

Hidden Concentration in ETF portfolios

Overlap, Weighting Effects, and Structural Diversification Risk

Introduction

Exchange-traded funds (ETF's) have fundamentally reshaped modern portfolio construction. Their appeal lies in simplicity, transparency, cost efficiency, and broad market access. Investors can achieve exposure to entire asset classes, sectors, or factors through a single instrument, enabling diversification strategies that were once operationally complex. As ETF adoption has accelerated, portfolio diversification has increasingly been framed at the fund level rather than at the underlying holdings level.

This evolution has introduced a subtle but important analytical gap. While ETFs are diversified vehicles individually, portfolios composed of multiple ETFs may not necessarily produce additive diversification benefits. The aggregation of diversified funds can unintentionally mask overlapping exposures, concentrated allocations, and structural redundancies. As a result, portfolios that appear diversified based on the number of funds held may exhibit significant concentration when evaluated through the lens of constituent securities.

This study investigates a central paradox within ETF portfolio construction: the presence of hidden concentration risk arising from overlapping holding and weight imbalances. The analysis moves beyond traditional holding-count comparisons and instead evaluates diversification quality using weighted overlapping metrics and concentration measures. By focusing on economic exposure rather than nominal diversification, the study aims to provide a more realistic assessment of portfolio structure.

Business Problem

The widespread adoption of ETFs has altered how investors conceptualize diversification. Portfolio construction decisions are often guided by fund classifications, investment themes, or index affiliations. Investors may combine ETFs representing growth, dividends or broad market exposure under the assumption that diversification naturally emerges through category differentiation. However, ETFs frequently share overlapping securities, particularly when tracking capitalization-weighted indices.

This dynamic creates a structural risk that is not immediately visible at the portfolio level. Overlapping holding may result in duplicated economic exposure, where the same securities exert disproportionate influence across multiple funds. In such cases, diversification becomes largely cosmetics, driven by fund labels rather than actual asset dispersion. Investors may

unknowingly amplify exposure to specific securities, sectors, or factors while believing risk has been mitigated.

The Implications of this misalignment are particularly relevant in environments characterized by elevated market concentration. As equity markets increasingly exhibit dominance by a relatively small subset of large cap securities, ETF portfolios may inherit and magnify these concentrations effects. Consequently, understanding overlap and concentration is not merely an academic exercise but a practical necessity for portfolio risk management.

Objective of the study

The primary objective of this analysis is to quantify and interpret the structural relationships between ETFs at the holdings level. Specifically, the study seeks to evaluate the magnitude of overlapping exposures, assess the economic significance of weight-adjusted overlap, and compare concentration characteristics across fund structures.

Rather than treating ETFs as discrete diversification units, the analysis reframes them as collections of weighted securities. This perspective enables a more precise evaluation of diversification quality and portfolio redundancy. By integrating overlap measurement with concentration metrics, the study provides a multidimensional understanding of ETF portfolio composition.

The analysis focuses on four widely held ETFs representing distinct investment narratives yet substantial market presence: QQQ, SCHD, SPY, and VOO. These funds collectively capture growth, dividend, and broad-market strategies, making them representative of common investor portfolio combinations.

Data & Analytics Framework

Holdings data were obtained directly from fund disclosures, ensuring alignment with actual portfolio compositions rather than derived estimates. Only constituent securities and their respective allocation weights were retained for analysis. This approach isolates the structural properties of each ETF while avoiding noise introduced by ancillary fund attributes.

The analytical workflow integrates Python, SQL, and Tableau, reflecting a hybrid quantitative and visualization-driven methodology. Python was employed for data ingestion, normalization, and transformation. SQL provided a structured environment for overlapping calculations and concentration metric derivation. Tableau served as the visualization layer, translating quantitative outputs into interpretable comparative frameworks.

Allocation weights were standardized into decimal format, and holding names were normalized to ensure consistent cross-fund matching. This normalization step is critical in overlap analysis,

as minor discrepancies in naming conventions can distort intersection calculations. The resulting dataset represents a unified holdings universe suitable for comparative evaluation.

Methodology

Overlap was measured using a weight-adjusted intersection framework designed to capture economic exposure duplication. Traditional overlap assessments often rely on holding counts or binary inclusion logic, which fail to account for allocation magnitude. In contrast, weighted overlap measurement evaluates the proportional significance of shared securities.

This methodology reflects a more economically meaningful interpretation of diversification. Two ETFs holding the same securities at materially different weights do not contribute equally to overlap risk. By incorporating allocation values, the analysis distinguishes between nominal overlap and economically dominant overlap.

Concentration metrics

Concentration was evaluated using complementary measures designed to capture both distributional structure and exposure dominance. Effective number of holdings metrics assess diversification quality beyond simple holding counts. Top allocation shares measures quantify weight concentration within leading securities. The Herfindahl-Hirschman Index (HHI) provides a formalized concentration measure widely used in competition and portfolio analysis.

Together, these metrics allow differentiation between broad holdings dispersion and weight-skewed exposure structures. This distinction is particularly important for capitalization-weighted ETFs, where diversification breadth may coexist with substantial concentration.

Key Findings

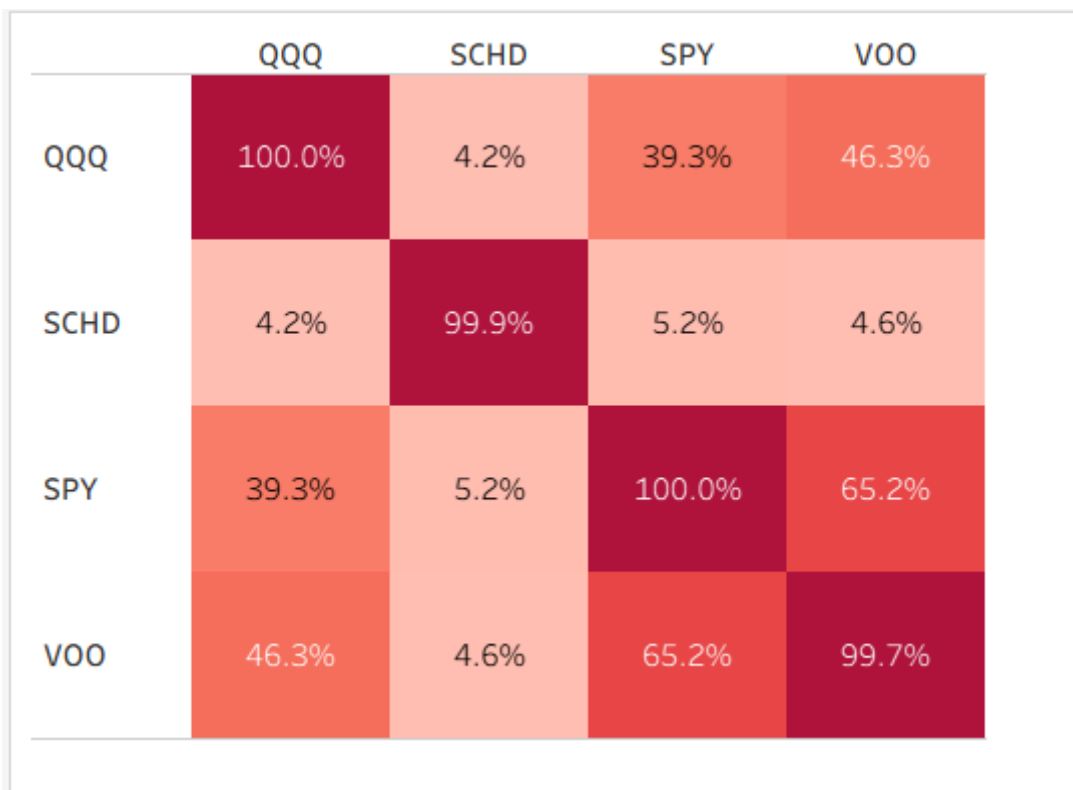
The weighted overlap matrix reveals that overlap between ETFs is not random but structurally embedded within index construction methodologies. The most prominent relationship emerges between SPY and VOO, which exhibit approximately sixty-five percent weighted overlap. This magnitude indicates that the funds function economically as near substitutes despite differing provider identities.

Similarly, substantial overlap is observed between QQQ and VOO, as well as between QQQ and SPY. These relationships underscore the pervasive influence of large-cap securities across multiple ETF strategies. Overlap is therefore not a marginal phenomenon but a defining characteristic of ETF ecosystem

Hidden concentration effects

The heatmap visualization illustrates the concentration of overlapping intensity among capitalization-weighted ETFs. While diversification is often assumed through fund stacking, the visual evidence indicates diminishing incremental diversification benefits. Combining SPY and VOO, for instance, introduces limited structural diversification despite increasing the number of funds held.

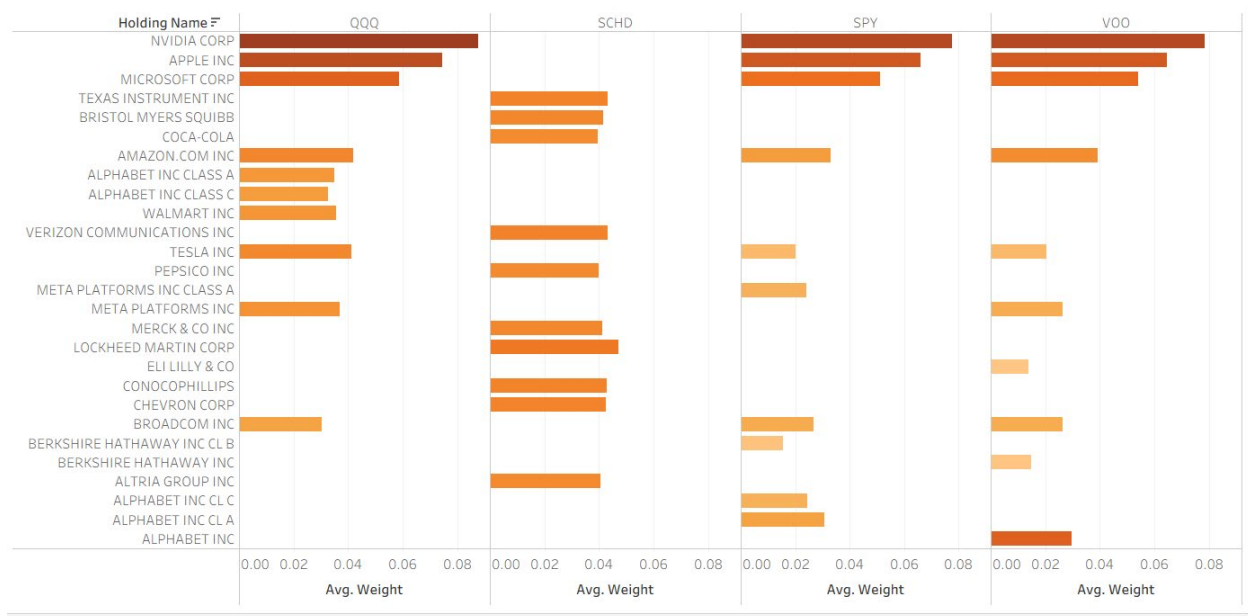
This finding highlights the distinction between fund-level diversification and holdings-level diversification. Investors may expand portfolio complexity without meaningfully reducing exposure concentration.



Weighted overlap dominance

The weighted overlap ranking chart reinforces the dominance of specific ETF pairings. SPY versus VOO overlap materially exceeds all other combinations, reflecting their shared structural foundations. The magnitude gap between leading and secondary pairings indicates that overlap risk is unevenly distributed rather than uniform across funds.

This concentration of overlap suggests that portfolio redundancy risk is highly sensitive to ETF selection rather than merely portfolio size.



Structural Concentration differences

Concentration metrics reveal meaningful differences across ETF structures. QQQ exhibits pronounced weight concentration within leading securities, reflecting its growth-oriented composition. SCHD displays a comparatively distributed weight profile, indicative of its dividend-focused methodology. SPY and VOO, while broadly diversified, retain allocation skewness inherent in market-cap weighting.

Effective holdings measures confirm that diversification quality varies significantly despite similar holding counts. Structural weighting frameworks therefore exert a stronger influence on concentration characteristics than nominal diversification breadth.

Visualization & interpretation

The weighted overlap heatmap provides a multidimensional representation of exposure duplication. Darker intensity regions signify greater economic redundancy. The near equivalence between SPY and VOO visually communicates structural similarity more effectively than numerical tables alone.

The weighted overlap ranking chart complements the matrix by translating relationships into magnitude comparisons. This visualization clarifies dominance patterns and highlights the disproportionate influence of specific ETF combinations. Together, these charts function as analytical evidence rather than supplementary graphics.

Implications for investors

The findings carry significant implications for portfolio construction philosophy. ETF diversification cannot be assumed based solely on fund variety. Structural overlap and weight concentration must be evaluated explicitly. Investors relying on fund stacking strategies may inadvertently replicate exposures, reducing diversification efficiency.

Moreover, weighting methodologies emerge as critical determinants of portfolio behavior. Funds with concentrated allocation structures may dominate portfolio risk despite representing a minority of holdings. Consequently, diversification quality is shaped more by allocation distribution than by security count.

Understanding these dynamics enables investors to differentiate between apparent diversification and effective diversification. Portfolio resilience depends not merely on asset inclusion but on structural dispersion.

Conclusion

This study demonstrates that hidden concentration risk is an inherent feature of ETF portfolios constructed without holding-level analysis. Weighted overlap measurement reveals that diversification assumptions based on fund labels may be misleading. Concentration metrics further illustrate that structural weighting frameworks materially influence portfolio exposure profiles.

ETFs remain powerful investment tools; however, their integration into portfolios requires analytical rigor beyond surface-level classification. Holdings-level evaluation provides a more economically grounded framework for assessing diversification integrity.