. After this, the portfolios quickly recovered upward in the years that followed until 2022, where there was yet another downtrend, followed by a sideways movement until around 2024. All of these were experienced by all portfolios, which suggests some degree of correlations and co-movements across all portfolios.

We performed all the methods of VaR and ES on a rolling 90-day basis to check the performance of all the models on the top 10th percentile portfolios (100 portfolios). Afterwards, we ran statistical tests such as the Kupiec test, the Christoffersen independence test, and the Lopez loss function. Interestingly, we found that historical VaR and ES models outperformed the other models in the rolling setup. This result is contrary to…

Interestingly, we found that Monte Carlo simulation using parametric assumption and parametric VaR and ES models outperformed the other models in the rolling setup. This result is contrary to when we used static estimation, where historical-based models appeared to perform better. The difference suggests that rolling windows, which introduce time-varying uncertainty and prevent look-ahead bias in VaR and ES estimation, reduce the apparent advantage of historical models and allow parametric models to perform relatively better.

Even though the statistics for EVT show no violations (Kupiec score = 0, Christoffersen score = 1), this outcome is due to the EVT model producing overly conservative estimates in short rolling windows, resulting in VaR and ES thresholds that are too wide to be practical. While this eliminates violations in backtesting, the excessive conservatism may limit the applicability of the EVT model.

Figure 15: Rankings of the Five EVT and ES Models based on Exception Rate, Kupiec Test, Christoffersen Test and Lopez Loss

	method	exc_rate	kupiec_p	christoffersen_p	lopez_loss	score
2	MC_Boot	0.0000	0.0000	1.0000	0.0000	1.0100
0	EVT	0.0000	0.0000	1.0000	0.0000	1.0100
1	Hist	0.0146	0.1099	0.2004	0.0000	1.6943
3	MC_Param	0.0241	0.0000	0.0088	0.0000	2.0053
4	Param	0.0241	0.0000	0.0088	0.0000	2.0053

We then separate our time series price data for all assets into two regimes, namely the “normal” regime and the “high volatility” regime, based on the VIX during each period, where we use the cutoff VIX of 30 to define which falls into normal and high volatility. After that, we re-run all the VaR and ES models and compare the number of exceptions and the exception rate.According to the results in Figure 15, during the normal period, all of the VaR and ES methods yield reasonable exception rate which does not differ much from expected violation of 1% (confidence level 99%), which suggests they are conservative but still useful in predicting VaR and ES. However, during high volatility period, the exception rate increased significantly for historical, parametric and MC parametric methods. Historical exception rate increased from 1.10% to 5.51% and parametric and MC parametric exception rate jumped from 1.72% to 10.17%. This suggests their reliability during high volatility periods is quite limited. As for MC historical and EVT model, even during the high volatility period, the number of exceptions and the exception rate remain at 0%, which suggests the models are still quite conservative. This highlights the opportunity where EVT-based techniques can be utilised, especially when navigating risk during a high-volatility period.

Figure 16: Comparisons of the Number of Exceptions and Exception Rate for Each VAR and ES methods During Normal and High Volatility Period.

Method	Normal Period		High Volatility Period	
	# of Exceptions	Exc. Rate	# of Exceptions	Exc. Rate
EVT	0	0%	0	0%
Historical	17.18	1.10%	7.6	5.51%
MC_Boot	0	0%	0	0%
MC_Param	26.86	1.72%	14.04%	10.17%
Param	26.86	1.72%	14.04%	10.17%

8. Conclusions

This study is focused on the comparison of traditional and Extreme Value Theory (EVT) techniques in calculating VaR and Expected Shortfall across a wide array of asset classes, namely equities, bonds, forex, commodities, and cryptocurrencies, throughout the span of 14 years from 2010 to 2024. After this, we ran several tests, such as the Christoffersen test and Lopez Loss function, to determine the most reliable VaR and ES technique to use. In our study, we test the models under various conditions, such as static time series, 90-day rolling VaR and ES, as well as separating the time series into two regimes, namely normal periods and high-volatility periods. In addition, we also construct portfolios configured by optimising the risk-return aspects (Sharpe Ratio, Treynor Ratio, and Jensen's Alpha) and the quantifiable measures of risk, such as VaR and ES.

In our preliminary findings, we found that none of the assets have completely Gaussian properties. Most assets, except for Treasury bonds, EUR/USD, and AUD/USD, have positive mean returns. Riskier assets tend to have higher volatility and excess kurtosis, which implies a higher likelihood of extreme events and thus a higher chance of major loss. Similarly, relatively defensive assets tend to exhibit platykurtic tendencies with lower volatility. From 2010 to 2024, the majority of assets within the respective asset classes exhibit an upward trend, with some commodities exhibiting mean-reverting behaviour. The assets have varying degrees of correlation intra-asset class and across asset classes to a varying degree. Assets belonging to the same asset class have relatively higher correlation compared to those between asset classes, where the correlation becomes weaker, with some even having slightly negative correlation. However, the volatility crowding effect becomes apparent during periods of market shock, as can be observed from 2020 to 2022, when the correlations between sector ETFs (SPDRs) and the market index (SPY) become inflated significantly.

Applying the parametric, historical, Monte Carlo simulation of historical bootstrap and parametric distribution, as well as EVT models that are set to have a 95% confidence interval, the results obtained are quite similar across all the techniques over the span of 14 years (2010 to 2024). Moreover, the EVT estimates of VaR and ES are not drastically different from the other techniques. Even though the results are quite close, applying the Christoffersen Conditional Coverage Test and Lopez Loss Function shows the Historical and Historical Monte Carlo VaR and ES techniques to have slightly better results, with parametric scoring the lowest, which can be foreseen as the return distribution for all the assets tends not to follow Gaussian distribution and, as such, the prediction from Gaussian distribution tends to underestimate the actual level of risk and thus experience slightly more exceedance breaches. However, there is also the possibility of look-ahead bias interfering with the results.

Evaluating the performance metrics of all the assets within the respective asset classes, we found that for the majority of assets, except for SHY and BND (bonds) and the EURUSD, AUDUSD, and USDCAD pairs (forex), the Sortino ratio is relatively higher than the Sharpe ratio, which suggests greater upside volatility compared to downside volatility. Moreover, the majority of assets have negative Jensen's alpha, which implies underperformance relative to the market risk. Another point worth observing is that assets such as equities, crude oil, and cryptocurrencies tend to be moderately to highly correlated, which may diminish the diversification benefit of holding these assets together in a portfolio. However, from the copula correlation matrix, there is relatively lower dependence across asset classes, which suggests that the risk of extreme co-movements (tail risk) can be reduced by holding a multi-asset-class portfolio.

Generating a list of 1000 randomly constructed multi-asset class portfolios and ranking them by VaR, ES, and Sortino ratio, we found that the great majority of portfolios exhibit co-movement and correlated drawdowns to a varying degree. This is especially pronounced during the early onset of COVID-19 at the beginning of 2020. Moreover, it can be observed that several trends occurred throughout the analysis period, such as downtrend, uptrend, and sideways movement. Re-applying the test on all the VaR and ES techniques on a rolling 90-day basis, we found that, unlike the previous finding, when using the 90-day basis, parametric and Monte Carlo simulation with Gaussian (parametric) assumptions exhibit slightly better results based on the Kupiec test, Christoffersen test, and Lopez Loss compared to Monte Carlo simulation using bootstrap historical distribution, and better results compared to historical VaR and ES, which challenged our previous finding that the historical method is better. A possible explanation is that the rolling approach introduces time-varying uncertainty and thus reduces the look-ahead bias that earlier advantaged the historical approach. Moreover, even though the EVT-based approach of VaR and ES does not have any exceptions and thus produces no exceedances when running the tests, this, if anything, implies that the EVT model is overly conservative.

Applying regime distinction based on the level of VIX (volatility index) throughout the analysis period, we found that all the models are quite conservative with the predictions, with the exception rates of the models being below the level of exceptions that is expected (confidence level is 95%, which means the exception rate anticipated is 5%). However, another story is depicted when we look into the high-volatility regime, where the exception rates increase several times above the expected level for all models except for EVT-based techniques, which shows that the predicted VaR and ES are underestimated by all the traditional models. This highlights the opportunity where EVT-based techniques can be utilised, especially when navigating risk during a high-volatility period.

Works Cited

Acerbi, Carlo, and Dirk Tasche. "On The Coherence of Expected Shortfall." Journal of Banking & Finance, vol. 26, no. 7, July 2002, pp. 1487–503. https://doi.org/10.1016/s0378-4266(02)00283-2.

Bank For International Settlements. "Minimum capital requirements for market risk." Basel Committee on Banking Supervision, 2019, www.bis.org/bcbs/publ/d352.pdf. Accessed 5 Aug. 2025.

Baran, Jaroslav, and Jiri Witzany. "A Comparison of EVT and Standard VaR Estimations." SSRN Electronic Journal, Jan. 2011, https://doi.org/10.2139/ssrn.1768011.

Bhattacharyya, Malay, and Gopal Ritolia. "Conditional VaR Using EVT – Towards a Planned Margin Scheme." International Review of Financial Analysis, vol. 17, no. 2, Sept. 2006, pp. 382–95. https://doi.org/10.1016/j.irfa.2006.08.004.

Chakraborty, Biplab. "Value at Risk." Journal of Rural Development, 2022. The Management Accountant, https://nirdprojms.in.

Chakraborty, Gourab, et al. "Measurement of Extreme Market Risk: Insights From a Comprehensive Literature Review." Cogent Economics & Finance, vol. 9, no. 1, Jan. 2021, https://doi.org/10.1080/23322039.2021.1920150.

Chen, James Ming. "On Exactitude in Financial Regulation: Value-at-Risk, Expected Shortfall, and Expectiles." Risks, vol. 6, no. 2, June 2018, p. 61. https://doi.org/10.3390/risks6020061.

Delis, Panagiotis, and John Hlias Plikas. "Comparing Methods to Forecasting Var: Fhs Versus Evt Approaches." IOSR Journal of Economics and Finance, vol.. 08, no. 02, Mar. 2017, pp. 63–68. https://doi.org/10.9790/5933-0802026368.

Financial Stability Board. "Risk Management Lessons from the Global Banking Crisis of 2008." SENIOR SUPERVISORS GROUP, 2009, www.fsb.org/uploads/r_0910a.pdf. Accessed 19 July 2025.

Haldane, Andrew. Risk off · The Hedge Fund Journal, thehedgefundjournal.com/risk-off/. Accessed 4 Aug. 2025.

Holton, Glyn A. "1.4 Value at Risk." Value-at-Risk Theory and Practice, Boston, Massachusetts, 2014, https://www.value-at-risk.net/value-at-risk/. Accessed 4 Aug. 2025.

Kamronnaher, Kanon, et al. Estimating Value at Risk and Expected Shortfall: A Brief and Some New Developments, 2024, https://doi.org/https://doi.org/10.48550/arXiv.2405.06798.

Loretan, Mico, and William B. English. “III. Special feature: Evaluating changes in correlations during periods of high market volatility *.” BIS Quarterly Review, Jan. 2000, www.bis.org/publ/r_qt0006e.pdf.

Markowitz, Harry. “PORTFOLIO SELECTION*.” The Journal of Finance, vol. 7, no. 1, Mar. 1952, pp. 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

Mossé Cyber Security Institute. Risk Management: Avoid, Accept, Mitigate, Transference.
library.mosse-institute.com/articles/2022/05/risk-management-avoid-accept-mitigate-transference/risk-management-avoid-accept-mitigate-transference.html.

Newall, Phillip W. S., and Leonardo Weiss-Cohen. “The Gamblification of Investing: How a New Generation of Investors Is Being Born to Lose.” International Journal of Environmental Research and Public Health, edited by Sabrina Molinaro, 2022, https://doi.org/10.3390/ijerph19095391.

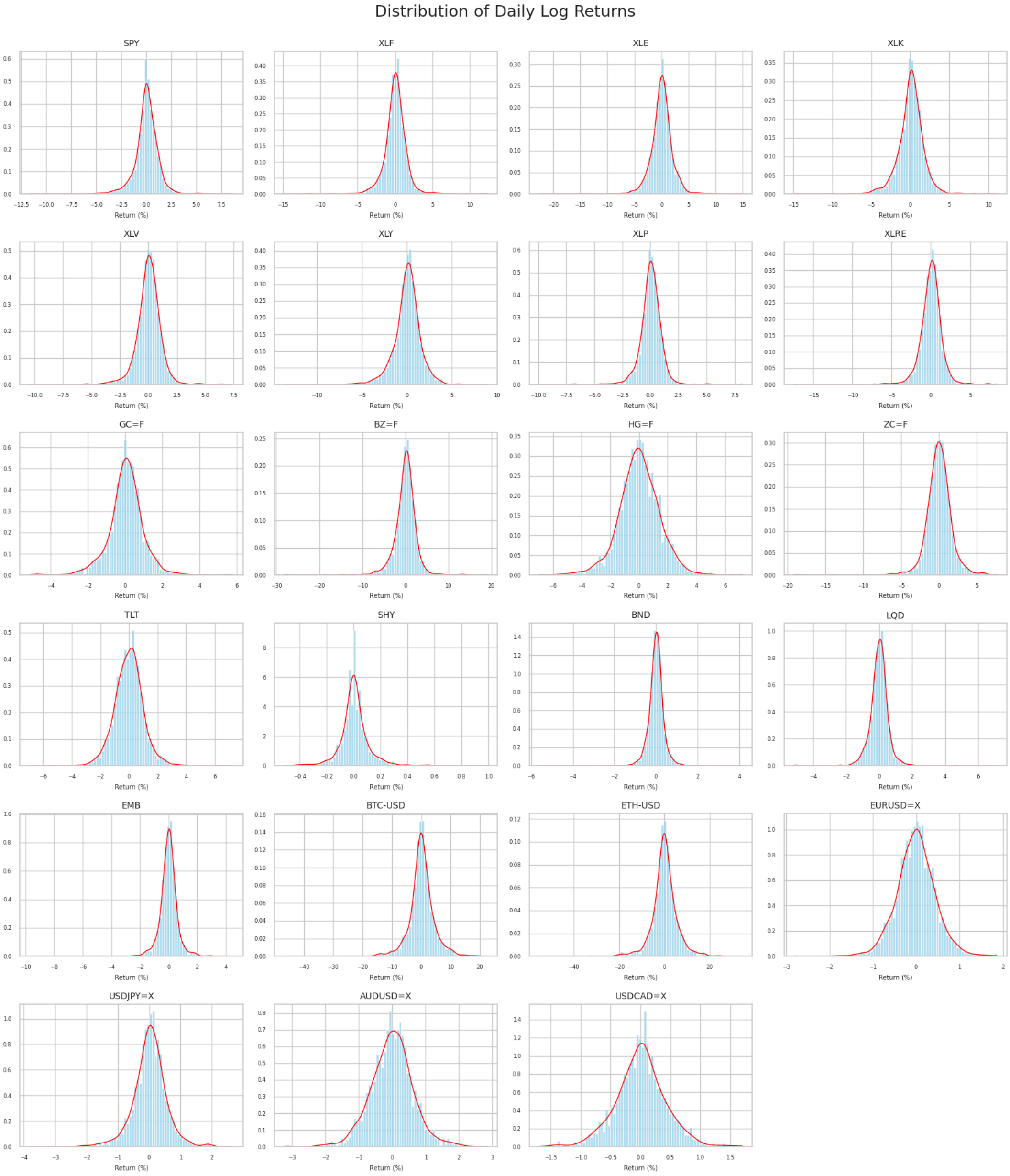
Noussair, Charles N., and Andreea Victoria Popescu. “Comovement and Return Predictability in Asset Markets: An Experiment With Two Lucas Trees.” Journal of Economic Behaviour & Organisation, vol. 185, Apr. 2021, pp. 671–87. https://doi.org/10.1016/j.jebo.2021.03.012.

Platanakis, Emmanouil, and Andrew Urquhart. “Should Investors Include Bitcoin in Their Portfolios? A Portfolio Theory Approach.” The British Accounting Review, vol. 52, no. 4, July 2019, p. 100837. https://doi.org/10.1016/j.bar.2019.100837.

RiskMetrics Group. “VaR: Parametric Method, Monte Carlo Simulation, Historical Simulation.” Yale School of Management, 2015, https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://elischolar.library.yale.edu/cgi/viewcontent.cgi%3Farticle%3D1605%26context%3Dypfs-documents&ved=2ahUKEwiM9JKYwvGOAxUwxDgGHbKnIfs4ChAWegQIKxAB&usg=AOvVaw0UhEqucjaoyQtRm_ocKXxK. Accessed 4 Aug. 2025.

Rockafellar, R. Tyrrell, and Stanislav Uryasev. “Conditional Value-at-risk for General Loss Distributions.” Journal of Banking & Finance, vol. 26, no. 7, July 2002, pp. 1443–71. https://doi.org/10.1016/s0378-4266(02)00271-6.

Appendix 1: Distribution of Daily Log Returns Across Assets



Appendix 2: Data Audit Table for Each Asset

Ticker	Start	End	Rows	Missing Days
SPY	2010-01-04	2024-12-30	3773	138
XLF	2010-01-04	2024-12-30	3773	138
XLE	2010-01-04	2024-12-30	3773	138
XLK	2010-01-04	2024-12-30	3773	138
XLV	2010-01-04	2024-12-30	3773	138
XLY	2010-01-04	2024-12-30	3773	138
XLP	2010-01-04	2024-12-30	3773	138
XLRE	2015-10-08	2024-12-30	2322	86
GC=F	2010-01-04	2024-12-30	3770	141
BZ=F	2010-01-04	2024-12-30	3740	171
HG=F	2010-01-04	2024-12-30	3771	140
ZC=F	2010-01-04	2024-12-30	3769	142
TLT	2010-01-04	2024-12-30	3773	138
SHY	2010-01-04	2024-12-30	3773	138
BND	2010-01-04	2024-12-30	3773	138
LQD	2010-01-04	2024-12-30	3773	138
EMB	2010-01-04	2024-12-30	3773	138
BTC-USD	2014-09-17	2024-12-30	3758	0
ETH-USD	2017-11-09	2024-12-30	2609	0
EURUSD=X	2010-01-01	2024-12-30	3907	5
USDJPY=X	2010-01-01	2024-12-30	3907	5
AUDUSD=X	2010-01-01	2024-12-30	3906	6
USDCAD=X	2010-01-01	2024-12-30	3906	6