Mockus, A. (2025). Performance evaluation of exchange-traded funds in the US. *Transformations and Sustainability, 1*(2), 132-149. https://doi.org/10.63775/ypw2q026

Performance evaluation of exchange-traded funds in the US

Arūnas Mockus*
Kaunas Faculty, Vilnius University, Kaunas, Lithuania

Abstract

This study investigates the performance of the largest U.S.-domiciled exchange-traded funds (ETFs) focused on large-cap equities by assets under management (AUM), covering the period from 2010 to 2024. ETFs play a central role in modern investing, offering low-cost, diversified, and liquid access to broad market exposures. While previous research has examined various ETF categories, this paper concentrates on the core group of top equity ETFs that dominate investor portfolios. Using a set of well-established performance metrics—total return, average annual return, standard deviation, alpha, beta, Sharpe ratio, and Treynor ratio—the analysis assesses both absolute and risk-adjusted performance across eight ETFs. The results reveal notable differences in return generation and risk efficiency, even among funds with similar investment styles. Growth-oriented ETFs generally delivered higher returns but also exhibited greater volatility, while more conservative or value-focused ETFs showed mixed outcomes. The ranking model applied in this study helps clarify which ETFs achieved superior overall performance based on a balanced evaluation of return and risk. The findings highlight the importance of multi-metric analysis when selecting ETFs for long-term investment strategies. Additionally, the study considers how ETF performance insights can inform the development of sustainability-aligned investment products, particularly in the context of growing ESG integration.

Keywords: exchange-traded funds, performance, risk-adjusted return, Sharpe ratio, Treynor ratio, alpha, beta, assets under management, ranking; large-cap equities.

1. Introduction

Exchange-traded funds have become one of the most significant innovations in modern financial markets, providing investors with low-cost, transparent, and diversified exposure to a wide range of asset classes. Since their inception in the early 1990s, ETFs have grown rapidly in both number and total assets under management, driven by demand from retail investors, institutional asset managers, and financial advisors alike. Their popularity stems from a combination of liquidity, tax efficiency, and ease of trading, making them a core component of many contemporary investment strategies.

As ETFs play an increasingly central role in portfolio construction and asset allocation, assessing their performance is vital. Understanding which ETFs deliver consistent returns, maintain efficient tracking of their benchmarks, and offer favourable risk-adjusted outcomes is essential for both individual investors aiming

* Corresponding author, arunas.mockus@knf.stud.vu.lt
Received April 2025; Revised May 2025; Accepted June 2025.

to build long-term wealth and institutional managers focused on risk control and return optimization.

In light of global concerns about sustainable development and responsible investing, ETF performance also carries implications for sustainability. The growing adoption of ESG-oriented ETFs and the integration of environmental, social, and governance criteria into fund selection highlight the need to evaluate how ETFs support sustainable investment goals. As capital increasingly flows into ESG-aligned products, understanding ETF performance becomes not only a financial concern but also a tool for promoting sustainable finance.

Despite a growing body of literature on ETF performance, much of the existing research focuses on specific asset classes, short-term returns, or niche market segments. Less attention has been paid to the comparative performance of the largest ETFs by AUM—those that dominate investor portfolios and influence market behaviour. Furthermore, there is limited synthesis on how these leading ETFs perform across different market conditions and investment styles.

This paper aims to address that gap by comparing the performance of top U.S.-domiciled exchange-traded funds based on assets under management. The analysis focuses on large-cap equity ETFs and evaluates long-term return, volatility, and risk-adjusted metrics. By examining these high-AUM funds, the study offers insights into which ETFs have consistently delivered superior performance and why they continue to attract substantial investor capital.

2. Literature review

2.1. Rise and Structural Advantages of ETFs

The performance of exchange-traded funds has been the subject of extensive academic and industry research, particularly as these investment vehicles have grown to become a dominant force in global financial markets. Since their launch in the early 1990s, ETFs have evolved from niche products designed for institutional arbitrage strategies to widely used investment tools accessible to both retail and institutional investors. Their structural advantages—such as intraday tradability, portfolio diversification, low management fees, and transparency—have contributed to their rising popularity, especially in the wake of increasing scrutiny of actively managed mutual funds and the broader shift toward passive investing strategies.

A key reason for their appeal lies in the comparative cost-efficiency and tax advantages they offer. ETFs typically maintain lower expense ratios than mutual funds, largely because most are passively managed and designed to replicate the performance of benchmark indices such as the S&P 500 or the MSCI World Index. This passive structure requires less intensive management, thereby reducing operational costs. Furthermore, the unique creation and redemption mechanism of ETFs—where market participants can exchange ETF shares for the underlying assets through in-kind transactions—minimizes the capital gains distributions that often burden mutual fund investors. Studies such as those by Elton et al. (2002) and Appel

(2003) confirm that this tax efficiency, combined with competitive net returns, has significantly contributed to ETF inflows in recent decades.

Beyond cost and tax considerations, performance studies have highlighted the relative effectiveness of ETFs in replicating index returns and offering liquidity across a range of market conditions (Elton et al., 2002; Gastineau, 2001). ETFs provide investors with real-time pricing and the flexibility to implement a wide range of trading strategies, including hedging, arbitrage, and sector rotation (Madhavan, 2016). This makes them especially attractive in volatile or rapidly changing market environments. In addition, their broad exposure capabilities allow investors to gain access to specific market segments, sectors, asset classes, and geographies, often with a single trade. For instance, an investor can quickly diversify into emerging markets, fixed income, or commodity sectors without the need for complex portfolio construction.

Moreover, ETFs are increasingly used as building blocks in sophisticated portfolio strategies by institutional investors such as pension funds, insurance companies, and hedge funds (Agapova, 2011). These institutions leverage the transparency and liquidity of ETFs to manage asset allocation dynamically and cost-effectively. As a result, ETFs are no longer seen merely as retail products but as foundational instruments within multi-asset portfolios.

The literature also emphasizes the role of ETFs in enhancing market efficiency. By facilitating arbitrage and reducing pricing discrepancies between the fund and its underlying assets, ETFs contribute to more accurate price discovery in the markets they track (Israeli, Lee, & Sridharan, 2017). This mechanism, however, is not without risks—particularly in times of extreme volatility, when liquidity constraints can widen bid-ask spreads and tracking errors. Still, on balance, the academic consensus supports the view that ETFs provide a highly efficient, flexible, and cost-effective means of gaining diversified market exposure.

Given this confluence of advantages—low fees, tax efficiency, real-time trading, and reliable performance—ETFs have attracted trillions of dollars in global assets under management. As of the mid-2020s, the ETF market encompasses thousands of funds across equity, fixed income, commodity, and multi-asset classes, serving the needs of an increasingly diverse investor base. These developments underscore the importance of closely examining ETF performance, particularly among the largest funds by AUM, to understand how these vehicles deliver value and shape investment outcomes.

2.2. ETFs vs Actively Managed Funds

Much of the early literature on ETFs centers around comparisons with actively managed mutual funds, particularly with regard to return performance, volatility, and cost. These studies consistently find that ETFs, especially those designed to passively track broad market indices, tend to outperform their actively managed counterparts over extended periods, especially after adjusting for fees and expenses. This outcome is largely attributable to the inherent cost advantages of passive

investing, including lower management fees, reduced turnover, and fewer transaction costs. For instance, Blitz and Huij (2012) conducted an extensive evaluation of global equity mutual funds versus ETFs and found that passive index-tracking ETFs frequently outperform active funds on a net return basis, especially within the large-cap equity segment. This performance differential is particularly stark in efficient markets like the U.S., where the opportunity to generate alpha is limited by the widespread availability of information and the intense competition among institutional investors. These findings reinforce a broader academic and industry consensus: active managers often fail to consistently deliver excess returns net of fees, especially in market segments where informational advantages are minimal or short-lived. The cumulative evidence has prompted a shift in investor preferences toward low-cost, passive ETF strategies as core portfolio holdings.

2.3. Common Performance Metrics and Tracking Accuracy

In evaluating ETF performance, researchers typically rely on a suite of risk-adjusted and absolute performance metrics to capture both return potential and downside exposure. Among the most commonly used are the Sharpe ratio, which measures excess return per unit of total risk; the Treynor ratio, which evaluates returns relative to market risk (beta); and the information ratio, which assesses returns over a benchmark relative to tracking error (Sharpe, 1994; Roll, 1978). These ratios help investors and analysts distinguish between funds that achieve higher returns through risk-taking versus those that deliver superior risk-adjusted performance (Elton, Gruber, & Blake, 2003; Ferri, 2009).

In addition to these, alpha and beta from the Capital Asset Pricing Model (CAPM) are also widely used to evaluate performance. Alpha measures the extent to which an ETF outperforms or underperforms its expected return given its market exposure, while beta captures the ETF's sensitivity to movements in the overall market. A high beta implies higher market-related risk, while a positive alpha indicates superior manager or strategy performance (Fama & French, 2004). Together, these metrics provide insight into whether an ETF's returns are driven by market exposure or by skill-based excess returns.

Standard deviation remains a cornerstone measure of total risk, offering a view of an ETF’s return volatility over time. Higher standard deviation implies greater uncertainty in returns, which can either reflect aggressive growth potential or heightened downside risk (Sharpe et al., 1999). While high volatility is not inherently negative, it requires adequate compensation through excess returns to be justified.

Another critical performance measure is tracking error—the standard deviation of the difference between an ETF’s returns and those of its benchmark. This metric reflects how accurately an ETF replicates its target index and is especially important for investors who use ETFs as passive tools for precise market exposure. DeFusco et al. (2011) highlight the significance of tracking error in evaluating ETF quality, noting that while most ETFs follow their indices closely, deviations can occur due to liquidity constraints, management practices, or differences in fund structure. Persistent or

high tracking error can erode investor confidence and dilute the intended portfolio exposure, particularly in asset classes where replication is more challenging (Rompotis, 2006).

Finally, total return and average annual return serve as the foundation for most performance evaluations. While total return reflects the cumulative growth of an investment over time, the average annual return (typically expressed as a simple arithmetic mean of daily or monthly returns annualized) offers an accessible summary of the fund's performance over a typical year. Though not adjusted for compounding, this method remains useful in comparative studies and as an input into risk-adjusted measures like the Sharpe and Treynor ratios (Sharpe et al., 1999; Hill et al., 2015).

2.4. Sectoral and Regional ETF Comparisons

Beyond broad-market comparisons, sectoral and regional studies have provided deeper insights into how ETF performance varies across geographies, industries, and market structures. Rompotis (2006), in his analysis of European equity ETFs, found notable differences in return behavior and tracking accuracy among sector-specific funds, suggesting that ETF performance is highly sensitive to the underlying market segment. This variance is often amplified in less liquid or more volatile sectors, such as small-cap equities or emerging technologies, where replication becomes more complex and transaction costs can significantly erode returns (Madhavan, 2016). Sector ETFs that concentrate on industries like biotechnology, clean energy, or cybersecurity tend to experience higher tracking error and risk, often reflecting greater idiosyncratic exposure and less stable index constituents (Charupat & Miu, 2013).

Similarly, research into emerging market ETFs—such as the study by Drenovak et al. (2012)—has revealed that while these instruments offer important diversification benefits, they also come with heightened volatility, greater currency exposure, and more pronounced tracking discrepancies compared to their developed market counterparts. These disparities are frequently attributed to less transparent financial markets, regulatory inconsistencies, and geopolitical risks, all of which complicate index replication and reduce the ability of ETFs to precisely mirror benchmark returns (Ben-David, Franzoni, & Moussawi, 2018). Moreover, lower liquidity in the underlying securities often results in wider bid-ask spreads, contributing to additional investor costs and execution slippage (Deville, 2008).

Even among developed markets, regional ETFs can differ significantly in their performance consistency, cost structure, and volatility profiles, depending on the degree of market integration, economic stability, and industry composition of the region being tracked. For instance, ETFs focused on the Asia-Pacific or Eurozone economies may display performance variations linked to interest rate policy, export dependency, or local macroeconomic shocks (Hill et al., 2015).

As a result, investors considering sector or region-specific ETFs must weigh the trade-off between potential diversification and increased risk or inefficiency. While

these funds can provide targeted exposure and enhance portfolio diversification, they demand greater due diligence in terms of fund structure, replication method (physical vs. synthetic), and market conditions. Despite these challenges, the growing availability of ETFs across global markets and asset classes continues to broaden access for retail and institutional investors alike, while raising important questions about how these products behave in different macroeconomic and financial environments (Madhavan, 2016; Hill et al., 2015).

2.5. ESG ETFs and sustainability

The rise of environmental, social, and governance (ESG)-focused exchange-traded funds represents a significant shift in passive investing. Unlike traditional index funds, ESG ETFs aim to align investment strategies with sustainability and ethical considerations, appealing especially to socially conscious investors. Their growth has been driven by increased awareness of climate change, corporate responsibility, and stakeholder-focused capitalism (Giese et al., 2019; Pastor, Stambaugh, & Taylor, 2021). ESG ETFs have gained popularity not only among individual investors but also with institutional asset managers seeking to meet environmental mandates and long-term sustainability goals (Amel-Zadeh & Serafeim, 2018; Schramade, 2016). Amel-Zadeh and Serafeim (2018) found that investors increasingly integrate ESG data into their decision-making not just for ethical reasons but also to manage risk, reduce long-term uncertainty, and uncover value drivers not captured in traditional models.

While ESG ETFs have attracted strong inflows, their financial performance remains mixed. Some outperform due to sector tilts—particularly toward technology—while others underperform after excluding profitable but controversial industries. Several studies suggest that ESG strategies can provide downside protection during crises but may lag during bull markets. For example, Broadstock et al. (2021) and Albuquerque et al. (2020) observed that ESG-aligned firms and funds exhibited greater resilience and lower volatility during the COVID-19 market downturn, indicating a potential risk mitigation role for ESG assets. However, this outperformance is not consistent across all ESG ETFs, which raises concerns about strategy design and benchmark selection.

A key challenge lies in the inconsistency of ESG definitions and screening criteria, which can differ significantly between providers and lead to divergent portfolio outcomes (Berg, Kölbel, & Rigobon, 2022; Kotsantonis & Serafeim, 2019). Christensen, Serafeim, and Sikochi (2022) demonstrate that disagreement among ESG rating agencies can be substantial, making it difficult for investors to identify high-quality sustainable investments. This lack of standardization complicates product comparison and may even lead to misallocation of capital if investors rely on flawed or misaligned ESG scores. Furthermore, ESG ratings are often based on voluntary disclosures, which can be biased or incomplete, especially in emerging markets or among smaller firms.

In addition to definitional inconsistencies, ESG scoring methodologies often rely on incomplete or subjective data, which further complicates performance evaluation.

As a result, two ESG funds tracking the same general theme may hold vastly different portfolios, leading to different exposures, risks, and returns. This variability has prompted calls for more transparent and harmonized frameworks for ESG index construction (Berg et al., 2022; De Franco et al., 2020). De Franco et al. (2020) also emphasize that the methodology used to filter ESG assets - whether based on simple screening or advanced machine learning - can materially affect the financial and ethical quality of ESG ETFs. Without clearer and standardized methodologies, the ability of ESG ETFs to deliver both ethical alignment and financial outperformance remains uncertain.

For investors, ESG ETFs represent both an opportunity and a challenge. On one hand, they offer a way to integrate values into portfolio construction without sacrificing the benefits of ETF liquidity and cost-efficiency. On the other, they require greater due diligence, particularly in assessing how ESG criteria are applied and how consistently those standards are maintained. Boffo and Patalano (2020) argue that ESG investing is at a crossroads: while investor demand is high, weak comparability and fragmented regulatory guidance risk undermining confidence in ESG-labelled financial products. As ESG ETFs become increasingly mainstream, aligning performance analysis with sustainability goals will be essential for ensuring that these products deliver both financial and societal value.

In the context of this study, sustainability considerations are increasingly relevant. As ESG principles become integrated into mainstream investment strategies, understanding how traditional high-AUM ETFs compare in risk-adjusted performance to ESG-focused alternatives becomes essential. Although this paper does not focus exclusively on ESG ETFs, the metrics used—such as Sharpe ratio, alpha, and volatility—are equally applicable to evaluating the financial viability of sustainability-aligned funds. Moreover, the methodologies and insights developed through performance analysis of conventional ETFs can inform the design, benchmarking, and evaluation of ESG funds, helping investors make informed choices in an evolving financial landscape. This intersection between performance evaluation and sustainable finance underscores the growing need for holistic metrics that capture both financial return and societal impact.

2.6. Focus of This Study: Comparing the Top ETFs by AUM

Given the central role that exchange-traded funds play in global investment portfolios, it is valuable to assess how the largest funds—those with the highest assets under management — perform in practice. These ETFs attract substantial investor capital and often serve as foundational components in both retail and institutional strategies. Their scale, liquidity, and popularity make them particularly relevant for performance evaluation.

This study focuses on a comparative analysis of the top ETFs by AUM in the US market over a fifteen-year period. Rather than examining a broad range of asset classes, sectors, or market conditions, the objective is to assess how these flagship funds perform relative to one another using standard measures of return and risk.

The analysis applies widely recognized performance metrics to evaluate risk-adjusted outcomes and provide insights into consistency, volatility, and efficiency across the selected funds. By comparing these core ETFs using well-established return and risk metrics, the study offers a concise yet meaningful overview of how the largest funds stack up against one another in terms of both absolute and risk-adjusted performance.

3. Methods

This study conducts a comparative performance analysis of the largest exchange-traded funds by assets under management over the period 2010–2024. The aim is to evaluate and rank these funds using well-established performance and risk metrics. Data was sourced from Yahoo Finance, which provided historical daily adjusted closing prices. The study excludes dividends and focuses solely on price returns, ensuring consistency across funds and simplifying comparisons, particularly where dividend reinvestment schedules may vary or be inconsistently reported.

3.1 Sample and Data Frequency

The sample consists of the top U.S.-domiciled ETFs by assets under management within the equity asset class, with a focus on large-capitalization companies. The ETFs were selected based on recent fund rankings and exclude those that directly track the performance of the S&P 500 index. These ETFs represent some of the most traded and widely held investment products, with extensive historical data and global investor reach. All ETFs in the sample were actively traded for the full duration of the study, ensuring data integrity and comparability.

The sample criteria were chosen to reflect a focused yet diverse group of highly influential investment vehicles in the U.S. equity market. By selecting U.S.-domiciled ETFs with the highest assets under management, the study targets funds that are most widely held and actively traded, ensuring both data availability and real-world relevance. High-AUM ETFs tend to have lower expense ratios, tighter bid-ask spreads, and more robust investor participation, which collectively enhance their practical appeal and analytical value (Elton et al., 2014).

Focusing exclusively on the equity asset class—and within that, on large-capitalization companies—provides consistency in terms of investment style, liquidity, and volatility. Large-cap stocks are generally more stable and represent a significant portion of institutional portfolios, making them ideal for comparative performance analysis (Bodie et al., 2014). The decision to exclude ETFs that directly track the S&P 500 index was made to avoid redundancy and allow for the inclusion of funds with distinct construction methodologies, such as growth/value tilts, factor-based strategies, or alternative large-cap indices. This enhances the comparative dimension of the analysis and avoids over-concentration in a single benchmark approach, which could bias the results.

Returns are calculated daily based on closing prices, which provides a high-resolution view of ETF performance and better captures the volatility and market

sensitivity of each fund. The use of daily data increases the precision of metrics such as standard deviation, alpha, and beta, which can be distorted or understated when based on monthly or annual figures (Elton et al., 2014). Since the analysis excludes dividends, the calculated returns reflect pure price appreciation, offering a clean measure of capital performance and avoiding inconsistencies in dividend timing, frequency, or reinvestment assumptions.

The S&P 500 index is used as the benchmark for calculating alpha, beta, and related risk-adjusted metrics. This index is widely regarded as a standard proxy for the U.S. equity market and provides a consistent baseline across all funds, many of which are heavily influenced by U.S. large-cap equity performance (Fama & French, 1992).

3.2 Performance Metrics

This section details the performance metrics used in the analysis. Each metric is grounded in academic theory and widely used in professional fund evaluation.

Total return represents the overall percentage change in an investment's value over a given period (Bodie et al., 2014):

$$Total\ Return = \frac{P_{end} - P_{start}}{P_{start}} * 100, \quad (1)$$

where P_{end} is the ETF's closing price at the end of the study period and P_{start}: is the ETF's closing price at the beginning of the study period.

In this study, it reflects the change in ETF's market price from 2010 to 2024, excluding dividends. This metric is intuitive and straightforward, making it a popular first-level indicator of fund performance in both academic and practitioner analyses (Bodie, Kane, & Marcus, 2014). Although it does not account for reinvested income or risk, it is essential for understanding the absolute return potential of an asset and forms the foundation for more advanced measures like compound returns and risk-adjusted performance.

Despite its simplicity, total return should not be interpreted in isolation. It lacks context regarding the volatility or consistency of returns and may be misleading for comparing funds with differing risk profiles. For example, two ETFs may generate similar total returns over a period, but one may do so with significantly higher volatility or exposure to drawdowns. Therefore, while total return provides an important measure of growth, it must be supplemented with other metrics to fully assess fund quality (Elton, Gruber, Brown, & Goetzmann, 2014).

The average annual return using the arithmetic (simple) method is a widely used measure to summarize an asset's average performance over a period, based on daily or periodic return data (Bodie et al., 2014):

$$Average\ Annual\ Return = \frac{1}{n}\sum_{t=1}^{n} r_t * 252, \quad (2)$$

where r_t is the return on day t, n is the total number of trading days in the sample, 252 is the average number of trading days in a year. It is calculated by taking the mean of daily returns and annualizing it by multiplying by the number of trading days in a year—typically 252 in the case of equity markets.

This approach provides a non-compounded estimate of the average return that an asset or portfolio generates on an annual basis. It is particularly useful in descriptive statistics and serves as the foundation for many financial ratios, such as the Sharpe ratio, which evaluates risk-adjusted performance (Sharpe et al., 1999).

According to Sharpe, Alexander, and Bailey (1999), the arithmetic average is appropriate when assessing expected return under the assumption of independent and identically distributed returns. It reflects what an investor might expect to earn in an average year, assuming each year is statistically similar to the historical sample. In practice, this measure is often used in conjunction with volatility metrics to form the basis for performance evaluation and portfolio optimization under modern portfolio theory.

However, the average annual return does not capture the effects of compounding or sequence of returns. As such, it may overstate long-term investment performance, particularly in the presence of high volatility. For investment decisions involving multi-period holding strategies, the compound annual growth rate (CAGR) is typically preferred. Nonetheless, in cross-sectional ETF performance comparisons or in constructing risk-adjusted ratios like Sharpe and Treynor, the simple average return remains analytically valid and interpretable (Strong, 2003).

Standard deviation quantifies the dispersion of returns around the mean, indicating how consistently an ETF has performed over time (Bodie et al., 2014):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(R_i - \bar{R})^2}{n-1}}, \tag{3}$$

where σ is standard deviation of returns, R_i is daily return at time i, $\bar{R}$ is the mean return over the period, n is the number of daily observations. This measure is the most widely used statistical measure of an investment's total risk. A high standard deviation implies greater volatility, while a low standard deviation suggests more stable and predictable returns. In portfolio theory, it represents total (systematic + unsystematic) risk and is a cornerstone of modern portfolio analysis (Markowitz, 1952; Elton et al., 2014).

While standard deviation is a powerful tool, it does not differentiate between upside and downside volatility. As such, it may penalize ETFs that experience large positive deviations from the mean, even though these may be desirable. This limitation has prompted the development of additional measures such as downside deviation or the Sortino ratio. Nonetheless, standard deviation remains a core metric in performance evaluation and is critical to calculating other risk-adjusted metrics such as the Sharpe ratio (Bodie et al., 2014).

Alpha measures an ETF's performance relative to what would be expected based on its exposure to the overall market, as measured by beta (Bodie et al., 2014):

$$\alpha = R_i - [R_f + \beta * (R_M - R_f)], \tag{4}$$

where R_i is the average return of the ETF, R_f is risk-free rate, R_m is market return (e.g., S&P 500), β is ETF's sensitivity to market movements. A positive alpha indicates that the ETF has generated excess returns not explained by market movements alone, potentially reflecting superior management, structural advantages, or factor

exposures. Originally proposed by Jensen (1968), alpha remains one of the most cited performance attribution measures in academic finance.

Although alpha is appealing as a measure of skill or inefficiency, it is highly sensitive to the specification of the benchmark and the accuracy of beta estimation. Misalignment in the chosen index can distort results. Furthermore, persistent alpha is difficult to sustain in efficient markets, as suggested by Fama and French (1992), who argue that most excess returns are explained by risk factors or chance. Despite these criticisms, alpha remains integral to evaluating whether a fund offers value beyond passive exposure.

Beta assesses the systematic risk of an ETF by measuring how much it moves relative to a benchmark market index (Bodie et al., 2014):

$$\beta = \frac{Cov(R_i,R_m)}{Var(R_m)}, \quad (5)$$

where Cov(R_i,R_m) is covariance between ETF and market returns, Var(R_m) is variance of market returns. A beta of 1 implies that the ETF tends to move in lockstep with the market. A beta greater than 1 indicates more amplified responses to market movements (i.e., higher volatility), while a beta below 1 implies a more conservative or defensive posture (Sharpe, 1964; Fama & French, 1992).

Understanding beta is essential in portfolio construction because it enables investors to manage exposure to market-wide shocks. It is particularly relevant when comparing ETFs that track different asset classes or regions. However, beta captures only market-related risk and ignores unsystematic (diversifiable) risk. For this reason, it is most effective when evaluating ETFs as part of diversified portfolios, where idiosyncratic risk is minimized (Elton et al., 2014).

The Sharpe ratio measures the efficiency of return per unit of total risk (Bodie et al., 2014):

$$Sharpe = \frac{R_i - R_f}{\sigma}, \quad (6)$$

where R_i is average ETF return, R_f is risk-free rate, σ is standard deviation of ETF returns. This measure is widely regarded as the gold standard for evaluating risk-adjusted performance, especially in academic and professional contexts. A higher Sharpe ratio suggests that the ETF is delivering more excess return (above the risk-free rate) for each unit of risk taken (Sharpe, 1994). It is particularly useful for investors evaluating single funds without broader portfolio context.

The Sharpe ratio assumes that all volatility is undesirable, which makes it ideal for standalone fund assessments but less effective for diversified portfolios where only systematic risk matters. Another limitation is that it assumes returns are normally distributed; in practice, many financial assets exhibit skewness or kurtosis, which may distort the ratio. Nevertheless, it remains an indispensable tool in fund comparison and is widely used in fund ratings and institutional due diligence processes (Bodie et al., 2014; Strong, 2003).

The Treynor ratio, developed by Jack Treynor (1965), provides a measure of risk-adjusted return based solely on systematic (market) risk (Bodie et al., 2014):

$$Treynor = \frac{R_i - R_f}{\beta}, \quad (7)$$

where R_i is average return of the ETF, R_f is risk-free rate, β is beta of the ETF relative to the benchmark. Unlike the Sharpe ratio, which uses total volatility, the Treynor ratio assumes that unsystematic risk has been diversified away. It is best suited for evaluating ETFs that are components of larger, well-diversified portfolios.

The Treynor ratio is particularly valuable when comparing ETFs that serve as strategic building blocks within broader asset allocations. It reveals how much return is being generated for each unit of market risk undertaken. Like alpha, the Treynor ratio's effectiveness depends on accurate beta estimation and appropriate benchmark selection. Still, it remains a foundational element of portfolio performance analysis, especially in institutional settings (Sharpe et al., 1999; Elton et al., 2014).

3.3 Ranking Methodology

To compare ETF performance across multiple dimensions, this study uses a point-based scoring system. Each ETF is assigned a score for each metric based on its rank among the top funds:

- 1 point for the best performance (e.g., highest Sharpe ratio, lowest standard deviation)
- 8 points for the worst performance
- Scores from all metrics are summed for each ETF

ETF with the lowest total score is identified as the top overall performer. This model assumes equal weighting across metrics to avoid subjective emphasis on return versus risk. While weighting returns more heavily might be justified in certain investor profiles, equal weighting reflects a balanced evaluation, aligning with modern portfolio theory, which stresses the joint importance of return and risk (Markowitz, 1952).

Equal weighting also increases transparency and methodological neutrality, avoiding biases that could arise from arbitrary weight assignment. It ensures that each dimension of performance—growth, volatility, market sensitivity, and efficiency—contributes equally to the final ranking.

4. Results

This section presents a comprehensive evaluation of the performance of eight U.S.-domiciled large-cap equity ETFs over the period 2010–2024. The analysis incorporates both absolute and risk-adjusted return metrics based on daily price data. By comparing total return, compound annual growth rate (CAGR), standard deviation, beta, alpha, Sharpe ratio, and Treynor ratio, this study offers a detailed view of how these ETFs performed across various dimensions of portfolio performance (Table 1).

Table 1
ETF Performance Metrics (2010-2024)

ETF	Total Return	Avg Annual Return	Standard Deviation	Alpha	Beta	Sharpe Ratio	Treynor Ratio
QQQ	10.08	0.1818	0.2046	0.1258	0.3183	0.7708	0.4955
IWF	6.7809	0.156	0.1934	0.1001	0.3166	0.6815	0.4164
FTCS	2.533	0.0981	0.1655	0.0573	0.1668	0.4472	0.4435
MGK	7.1407	0.1581	0.1885	0.1065	0.2732	0.7103	0.4901
VUG	2.253	0.0936	0.1711	0.0508	0.1858	0.4058	0.3737
VV	4.33	0.1269	0.1727	0.0779	0.2477	0.5948	0.4146
VTV	7.267	0.1602	0.1944	0.1021	0.3393	0.7	0.401
IWD	3.2446	0.1098	0.162	0.0676	0.1799	0.529	0.4762

Source: designed by the author.

Absolute performance. ETF with the highest total return over the 15-year period was the Invesco QQQ Trust (QQQ). This fund, which tracks the Nasdaq-100 Index, benefited significantly from its exposure to high-growth technology stocks and large-cap innovators such as Apple, Microsoft, and Amazon. QQQ's performance reflects the broader dominance of growth-oriented strategies throughout the 2010s and early 2020s.

The Vanguard Mega Cap Growth ETF (MGK), Vanguard Growth ETF (VUG) and iShares Russell 1000 Growth ETF **(**IWF) also delivered strong absolute returns, closely following QQQ. Both funds focus on large-cap growth companies and have substantial overlap with QQQ in sector allocation, particularly in technology and consumer discretionary stocks. These funds achieved total returns well in excess of 500% over the sample period, underscoring the strength of growth equity strategies during this era.

On the other hand, the First Trust Capital Strength ETF (FTCS), iShares Russell 1000 Value ETF (IWD) and Vanguard Value Index Fund ETF Shares (VTV) posted the lowest total returns in the sample. Although FTCS is constructed to emphasize financial stability and balance sheet strength, this defensive orientation resulted in reduced exposure to the growth themes that propelled the market over the last decade. IWD, with its value tilt, similarly underperformed during a period when value stocks lagged behind their growth counterparts.

Average Annual Return and Volatility. The pattern of average annual returns closely mirrored total returns. QQQ again led with an average annual return of over 18%, followed by MGK, VUG and IWF, which each posted returns above 15%. These results highlight the consistency and magnitude of returns achieved by growth-focused ETFs during a long-term bull market.

Volatility, as measured by standard deviation, revealed important differences in risk profiles across ETFs. QQQ exhibited the highest volatility in the sample, with an annualized standard deviation exceeding 20%, reflecting the high beta nature of its tech-heavy portfolio. In contrast, ETFs such as VTV, FTCS, and IWD showed considerably lower volatility, ranging from 15% to 17% annually. These lower-volatility funds may appeal to risk-averse investors, though their lower returns present a trade-off.

Risk-Adjusted Performance. While QQQ dominated in absolute returns, it also excelled in risk-adjusted performance. Its Sharpe ratio — a measure of return per unit of total volatility — was the highest among all ETFs, indicating that it offered the best reward relative to risk. Its Treynor ratio, which considers only market-related (systematic) risk, was also the strongest, underscoring QQQ's efficiency in delivering returns given its beta exposure.

MGK, FTCS and IWF also performed strongly on risk-adjusted metrics, with high Sharpe and Treynor ratios that indicate strong compensation for both total and market-specific risk. These findings suggest that the best-performing growth ETFs were not only delivering high returns but were doing so efficiently in terms of risk.

Conversely, IWD and VTV scored poorly in both Sharpe and Treynor ratios. While they offered a lower-volatility ride, their returns were insufficient to justify even their more conservative risk profiles. These results are consistent with broader market narratives over the past decade, where value and defensive strategies struggled to keep pace with the high-growth segment of the market.

Composite Ranking. To integrate all performance aspects into a single evaluation, each ETF was scored from 1 to 8 on every metric, with the best fund receiving the lowest total score. Rankings were reversed for metrics where lower is better (standard deviation and beta), and total scores were summed to calculate a composite performance score (Table 2).

Table 2
Ranking ETF performance

ETF	Total Return Rank	Avg Annual Return (Simple) Rank	Standard Deviation Rank	Alpha Rank	Beta Rank	Sharpe Ratio Rank	Treynor Ratio Rank	Total Score	Overall Rank
QQQ	1	1	8	1	7	1	1	20	1
IWF	3	3	5	2	5	2	2	22	2
FTCS	6	6	1	6	2	6	3	30	3
MGK	2	2	7	3	8	3	7	32	4
VUG	4	4	6	4	6	4	5	33	5
VV	5	5	4	5	4	5	6	34	6
VTV	7	7	2	7	1	7	4	35	7
IWD	8	8	3	8	3	8	8	46	8

Source: designed by the author.

This ranking system revealed that Invesco QQQ Trust (QQQ) was the top overall performer, excelling in both return and risk-adjusted categories. It consistently outperformed its peers not only in raw numbers but also in how efficiently it translated risk into return.

The runners-up were IWF and FTCS, both of which demonstrated excellent balance between return and risk. These ETFs offer compelling options for investors seeking exposure to large-cap growth with slightly more diversification than QQQ. At the bottom of the ranking were VTV and IWD. They may be suitable for very conservative investors, but over the 15-year horizon studied here, they underperformed compared to growth-oriented peers.

5. Conclusions

This study provided a comparative analysis of the largest U.S.-domiciled, large-cap equity ETFs by assets under management over the period 2010–2024. Using a range of absolute and risk-adjusted performance metrics—including total return, average annual return, standard deviation, beta, alpha, Sharpe ratio, and Treynor ratio—the research evaluated how these widely held funds perform relative to one another.

The findings show that while all funds in the sample offer liquid, diversified exposure to the large-cap segment, notable differences exist in performance outcomes. Growth-oriented ETFs generally produced higher returns but also carried greater volatility and beta. In contrast, defensive or value-oriented funds displayed lower volatility, with mixed results in terms of efficiency. Importantly, some ETFs with moderate risk profiles achieved strong risk-adjusted returns, suggesting that superior performance is not exclusive to high-risk strategies.

From a policy and managerial perspective, these findings have implications for the expanding field of sustainable investing. As investors increasingly prioritize ESG performance, fund structure, benchmark tracking accuracy, and transparent methodologies become essential factors in supporting long-term sustainability goals. High-performing, low-volatility ETFs may serve as models for developing ESG-aligned products that deliver competitive financial outcomes without compromising risk control. Supporting the growth and standardization of such ETFs could facilitate more responsible capital allocation and contribute meaningfully to environmental and social objectives.

These results also underscore the importance of evaluating ETFs using multiple performance lenses rather than relying on total return alone. For long-term investors, incorporating risk-adjusted metrics into fund selection is vital for building consistent and efficient portfolios. By demonstrating the variability in performance even among high-AUM ETFs, this study reinforces the need for thorough, data-driven evaluation.

Overall, this research contributes to the understanding of ETF performance and its intersection with sustainable investing priorities. Future studies could explore similar analyses across ESG-specific ETFs or examine how investor preferences shift in response to sustainability disclosures and ratings.

References

Agapova, A. (2011). Conventional mutual index funds versus exchange-traded funds. *Journal of Financial Markets, 14*(2), 323–343.

Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. *Review of Corporate Finance Studies*, 9(3), 593–621.

Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. Financial Analysts Journal, 74(3), 87–103.

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility? *The Journal of Finance, 73*(6), 2471–2535.

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance, 26*(6), 1315–1344.

Blitz, D., & Huij, J. (2012). Evaluating the performance of global emerging markets equity ETFs. *Emerging Markets Review, 13*(2), 149–158.

Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments* (10th ed.). McGraw-Hill Education.

Boffo, R., & Patalano, R. (2020). ESG Investing: Practices, Progress and Challenges. *OECD Paris: OECD Publishing.*

Broadstock, D. C., Chan, K., Cheng, L. T., & Wang, X. (2021). The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. *Finance Research Letters*, 38, 101716.

Charupat, N., & Miu, P. (2013). The pricing and performance of leveraged exchange-traded funds. *Journal of Banking & Finance, 37*(12), 4760–4777.

Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. Accounting Review, 97(1), 147–175.

De Franco, C., Geissler, C., Margot, V., & Monnier, B. (2020). ESG investments: Filtering versus machine learning approaches.

DeFusco, R. A., Ivanov, S. I., & Karels, G. V. (2011). The exchange-traded funds' price deviation from net asset value and bid-ask spreads. *Journal of Economics and Finance, 35*(2), 181–197. DOI:10.1007/s12197-009-9090-6

Deville, L. (2008). Exchange traded funds: History, trading, and research. *SSRN Electronic Journal*. DOI:10.1007/978-0-387-76682-9_4

Drenovak, M., Urošević, B., & Jelic, R. (2012). European bond ETFs: Tracking error and performance. *Journal of Banking & Finance, 36*(6), 1609–1618.

Elton, E. J., Gruber, M. J., & Blake, C. R. (2003). Incentive fees and mutual funds. *Journal of Finance, 58*(2), 779–804. https://doi.org/10.1111/1540-6261.00545

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). *Modern Portfolio Theory and Investment Analysis* (9th ed.). Wiley.

Elton, E. J., Gruber, M. J., Comer, G., & Li, K. (2002). Spiders: Where are the bugs? *Journal of Business, 75*(3), 453–472. DOI:10.1007/3-540-27637-8_4

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance, 47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and evidence. *Journal of Economic Perspectives, 18*(3), 25–46. DOI: 10.1257/0895330042162430

Ferri, R. A. (2009). *The ETF Book: All You Need to Know About Exchange-Traded Funds*. Wiley.

Gastineau, G. L. (2001). An introduction to exchange-traded funds. *Journal of Portfolio Management, 27*(3), 88–96. DOI: 10.3905/jpm.2001.319804

Giese, G., Lee, L. E., Melas, D., Nagy, Z., & Nishikawa, L. (2019). Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *Journal of Portfolio Management, 45*(5), 69–83. DOI: 10.3905/jpm.2019.45.5.069

Hill, J. M., Nadig, D., & Hougan, M. (2015). *A Comprehensive Guide to Exchange-Traded Funds*. Wiley.

Israeli, D., Lee, C. M. C., & Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? An information perspective. *Review of Accounting Studies, 22*, 1048–1083. http://dx.doi.org/10.2139/ssrn.2625975

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance, 23*(2), 389–416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x

Appel, M.. (2006). *Investing in Exchange Traded Funds Made Easy Higher Returns With Lower Costs-do It Yourself Strategies Without Paying Fund Managers*. FT Press.

Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance, 31*(2), 50–58.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77–91. https://doi.org/10.2307/2975974

Madhavan, A. (2016). *Exchange-Traded Funds and the New Dynamics of Investing*. Oxford University Press.

Pastor, L., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics, 142*(2), 550–571. DOI: 10.1016/j.jfineco.2020.12.011

Rompotis, G. G. (2006). Performance and trading characteristics of iShares. *SSRN Electronic Journal. DOI:10.2139/ssrn.946732*

Roll, R. (1978). Ambiguity when performance is measured by the securities market line. *Journal of Finance, 33*(4), 1051–1069. https://doi.org/10.2307/2326939

Schramade, W. (2016). Integrating ESG into valuation models and investment decisions: The value-driver adjustment approach. Journal of Sustainable Finance & Investment, 6(2), 95–111.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance, 19*(3), 425–442. https://doi.org/10.2307/2977928

Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management, 21*(1), 49–58. DOI: 10.3905/jpm.1994.409501

Sharpe, W. F., Alexander, G. J., & Bailey, J. V. (1999). *Investments* (6th ed.). Prentice Hall.

Strong, R. A. (2003). *Practical Investment Management* (3rd ed.). Cengage Learning.

Treynor, J. L. (1965). How to rate management of investment funds. *Harvard Business Review, 43*(1), 63–75.