The Price of Passive: Valuation Distortion and Systematic Fragility in the S&P 500

Master Thesis

by

Alexander Moe Halvorsrød and Benjamin Egeland Billington

MSc in Finance

Supervisor

Ignacio Garcia de Olalla Lopez

Oslo, July 1, 2026

ABSTRACT

This thesis examines how the rise of passive investing has affected US equity market efficiency over the period 1993 to 2024. Using S&P 500 constituent data from CRSP and Compustat alongside Morningstar fund flow data, we test two hypotheses: that passive capital inflows inflate valuation multiples beyond fundamental justification, and that passive dominance increases systematic fragility through reduced price informativeness, risk concentration, and correlation asymmetry. We find evidence of a long-run cointegrating relationship between passive share and key price-to-earnings and price-to-book measures, with error correction models confirming mean reversion over horizons of four to twelve months. Panel fixed-effects regressions show that index members command a valuation premium over non-index peers that grows with passive share, consistent with a price-pressure mechanism rather than a general market-wide trend. For systematic fragility, we document a significant positive relationship between passive share and price synchronicity, a more than doubling of risk concentration among the ten largest constituents over the sample period, and persistent correlation asymmetry between up and down markets, though causal attribution of the concentration trend remains incomplete given identification constraints. Taken together, the evidence suggests that passive dominance has measurably altered both the valuation level and the risk structure of the S&P 500, with costs to market efficiency that accumulate gradually and are not visible in short-run return comparisons.

Acknowledgements

We would like to express our gratitude to our supervisor, Ignacio, for his guidance, valuable feedback, and continued support throughout this process.

We also wish to thank the portfolio managers and analysts at Fondsfinans, and Sissener for taking the time to meet with us and share their professional perspectives on passive investing and market dynamics. These conversations have meaningfully informed our thinking.

Finally, we are grateful to our fellow students for their constructive discussions and encouragement during the writing of this thesis.

Contents

List of Abbreviations

ADF Augmented Dickey-Fuller.

AIC Akaike Information Criterion.

AUM Assets Under Management.

CRSP Center for Research in Security Prices.

DW Durbin-Watson statistics.

ECM Error Correction Model.

ECT Error Correction Term.

EMH Efficient Market Hypothesis.

ETF Exchange-Traded Funds.

EW Equal-Weighted.

GFC Global Financial Crisis.

ITT intent-to-treat.

KPSS Kwiatkowski-Phillips-Schmidt-Shin.

LATE Local Average Treatment Effect.

MCR Marginal Contribution to Risk.

OLS Ordinary Least Squares.

P/B Price to Book.

P/E Price to Earnings.

RD Regression Discontinuity.

S&P 500 Standard and Poor's 500.

VW Value-Weighted.

WRDS Wharton Research Data Services.

List of Tables

List of Figures

1 Introduction and motivation

The Efficient Market Hypothesis (EMH) (Fama (1970)) has long served as the theoretical benchmark for financial markets, positing that in an efficient market, prices fully reflect all available information. This model relies on the assumption that active, informed traders compete to correct mispricing and keep prices aligned with fundamental values. However, the practical evolution of the market has diverged significantly from this theoretical ideal. While the concept of market efficiency was defined in 1970, the vehicle that would challenge it, the index fund, did not exist until 1976, when John Bogle launched the first index fund at Vanguard (Lawrence and Plagge (2023)).

Today, this structural shift has reached a tipping point. Passive funds now account for the majority of US equity fund assets, with passive ownership crossing the 50% in 2019 (Bloomberg News (2025)). Our analysis of Morningstar asset flow data, isolating the US equity category, finds passive share reaching 63.4% by December 2024. This dominance implies that the majority of capital is now allocated according to fixed index rules rather than the informed price discovery assumed by Fama. This thesis investigates the friction between the theoretical assumption of market efficiency and the empirical reality of a passive-dominated market, a tension formalised by Asness (2024) as the Less-Efficient Market Hypothesis.

This study seeks to answer the following research question:

"How does the increasing demand in index investing affect market efficiency?"

Specifically, we examine two channels through which passive investing may degrade market efficiency. First, we examine valuation distortion driven by non-informational capital flows, and second, systematic fragility arising from reduced price informativeness, mechanical risk concentration, and correlation asymmetry during market stress.

Understanding whether passive capital structurally alters market function is of broad relevance to investors, asset managers, and policymakers alike. If passive inflows systematically inflate the valuations of index constituents beyond what fundamentals justify, the widespread advice to invest passively may carry hidden costs that are not visible in short-run return comparisons (Israel et al. (2021)). If the concentration of passive capital in the largest index constituents increases risk concentration and correlation asymmetry during market stress, the diversification benefits attributed to index funds may be weaker than commonly assumed, and weakest precisely when investors need them most (Chua et al. (2009); von Moltke and Sløk (2024)). Nearly all mainstream investment advice today emphasises index funds as the sensible default choice for retail investors. Whether that advice remains sound in a market where passive dominates informed price discovery is a question with material implications for how capital is allocated and how risk is understood.

We find evidence supporting both channels. Passive share is cointegrated with key Price-to-Earnings and Price-to-Book ratios, with Error Correction Models confirming a genuine long-run equilibrium and mean reversion over horizons of four to twelve months. Panel fixed-effects regressions show that index members command a valuation premium over comparable non-index peers that grows with passive share, consistent with a price-pressure mechanism. For systematic fragility, we document rising price synchronicity, an increase in risk concentration among the largest index constituents, though the causal role of passive investing in this trend cannot be cleanly separated from the concurrent rise of mega-cap technology firms, and persistent correlation asymmetry between up and down markets.

The scope of this study is deliberately narrow. We focus exclusively on US large-cap equities using the S&P 500 as our primary laboratory, covering the period February 1993 to December 2024. International equity markets, fixed income, and small-cap equities are excluded. The S&P 500 is the most heavily tracked equity index globally and the primary vehicle through which passive capital enters the US equity market, making it the natural setting in which the effects of passive dominance should be most observable. Fund flow data is sourced from Morningstar Direct and covers open-end funds and ETFs in the US equity category.

The remainder of this thesis is structured as follows. Section 2 reviews the relevant literature on market efficiency, passive investing, and systematic fragility. Section 3 presents the hypotheses and methodology. Section 4 describes the data. Section 5 reports and discusses the empirical results. Section 6 concludes.

2 Literature review

2.1 Evolution of Market Efficiency

The theoretical anchor for this thesis is the EMH. Fama (1970) defines an efficient market as one where prices fully reflect all available information, relying on the assumption that active arbitrageurs will compete to correct mispricings. This condition creates a paradox. Grossman and Stiglitz (1980) argue that if markets are perfectly efficient, then the incentive to gather information disappears, leading investors to stop analysing, resulting in inefficiency.

Since the launch of the first index fund in 1976 (Lawrence and Plagge (2023)), the market has undergone a regime shift. With passive ownership now accounting for a dominant share of US equity assets (Bloomberg News (2025)), the market has pivoted from one defined by informed price discovery to one driven by mechanical capital allocation.

2.2 Evidence of Structural Distortion

2.2.1 Price Distortion

A growing body of literature argues that the dominance of passive capital has structurally degraded market efficiency. Asness (2024) formalises this as the Less-Efficient Market Hypothesis, arguing that the reduction in informed capital relative to passive flows has allowed mispricings to persist longer than in previous decades. Israel et al. (2021) find that the rise in passive investing has weakened the link between stock prices and fundamental metrics. They argue that index-agnostic flows overwhelm value signals, causing prices to detach from earnings and book value. The price-detachment mechanism is further supported by von Moltke and Sløk (2024), who argue that the mechanical nature of passive capital allocation causes prices to increasingly reflect index rather than fundamental value.

Several studies use the annual reconstitution of the Russell 1000 and Russell 2000 indices as a natural experiment for identifying the causal effects of passive ownership. FTSE

Russell assigns index membership solely based on end-of-May market capitalisation rank, creating a sharp discontinuity in passive ownership at the boundary. Because the assignment is mechanical and based on a continuous ranking variable, firms in the neighbourhood of the boundary are as-good-as randomly assigned. Appel et al. (2016) provide the foundational application of this design, demonstrating that the boundary generates a discontinuity in quasi-index ownership that can be used to identify causal effects on corporate governance outcomes. Their two-stage least squares framework, where the Russell 2000 dummy instruments for passive ownership, is the methodical template we follow. Their outcome variables are governance-related rather than valuation-related, leaving open the question of whether the price-pressure mechanism they documented extends to systematic valuation premiums. Boone and White (2015) use the same design to study institutional ownership and firm transparency, providing additional validation that the boundary generates a material first-stage discontinuity in institutional ownership.

The Russell reconstitution evidence bears directly on the debate between Greenwood and Sammon (2024) and Israel et al. (2021). Greenwood and Sammon (2024) document that the short-run index inclusion premium has disappeared as active traders front-run predictable flows, suggesting that discrete event-driven arbitrage has become efficient. But Russell boundary captures a structural rather than event-driven discontinuity, the same firms face different passive ownership levels year after year based on which side of the boundary they fall on. The design is therefore directly suited to identify the slow-moving structural overpricing that Israel et al. (2021) and Asness (2024) theorise.

This valuation decoupling is further compounded by a degradation in information processing. Sammon (2024) provides evidence that passive ownership reduces the market's informational efficiency, showing that high passive ownership correlates with a reduced anticipation of earnings announcements, because passive funds do not gather firm-specific information ahead of news events. Together, these findings suggest that as passive capital grows, prices increasingly reflect capital allocation mechanics rather than fundamental analysis, creating conditions under which valuation multiples may expand independently of earnings or growth.

2.2.2 Systematic Fragility

Beyond valuation distortion, a parallel strand of literature examines how the rise of passive investing has altered the risk structure of equity markets. Three distinct but related mechanisms have been identified: reduced price informativeness, mechanical risk concentration, and asymmetric correlation during market stress.

Price Informativeness and synchronicity: Roll (1988) establishes that the fraction of individual stock return variation unexplained by market-wide movements serves as a proxy for the extent of informed, firm-specific trading. When this residual is large, firm-specific information is being actively incorporated into prices. When it is small, stocks move predominantly with the market, suggesting that price discovery has been crowded out by non-informational trading. Morck et al. (2000) extends this framework cross-sectionally, showing that stock price synchronicity is systematically lower in markets with more active informed trading, and interprets high synchronicity as evidence of impaired price informativeness. Applied to the time-series dimension, this framework predicts that the rise of passive investing should increase synchronicity as the share of non-informational trading grows. Sammon (2024) provides direct evidence consistent with this mechanism, showing that high passive ownership reduces the market's anticipation of earnings announcements, a hallmark of firm-specific informed trading.

Mechanical risk concentration: Index funds allocate capital in proportion to market capitalisation, meaning that every dollar flowing into a passive fund is distributed across constituents in proportion to their size. As passive inflows grow, the largest constituents receive disproportionately large capital allocations simply by virtue of their weight in the index, independent of their fundamental value or risk characteristics. von Moltke and Sløk (2024) argue that this mechanical process has created inelastic demand concentrated in the top index constituents, where liquidity and capital increasingly pool regardless of underlying fundamentals. The implication for risk is direct. As the largest stocks absorb an ever-growing share of passive capital, their contribution to total portfolio risk grows with them, producing a concentration of marginal risk that is driven by capital allocation mechanics rather than by any change in the underlying economic risks of those firms. This dynamic is self-reinforcing. As the largest constituents grow in weight, future passive inflows further amplify their dominance, creating a feedback loop between size, weight, and risk concentration.

Correlation asymmetry and diversification breakdown: Chua et al. (2009) documents a robust empirical regularity across equity markets. Pairwise correlations among stocks are significantly higher during market downturns than during upturns. This asymmetry has direct implications for portfolio construction, as diversification benefits are weakest precisely when investors need them most. The authors link this pattern to the behaviour of investors under stress, where forced selling, margin calls, and risk-off positioning cause broad, indiscriminate liquidation that overrides the firm-specific factors that normally differentiate stock returns. The passive investing mechanism amplifies this dynamic in two ways. First, passive funds hold all index constituents simultaneously and cannot reduce exposure to the most correlated stocks during stress. Second, as passive ownership grows, the marginal seller during a market downturn is increasingly a passive fund responding to redemptions rather than an informed investor making a fundamental judgment. This mechanical, correlation-blind selling reinforces the co-movement of index constituents precisely during periods of market stress, deepening the asymmetry between up and down market correlations. von Moltke and Sløk (2024) connect this argument directly to the passive investing context, suggesting that the concentration of liquidity in index constituents creates fragility during downturns that does not manifest during upturns, consistent with the asymmetry documented by Chua et al. (2009).

2.3 Evidence of Market Adaptation

However, the evidence is not one-sided. Some scholars argue that the market has adapted to the passive shift, essentially arbitraging away simple distortions. While early studies documented massive price jumps upon S&P 500 inclusion, Greenwood and Sammon (2024) demonstrate that this *Index Effect* has largely vanished in recent years. They find that abnormal returns around index additions have declined significantly from the 1990s to the 2020s, attributing this to faster arbitrage and improved liquidity provisioning. This suggests that active traders have become more efficient at anticipating index changes, effectively "front-running" the passive flows and keeping prices closer to equilibrium.

2.4 Synthesis and Research Gap

On one side, Israel et al. (2021) and Asness (2024) argue that valuations are permanently decoupling from fundamentals as passive capital crowds out informed price discovery. On the other side, Greenwood and Sammon (2024) argue that specific, predictable event-driven arbitrage is becoming *more* efficient, with the abnormal return earned upon Standard and Poor's 500 (S&P 500) index inclusion having largely vanished as hedge funds and high-frequency traders front-run the mechanical passive buying. This thesis argues that these two findings are not in conflict. They describe the same phenomenon operating at different time horizons. The short-term announcement premium has been arbitraged away, but the underlying valuation distortion has not disappeared. It has shifted from a discrete event-driven spike into a persistent, structural overpricing of index constituents relative to their fundamentals. By testing cointegration between passive share and valuation multiples, alongside a panel comparison of index and non-index firms, we seek to show that the market has not adapted to passive dominance. It has simply embedded the inefficiency into the long-run level of prices rather than short-run price reactions.

3 Methodology and testable hypotheses

3.1 Hypotheses

The structural shift toward passive investing has introduced a fundamental tension in how capital is allocated. Index funds do not evaluate individual securities. They mechanically purchase all constituents in proportion to their market capitalisation (von Moltke and Sløk (2024)). As passive ownership grows, an increasing share of capital flows into equities regardless of their fundamental value. The pool of active investors who would otherwise correct mispricings shrinks accordingly, as Grossman and Stiglitz (1980) argue. Based on the Less-Efficient Market Hypothesis proposed by Asness (2024), we formulate two primary hypotheses:

***H1*, Price Distortion:** As passive investing grows, an increasing share of passive capital flows into equities mechanically, without regard to fundamental value. This non-informational demand inflates the prices of index constituents beyond what earnings, book value, or growth prospects justify, causing valuation multiples to expand independently of fundamentals as passive share rises.

***H2*, Systematic Fragility:** As passive investing grows, the market becomes structurally more fragile in three ways. Stocks move increasingly in lockstep because passive funds do not process firm-specific information, reducing price informativeness. Risk becomes concentrated in the largest constituents because market-cap weighting mechanically channels capital to the largest constituents. Finally, correlations rise more sharply during downturns than upturns because passive funds sell indiscriminately during redemption pressure, amplifying co-movement precisely when diversification is needed.

3.2 Methodology

3.2.1 Price Distortion

To test *H1*, we examine whether passive share is associated with elevated valuation multiples, specifically the Price to Earnings (P/E) and Price to Book (P/B) ratios of S&P

500 constituents, beyond what is explained by macro-economic and fundamental controls. We employ three complementary time-series approaches and one cross-sectional panel approach, each addressing a different dimension of the relationship.

Time-series properties and cointegration. Passive share follows a clear upward trend over the sample period, and valuation multiples exhibit similar persistent behaviour, suggesting both series may be non-stationary. Estimating a levels regression between two I(1) series without first establishing cointegration risks producing spurious results with inflated significance. We therefore test all key series for stationarity using the Augmented Dickey-Fuller (ADF) test, where the null hypothesis is a unit root, and the KPSS test, where the null hypothesis is stationarity. We classify a series as I(1) if at least one test indicates non-stationarity. Where both passive share and a valuation multiple are I(1), we apply the Engle-Granger two-step cointegration test with the null hypothesis of no cointegration:

$$\hat{\epsilon}_t = \rho\hat{\epsilon}_{t-1} + u_t \quad H_0 : \rho = 0 \tag{1}$$

where $\hat{\epsilon}_t$ are the residuals from the long-run Ordinary Least Squares (OLS) regression of the valuation multiple on passive share and controls. Rejection of the null implies the series share a common stochastic trend and a genuine long-run equilibrium relationship exists. The choice of estimator for each dependent variable follows directly from these tests.

Long-run OLS. Where cointegration is established, we estimate the long-run relationship using OLS with Newey-West heteroskedasticity and autocorrelation consistent standard errors with 12 lags:

$$M_t = \alpha + \beta_1 PS_t^{\perp} + \gamma X_t + \varepsilon_t \tag{2}$$

where M_t is the aggregate valuation multiple at month t, $PS_t^{\perp}$ is the passive share orthogonalised with respect to the 10-year Treasury yield, and X_t is a vector of controls comprising the 10-year Treasury yield, CPI inflation, real GDP growth, VIX, and the five Fama-French factors. The coefficient β_1 captures the association between passive share and valuations after removing the variation shared with long-term interest rates. The long-run OLS is presented as a baseline characterisation of the relationship. Given the autocorrelation inherent in levels regressions between trending series, the primary inferential tool is the Error Correction Model (ECM).

Error Correction Model. The ECM captures both the long-run equilibrium and the short-run adjustment dynamics. Following the Engle-Granger two-step procedure, we first recover the long-run residuals $\hat{\epsilon}$ from Equation 2, then estimate:

$$\Delta M_t = \alpha + \delta\hat{\epsilon}_{t-1} + \beta_1 \Delta PS_t^{\perp} + \gamma \Delta X_t + u_t \tag{3}$$

where Δ denotes the first difference operator and $\hat{\epsilon}_{t-1}$ is the lagged error correction term. The coefficient δ measures the speed of adjustment back to the long-run equilibrium following a deviation. A negative and significant δ confirms that the cointegration relationship is genuine and mean-reverting. Standard errors are Newey-West with 12 lags.

Panel fixed-effects regression. To complement the time-series analysis, we exploit the cross-sectional dimension of our data by comparing index members against a matched universe of non-index large-cap stocks. This approach controls for all time-invariant firm characteristics through entity fixed effects and identifies the effect of passive investing through the differential valuation premium commanded by index members as passive share grows. The specification is:

$$\ln(\text{Multiple}_{i,t}) = \alpha_i + \beta_1(1_{\text{index}} \times \text{PS}_t^{\perp}) + \beta_2 \text{PS}_t^{\perp} + \gamma X_t + u_{i,t} \tag{4}$$

where α_i denotes firm fixed effects, 1_{index} is a binary indicator equal to one if firm i is an S&P 500 constituent at time t, and X_t contains the macro controls. The coefficient of interest is β_1, which captures the incremental valuation premium of index members relative to non-index peers as passive share increases. Standard errors are clustered at the entity level. An alternative specification replaces $PS_t^{\perp}$ with the passive flow rate to provide a flow-based rather than stock-based measure of passive demand.

Granger Causality. To examine the direction of causality, we estimate a bivariate VAR in first differences and perform Granger causality tests in both directions, testing whether passive share changes predict future valuation changes and vice versa. This addresses the concern that rising valuations attract passive inflows rather than passive inflows causing rising valuations.

Causal Identification: Russell Reconstitution Regression Discontinuity. To address the endogeneity concern that passive capital may flow into indices already holding overvalued firms, we implement Regression Discontinuity (RD) design exploiting the annual Russell 1000/2000 reconstitution boundary. The running variable is each firm's end-of-May market capitalisation rank minus 1000, with negative and positive values corresponding to the Russell 1000 and Russell 2000, respectively. The treatment indicator D equals one for Russell 2000 assignment. The core estimation is:

$$\log(\text{Valuation})_{it} = \alpha + \beta D_{it} + f(\text{running\_var}_{it}) + \gamma \log(\text{mktcap})_{it} + \lambda_t + \varepsilon_{it} \tag{5}$$

where *f(running_ var)* is a quadratic polynomial estimated separately on each side of the cutoff, λ_t denotes year fixed effects, and standard errors are clustered at the firm level. Three bandwidths are used: ± 250, ± 100, and ± 50 ranks. The Russell 2000 dummy serves as both instrument and treatment proxy, yielding intent-to-treat (ITT) estimates that are conservative lower bounds on the Local Average Treatment Effect (LATE). The key coefficient of interest is β, which captures the discontinuous jump in valuations at rank zero.

3.2.2 Systematic Fragility

Beyond valuation, we examine whether the dominance of passive capital alters the risk structure of the market. We test three distinct but related manifestations of systematic fragility: price synchronicity, concentration of marginal risk contribution, and correlation asymmetry between up and down markets.

Price synchronicity. If passive investing reduces the role of informed, firm-specific price discovery (Sammon (2024)), stocks should move increasingly in lockstep with the market regardless of their individual fundamentals. We measure this using the average market model R^2, following Roll (1988) and Morck et al. (2000). For each t, we estimate a rolling 90-day market model for each index constituent:

$$r_{i,d} = \alpha_i + \beta_i r_{m,d} + \epsilon_{i,d} \tag{6}$$

where $r_{i,d}$ is the daily return of stock i and $r_{m,d}$ is the equal-weighted index return on day d. The average R^2 across all stocks in month t serves as our synchronicity measure. Higher values indicate greater co-movement and less firm-specific information being priced in. We test the integration properties of the synchronicity measure and select the appropriate specification accordingly. Where the series is I(1) and cointegration with passive share is absent, we estimate the relationship in first differences:

$$\Delta \overline{R_t^2} = \alpha + \beta_1 \Delta PS_t + u_t \tag{7}$$

with Newey-West standard errors and 12 lags. A positive and significant β_1 is consistent with passive investing crowding out firm-specific price discovery.

Concentration risk. We quantify risk concentration using the Marginal Contribution to Risk (MCR), which measures the sensitivity of total portfolio volatility to a marginal increase in each constituent's weight:

$$MCR_i = \frac{\partial \sigma_p}{\partial w_i} = \frac{(\Sigma w)_i}{\sigma_p} \tag{8}$$

where Σ is the covariance matrix of returns, w is the vector of market-cap weights, and $\sigma_p = \sqrt{w^\top \Sigma w}$ is total portfolio volatility. We compute this monthly using a rolling 60-

day return window and record the share of total risk attributable to the ten highest-MCR constituents. We then regress this top-10 MCR share on passive share:

$$MCR_{10,t} = \alpha + \beta_1 PS_t + u_t \tag{9}$$

with Newey-West standard errors. An increasing top-10 MCR share confirms that the index has become disproportionately sensitive to the idiosyncratic risks of a small number of large constituents.

Correlation asymmetry. We adopt the framework of Chua et al. (2009) to test whether diversification breaks down precisely when it is most needed. We partition all trading days into up-market days, where the equal-weighted index return is positive, and down-market days, where it is negative. We compute the average pairwise correlation among all index constituents separately for each regime. A Welch t-test is used to test whether down-market correlations are significantly higher than up-market correlations:

$$H_0 : \bar{\rho}_{\text{down}} \leq \bar{\rho}_{\text{up}} \quad \text{vs.} \quad H_1 : \bar{\rho}_{\text{down}} > \bar{\rho}_{\text{up}} \tag{10}$$

To examine whether this asymmetry has grown with the rise of passive investing, we compute the annual correlation asymmetry as the difference between average down-day and up-day correlations within each calendar year, and regress it on the annual average passive share with Newey-West standard errors.

Crisis Period, Synchronicity Analysis. To examine whether the amplification of price synchronicity during acute market stress grows in proportion with passive share, we extend the synchronicity analysis to daily frequency and compare three crisis episodes, with different passive share levels. For each trading day, we compute the fraction of S&P 500 constituents moving in the same direction as the Equal-Weighted (EW) index return, consistent with Roll (1988). We estimate:

$$\text{sync}_d = \alpha + \beta_1 \text{Panic}_d^{\text{COVID}} + \beta_2 \text{Panic}_d^{\text{GFC}} + \beta_3 \text{Panic}_d^{\text{Dot-com}} + \gamma \text{VIX}_d + \delta \left|\text{MktRet}_d\right| + \varepsilon_d \quad (11)$$

Where $Panic^{\text{k}}$ equals one within crisis window k and $|\text{MktRet}_d|$ is the absolute EW return, which controls for the mechanical relationship between market magnitude and co-movement. Standard errors are Newey-West with five lags. A third specification replaces period indicators with down-day interactions to isolate amplification, specifically on selling-pressure days. All results are replicated with a Value-Weighted (VW) benchmark as a robustness check.

4 Data

4.1 Data Strategy and Scope

To ensure the reliability and relevance of our analysis, we limit the scope of our study to US large-cap equities using the S&P 500 index as our primary proxy for the broader market. This focus is methodologically deliberate since the S&P 500 is the most heavily tracked equity index globally, and consequently the primary vehicle through which passive capital enters the equity market. Any structural distortions or efficiency degradations caused by the rise of passive investing should therefore be most observable within this index and its constituent stocks.

Our sample period runs from January 1993 to December 2024, covering 32 years of monthly end-of-month observations. The start date is determined by the earliest availability of comprehensive fund flow data from Morningstar Direct, with the first flow observation recorded in February 1993. Security-level constituent data from Wharton Research Data Services (WRDS) extends back to January 1992, providing one additional year of daily price history. While the WRDS constituent database extends as far back as 1958, we set 1992 as the start of our constituent history as a deliberate methodological choice, providing sufficient pre-sample history while maintaining data quality and consistency across all sources. The end date of December 2024 is equally deliberate: while the Morningstar net flow series records its final observation on December 1, 2024, and all remaining data sources extend into 2025, we terminate the sample at December 2024 to maintain a clean calendar-year boundary and ensure consistency across all data sources.

4.2 S&P 500 Index Constituents

A critical methodological challenge in studying the S&P 500 over long horizons is survivorship bias: the index composition changes continuously through additions and removals, and analysing only the current set of constituents would systematically exclude firms that declined and were removed. Addressing this requires a point-in-time constituent history.

The foundation of our constituent data is drawn from WRDS, which provides annual snapshots of S&P 500 membership as of January 1st each year from 1992 to 2022. For each company, this source records its PERMNO identifier, the permanent numerical identifier used by the Center for Research in Security Prices (CRSP), along with its ticker symbol, company name, and exact dates of entry into and exit from the index. The PERMNO is critical because ticker symbols are not stable over time. Companies rename themselves, merge, go bankrupt, and re-emerge under entirely new symbols, while the PERMNO remains constant throughout a company's history. The WRDS dataset was downloaded as 32 annual files and stacked into a single longitudinal file without deduplication, preserving all individual membership periods. This is important for companies that left the index and subsequently rejoined. For example, Maxim Integrated Products was a constituent from 2000 to 2007, and rejoined in 2018. Collapsing such records to a single row per PERMNO would incorrectly assign continuous membership across the gap. After removing exact duplicates arising from a company appearing in multiple annual snapshots, we retain 1,256 unique membership periods across 1,226 PERMNOs, of which 30 companies have two distinct membership periods, ensuring their gap years are correctly excluded from the analysis.

To extend the constituent history beyond December 2022, the last confirmed update date in the WRDS database, we supplement with a publicly available dataset maintained on GitHub, originally compiled by Clenow (2019) and subsequently updated by the repository maintainer to reflect index changes through December 2024 (fja05680 (2026)). This file records the full list of index constituents for each date on which a compositional change occurred. By comparing the index composition before and after the end of 2022, we identify all additions and removals occurring between January 2023 and December 2024, yielding 31 new additions not present in the WRDS data. For each of these companies, we manually identified the corresponding PERMNO using the CRSP stocknames table, which records every ticker a company has ever traded under, along with the exact dates each ticker was valid, allowing unambiguous identification across name changes and ticker reassignments.

Three companies required additional manual handling outside of the standard pipeline. Berkshire Hathaway Class B shares (BRK.B) were added to the S&P 500 in February 2010 but are not captured in the WRDS snapshots, which record only the Class A shares under a different PERMNO. We manually assign BRK.B its correct PERMNO of 83443, corresponding to the Class B listing in CRSP, with a membership window of February 16, 2010 through December 2023. Brown-Forman Class B shares (BF.B) are similarly absent from the WRDS file. CRSP records both share classes under the ticker BF, and we identify the Class B listing as PERMNO 29946 based on shares outstanding, covering membership from January 1992 through December 2024. Finally, Washington Mutual (WAMUQ) was absent from the WRDS constituent file entirely. As the largest bank failure in US history, it represented a significant index constituent during the pre-crisis period and is material to any analysis of market risk. We manually assign it PERMNO 81593, covering its S&P 500 membership from July 1997 until its government seizure on September 26, 2008.

The resulting master constituent file is expanded to a daily panel by assigning each PERMNO to every trading day within its confirmed membership window, separately for each membership period in the case of companies with multiple periods. This yields 4,165,848 stock-day observations covering 1,257 unique PERMNOs across 8,311 trading days from January 1992 to December 2024. The average number of constituents per day is 501.1, with a minimum of 498 and a maximum of 507. The excess above 500 is fully accounted for by dual share class listings. Eight company pairs hold two separate PERMNOs in the index simultaneously, including Alphabet (GOOG/GOOGL), Fox Corp (FOXA/FOX), News Corp (NWSA/NWS), Under Armour (UA/UAA), and Comcast (CMCSA/CMCSK). This is consistent with S&P's own reporting, which distinguishes between the number of companies and the number of securities in the index.

As a validation exercise, we compare our year-end constituent counts against the WRDS January 1st snapshots for each year from 1992 to 2022. The maximum discrepancy across all years is +3, which arises in years where companies were added to the index on January 2nd, the first trading day. They therefore appear in our data but not in the WRDS January 1st snapshot. This is expected and confirms the accuracy of our constituent construction.

4.3 Security-Level Data

4.3.1 S&P Constituent Data

Using the constituent history as a lookup table, we retrieve daily security-level data from the CRSP Daily Stock File (crsp.dsf) via WRDS. For each stock, we collect closing price, daily return, shares outstanding, and trading volume exclusively for the dates during which the stock was an active S&P 500 member. Observations outside a stock's confirmed membership window are excluded. Market capitalisation is computed as the absolute value of the closing price multiplied by shares outstanding, where the absolute value is applied because CRSP uses negative prices as a flag for bid-ask averages rather than actual transaction prices.

The resulting daily panel contains 4,165,672 stock-day observations across 1,257 unique PERMNOs. Of the 4,165,848 membership-panel observations, 176 have no corresponding record in the CRSP Daily Stock File and are dropped on the left merge. These arise where a PERMNO-day falls within a confirmed membership window but CRSP carries no entry for that date, most plausibly attributable to trading halts or delistings at the boundary of a membership period. Among matched observations, a negligible number return null prices or market capitalisations, attributable to similar trading interruptions. All missing observations are retained as missing rather than imputed. For the panel regression analysis in *H1*, the daily price data is aggregated to a monthly frequency by taking end-of-month values for prices and market capitalisation, and compounding daily returns within each calendar month to obtain monthly returns. Index weights are computed as each stock's end-of-month market capitalisation divided by the sum of all constituent market capitalisations in that month.

4.3.2 Valuation Ratios

P/E and P/B ratios serve as the primary dependent variables for *H1*. These are sourced from two complementary sources within WRDS, with pre-computed ratios used as the primary source and manually computed ratios filling remaining gaps.

The primary source is the WRDS Financial Ratios Suite (comp.wrds_ratios), which provides monthly pre-computed ratios linked to Compustat fundamentals, available from January 1993 to December 2024. We use pe_inc, representing the P/E ratio including extraordinary items, and ptb, representing the P/B ratio. These are reported at month-end and applied to all trading days within each calendar month.

For companies not covered by wrds_ratios, primarily foreign-incorporated US-listed companies such as Medtronic, Accenture, and Eaton Corporation, which are incorporated in Ireland but listed in the United States, we compute P/E and P/B directly from Compustat quarterly fundamentals (comp.fundq). P/E is computed as quarterly closing price (prccq) divided by earnings per share excluding extraordinary items (epspxq), and P/B as closing price divided by book value per share, where book value per share is total common equity (ceqq) divided by shares outstanding (cshoq). Observations with negative earnings, producing negative or undefined P/E ratios, are excluded. The two sources are combined using a waterfall approach. wrds_ratios are used where available, with computed values filling remaining gaps.

The PERMNO-to-gvkey link required for merging Compustat data with CRSP is established using the CCM link table (crsp.ccmxpf_linktable), retaining links of type LC, LU, and LS with primary link classification P or C. Final P/E coverage across the full panel is 94.6% and P/B coverage is 92.7%. The approximately 5-7% missing coverage is concentrated in foreign-incorporated companies and negative-earnings periods, both of which are standard limitations in the empirical asset pricing literature (Israel et al. (2021)). All main results are verified to be robust to the exclusion of these observations.

4.4 Fund Flow Data

4.4.1 Aggregate Market-Level Flows

Data on aggregate fund flows, total net assets, and organic growth rates are sourced from Morningstar Direct, using the US Open-End and Exchange-Traded Funds (ETF) Asset Flows dataset. We isolate the US Equity category group and separate observations by active and passive management classification. To ensure accurate representation, we exclude Funds of Funds, Feeder Funds, and Money Market accounts. Crucially, to mitigate survivorship bias in the fund universe, we include Obsolete funds that were liquidated or merged during the sample period.

This dataset provides three distinct monthly variables covering February 1993 to December 2024, 383 monthly observations. Estimated Net Flows capture new capital entering or leaving funds in a given month, expressed in millions of US dollars. Total Net Assets represent cumulative assets under management at month-end. Organic Growth Rate measures asset growth attributable to flows alone, excluding market appreciation, making it a purer measure of investor demand independent of market returns.

We compute the Passive Share of US Equity fund assets as the ratio of passive total net assets to total assets under management, which is our primary independent variable for Specification 1 of *H1*. This share grew from 4.4% in February 1993 to 63.4% by December 2024, crossing the 50% threshold in April 2019, consistent with widely reported figures in the financial press (Bloomberg News (2025)). Aggregate net flows into passive funds serve as the primary independent variable for Specification 2, offering a flow-based rather than stock-based measure of passive demand. Organic growth rate is retained as a robustness variable, capturing the intensity of passive demand growth independently of the level of assets already accumulated.

4.5 Panel Construction

The panel-fixed effects analysis in *H1* requires a combined dataset of index and non-index firms observed over the same time period. We construct this by stacking two monthly panels. The first consists of S&P 500 constituent observations drawn from *daily_panel*, aggregated to monthly frequency by taking end-of-month values for price, market capitalisation, and valuation ratios, and compounding daily returns within each calendar month to obtain monthly returns. The second consists of the non-index large-cap universe from *crsp_nonindex*. Both panels carry a binary indicator *in_index* equal to one for index members and zero otherwise, which serves as the basis for the interaction terms in the panel fixed-effects specifications.

The two panels are concatenated and indexed by firm identifier (PERMNO) and date, yielding an unbalanced panel of approximately 1.23 million firm-month observations spanning 1993 to 2024. The panel is unbalanced because index membership changes over time and because the non-index universe varies with market cap and listing requirements. Valuation ratios are winsorised annually at the 1st and 99th percentiles within each panel separately before concatenation to prevent extreme observations from driving results. Log transformations of P/B and P/E are used as dependent variables in all panel specifications to reduce skewness and facilitate interpretation of coefficients as proportional effects.

Three interaction terms are constructed to capture the differential passive investing effect on index members. *index_x_passive* is the product of *in_index* and *passive_share_orth*. *index_x_flow* is the product of *in_index* and the contemporaneous passive flow rate. *index_x_flow_lag* is the product of *in_index* and the one-month lagged passive flow rate. These three terms serve as the key variables of interest across the six panel fixed-effects specifications.

4.6 Data Summary

Table 1 summarises the data sources, frequencies, and coverage periods used throughout the analysis. The full sample spans February 1993 to December 2024, yielding 383 monthly observations for the time-series analysis and approximately 757,000 firm-month observations with available valuation data for the panel fixed-effects specifications. This figure reflects firm-months with available valuation data. The log specifications in Table 8 further exclude observations with non-positive earnings, yielding final estimation samples of approximately 608,000 to 622,000 firm-months depending on the dependent variable.

Table 1: Summary of Data Sources

Variable	**Source**	**Freq.**	**Coverage**
S&P 500 constituents	WRDS + GitHub	Annual	1992–2024, 1,257 stocks
Prices, mkt cap, returns	WRDS / CRSP	Daily → Monthly	1992–2024, 1,257 stocks
P/E, P/B (index)	WRDS / Compustat	Monthly	1993–2024, 94.6% / 92.7% coverage
CRSP control universe	WRDS / CRSP msf	Monthly	1993–2024
Aggregate fund flows	Morningstar Direct	Monthly	1993–2024, 383 months
Fama-French 5 factors	Kenneth French library	Monthly	1993–2024
10-year Treasury yield	FRED (GS10)	Monthly	1993–2024
CPI inflation	FRED (CPIAUCSL)	Monthly	1993–2024
Real GDP growth	FRED (A191RL1Q225SBEA)	Quart. → Monthly	1993–2024
VIX	CBOE	Daily → Monthly	1993–2024

CRSP control universe includes all US common stocks (share codes 10, 11) on NYSE, AMEX, and NASDAQ with market cap > \$500M. Fund flow data excludes Funds of Funds, Feeder Funds, Money Market accounts, and includes Obsolete funds to mitigate survivorship bias. GDP growth is reported quarterly and forward-filled to monthly frequency. VIX is averaged within each calendar month.

Passive share grew from 4.4% in February 1993 to 63.4% by December 2024, crossing the 50% threshold in April 2019. The passive flow rate averaged 0.17% of total assets per month over the sample, with considerably higher volatility in the early period reflecting rapid adoption from a small base. Valuation multiples show substantial variation over the sample. Median P/B ranged from 1.44 to 3.93 and median P/E from 8.88 to 25.27, capturing both the Dot-com bubble, the post Global Financial Crisis (GFC), and the recent technology-driven expansion. Coverage for P/E and P/B ratios across the index constituent panel is reported in Section 4.3.2.

The control variables are sourced from FRED and the Kenneth French data library and span the full sample period without gaps, with the exception of GDP growth, which is reported quarterly and forward-filled to monthly frequency, yielding 381 usable observations. Pairwise correlations among all key variables are examined in Section 5.1.1 before estimation.

4.7 Russell Reconstitution Data

The causal identification analysis uses two additional data sources not used in the main thesis analysis; annual Russell index membership reconstructed from the CRSP and firm-level valuation ratios from the WRDS Financial Ratios database. Both data sets are constructed using the internal *Python Jupyter Notebook* in the WRDS portal.

4.7.1 Russell Index Membership

Annual Russell 1000 and 2000 membership is reconstructed by replicating the official FTSE Russell ranking methodology. For each calendar year t, the last available closing price in May is identified for each security in the CRSP Daily Stock File. End-of-May market capitalisation is computed as the product of the closing price and shares outstanding. All firms are ranked in descending order within each year. Ranks 1-1000 are assigned to the Russell 1000, ranks 1001-3000 to the Russell 2000. The running variable is rank minus 1000. The data covers 1988-2024, restricted to 1992-2024 for the RD analysis. After 2007, Russell introduced a banding rule; we follow Appel et al. (2016) in not modelling it explicitly but include a pre/post 2007 robustness check.

4.7.2 Valuation Ratios and Panel

Valuation outcomes are drawn from the WRDS Financial Ratios database. The primary outcome variables are log P/E and log P/B, using the Q4 observation as the annual figure, winsorised at 1/99th percentiles. The final RD panel contains 61 491 firm-year observations (1992-2024), comprising 22 281 Russell 1000 and 39 210 Russell 2000 observations. The ± 250, ± 100, and ± 50 bandwidth samples contain 10 508, 4 214 and 2 111 observations respectively.

5 Results and analysis

5.1 Preliminary Analysis

5.1.1 Descriptive statistics and correlations

Before proceeding to hypothesis testing, we examine the pairwise correlations among the variables in our analysis. This serves two purposes. First, high correlations between independent variables may indicate multicollinearity, which inflates standard errors and complicates inference. Second, and more importantly for our identification strategy, correlations between the passive share variable and other trending series may reflect common time trends rather than genuine economic relationships, which would bias our coefficient estimates if left unaddressed. Table 2 reports the lower-triangle correlation matrix for the main variables over the full sample period.

Table 2: Correlation Matrix: Key Variables (1993 – 2024, Monthly)

Variable	PS	Flow	T10y	Inf	GDP	VIX	Mkt-RF	PB	PE
passive_share	1.00								
passive_flow_rate	0.12	1.00							
treasury_10y	**-0.76**	-0.06	1.00						
inflation	0.17	0.05	0.13	1.00					
gdp_growth	-0.05	-0.00	0.09	0.05	1.00				
vix	-0.04	0.00	-0.10	-0.05	-0.26	1.00			
Mkt-RF	0.04	0.14	-0.06	-0.15	0.10	-0.25	1.00		
pb_median	**0.59**	0.15	-0.27	0.33	0.18	-0.25	0.08	1.00	
pe_median	**0.65**	0.13	-0.34	0.26	0.19	-0.44	0.12	**0.86**	1.00

Lower triangle pairwise Pearson correlations. **Bold** entries indicate $|r| \geq 0.50$. Passive share (PS) is orthogonalised against the 10-year Treasury yield (T10y) in all main specifications; see Section 3.2. $N = 381$ monthly observations.

The most prominent feature of the table is the strong negative correlation between passive share and the 10-year treasury yield ($r = -0.76$). Passive share has risen monotonically over the sample period while long-term yields have declined from their early 1990s highs, and their high correlation likely reflects this shared trend rather than a direct economic link. To prevent this mechanical co-movement from contaminating the passive share coefficient in our regressions, we orthogonalise passive share against the 10-year treasury yield prior to estimation, retaining only the residual variation that is uncorrelated with yields. The resulting variable, $passive\_share^{\perp}$, is used throughout the *H1* analysis.

The raw correlations between passive share and the valuation multiples are 0.59 for *pb_median* and 0.65 for *pe_median*. These are consistent with the direction predicted by *H1*, but should not be interpreted causally at this stage. Both passive share and the valuation multiples exhibit a visible upward trend over the sample period, meaning their raw correlation partly reflects this shared trend. The appropriate test for a genuine long-run relationship is cointegration, which we address in Section 5.1.3 following the stationarity analysis in Section 5.1.2.

Among the control variables, correlations are generally low, indicating no serious multicollinearity concerns in the main specifications. The correlation between *pb_median* and *pe_median* of 0.86 confirms that the two multiples move closely together over the sample period, though they are not identical, which justifies reporting results for both throughout the analysis.

5.1.2 Stationarity tests

A necessary step before estimating any relationship between passive share and valuation multiples is to establish the time-series properties of the key variables. Regressing one non-stationary series on another without accounting for their integration order risks producing spurious results, where high R^2 and significant t-statistics emerge purely from shared trends rather than a genuine economic relationship. We apply two complementary tests to all key series: the ADF test, where the null hypothesis is a unit root, and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, where the null hypothesis is stationarity. A series is classified as I(1) if at least one test indicates non-stationarity. Results are reported in Table 3.

Table 3: Stationarity Tests: ADF and KPSS (1993 – 2024)

	ADF		KPSS		
Variable	p-value	Concl.	p-value	Concl.	Order
Valuation multiples					
P/B Median	0.057	I(1)	0.010	I(1)	**I(1)**
P/B Equal-weighted	0.695	I(1)	0.010	I(1)	**I(1)**
P/B Value-weighted	0.991	I(1)	0.010	I(1)	**I(1)**
P/E Median	0.192	I(1)	0.010	I(1)	**I(1)**
P/E Equal-weighted	0.006	I(0)	0.025	I(1)	**I(1)**
P/E Value-weighted	0.035	I(0)	0.100	I(0)	I(0)
Passive share					
Passive share	1.000	I(1)	0.010	I(1)	**I(1)**
Passive share (orth.)	0.945	I(1)	0.010	I(1)	**I(1)**
Passive flow rate	0.001	I(0)	0.091	I(0)	I(0)
Controls					
Treasury 10y	0.347	I(1)	0.010	I(1)	**I(1)**
Inflation	0.009	I(0)	0.100	I(0)	I(0)
GDP growth	0.000	I(0)	0.100	I(0)	I(0)
VIX	0.000	I(0)	0.100	I(0)	I(0)
Mkt-RF	0.000	I(0)	0.100	I(0)	I(0)
First differences					
Δ P/B Median	0.000	I(0)	0.100	I(0)	I(0)
Δ P/B Equal-weighted	0.000	I(0)	0.100	I(0)	I(0)
Δ P/B Value-weighted	0.000	I(0)	0.063	I(0)	I(0)
Δ Passive share	0.000	I(0)	0.010	I(1)	**I(1)**
Δ Passive share (orth.)	0.000	I(0)	0.100	I(0)	I(0)
Zivot-Andrews (structural break)					
Δ Passive share	0.000	I(0)	Break: 2008		I(0)
Δ Passive share (orth.)	0.000	I(0)	Break: 2020		I(0)

ADF null hypothesis: unit root (non-stationary). KPSS null hypothesis: stationary. A series is classified as **I(1)** if at least one test indicates non-stationarity. ADF lag length selected by AIC. KPSS bandwidth selected automatically. Δ denotes first difference. Passive share (orth.) is residualised against the 10-year Treasury yield. The Zivot-Andrews test allows for a single structural break in the intercept and tests the null of a unit root. p-values reported in the ADF column. The conflict between ADF and KPSS for Δ Passive share is resolved by the Zivot-Andrews test (t-stat $= -16.845$, $p = 0.000$, break: 2008), confirming stationarity. Δ Passive share (orth.) is similarly confirmed stationary (t-stat $= -13.375$, $p = 0.000$, break: 2020).

The valuation multiples are uniformly I(1), with one exception. P/B median, P/B EW, and P/B VW, and P/E median and P/E EW are all classified as non-stationary, confirmed by both tests in most cases. P/E VW is the exception, classified as I(0) by both tests, and is therefore treated separately in the analysis using standard OLS rather than cointegration-based methods.

Passive share and its orthogonalised counterpart are strongly I(1), consistent with their monotonic upward trend over the sample period. The passive flow rate, by contrast, is stationary. This confirms the conceptual distinction between the two measures. Passive share is a slowly accumulating stock variable while the flow rate captures month-to-month variation in new demand, which oscillates around a mean. Among the controls, the 10-year treasury yield is I(1), while inflation, GDP growth, VIX, and the market factor are all stationary.

The first differences of all I(1) valuation multiples and passive share orthogonalised are confirmed stationary, establishing that these series are integrated of order one rather than higher orders. One borderline result requires additional attention. The first difference of raw passive share is classified as I(1) by the KPSS test despite a near-zero ADF p-value, producing a conflict between the two tests. This is a known failure mode of the KPSS test when a stationary series contains a structural break. The break causes the KPSS to spuriously reject stationarity even when the series is genuinely stationary around a shifting mean.

To resolve this, we apply the Zivot-Andrews test, which explicitly allows for a single structural break in the intercept when testing for a unit root. The test rejects the null of a unit root at $p = 0.000$, identifying a structural break in 2008, consistent with the acceleration of passive share growth following the GFC, when retail and institutional investors shifted rapidly toward index products in the aftermath of active fund underperformance. The first difference of raw passive share is therefore confirmed stationary once the structural break is accounted for. The orthogonalised measure differences cleanly to I(0) under all three tests, with the Zivot-Andrews identifying a break in 2020 associated with the COVID-era surge in passive inflows.

These results shape the estimation strategy for *H1* directly. Where passive share and a valuation multiple are both I(1), we test for cointegration before proceeding, as described in Section 5.1.3. Where a series is I(0), standard OLS applies without risk of spurious inference.

5.1.3 Cointegration test

Having established that passive share and the valuation multiples are integrated of order one, we test whether they share a common long-run equilibrium using the Engle-Granger two-step cointegration test. Two I(1) series that are cointegrated do not drift apart permanently. Deviations from their equilibrium relationship are temporary and mean-reverting. The presence of cointegration justifies estimating the long-run relationship in levels and motivates the use of an ECM to capture the short-run adjustment dynamics. The null hypothesis is no cointegration, and rejection at the 5% level indicates a stable long-run relationship. Results are reported in Table 4.

Table 4: Engle-Granger Cointegration Tests (1993 – 2024)

Pair	Test statistic	p-value	Conclusion
Passive share – P/B Median	-3.549	0.0283	**Cointegrated**
Passive share – P/B Equal-weighted	-2.503	0.2778	Not cointegrated
Passive share – P/B Value-weighted	-0.690	0.9477	Not cointegrated
Passive share – P/E Median	-4.016	0.0068	**Cointegrated**
Passive share – P/E Equal-weighted	-4.138	0.0045	**Cointegrated**

Engle-Granger two-step cointegration test. Null hypothesis: no cointegration. Rejection at the 5% level indicates a stable long-run equilibrium relationship between the two series. The valuation multiple is regressed on passive share in the first step; residuals are tested for stationarity using the ADF test. P/E value-weighted excluded as it is classified I(0) in Table 3 and does not require cointegration testing.

Three of the five pairs are cointegrated with passive share at the 5% level: P/B Median (p = 0.028), P/E Median (p = 0.007), and P/E EW (p = 0.005). P/B EW and P/B VW are not cointegrated. This mixed pattern is informative. The median-based measures, which are less sensitive to the outsized influence of a small number of mega-cap constituents, show the strongest long-run link to passive share. The value-weighted P/B, by contrast, is dominated by the largest index members whose valuations are driven by firm-specific factors that passive share alone cannot fully capture in a cointegrating relationship.

These results directly shape the estimation strategy for the remainder of Section 5.2. For the cointegrated pairs, we present the long-run OLS as a baseline characterisation of the relationship and rely on the ECM as the primary inferential tool, with the lagged error correction term capturing how quickly valuations revert to their long-run equilibrium following a deviation. For the non-cointegrated pairs, P/E EW and P/B VW, we report results for completeness but interpret them with caution. P/E VW, classified as I(0) in Section 5.1.2, is estimated using standard OLS throughout.

5.2 *H1*: Price Distortion

This section presents and discusses the empirical findings. *H1* examines whether an increase in the passive share of total equity assets under management is positively associated with aggregate equity valuations. Three complementary time-series estimators are run, long-run OLS and ECM, supplemented by a fixed effects panel regression. Together, they provide a robust image of the long-run relationship.

5.2.1 Sequential OLS

Before examining the full model, Table 5 builds the specification sequentially, adding controls one at the time to the baseline passive share regression. This allows direct assessment of whether the passive share coefficient is sensitive to the inclusion of macro and factor controls.

Table 5: Sequential OLS, Dependent Variable: Median P/B Ratio

Variable	**(1) PS only**	**(2) + Treasury**	**(3) + Inflation**	**(4) + VIX**	**(5) + Mkt-RF**	**(6) Full**
Constant	2.6369	2.9283	2.8402	3.0234	3.0151	3.0122
	(0.0591)***	(0.1412)***	(0.1273)***	(0.1572)***	(0.1617)***	(0.1637)***
passive_share_orth	2.4898	2.4898	2.2393	2.0753	2.0700	2.0790
	(0.4690)***	(0.4616)***	(0.4276)***	(0.3995)***	(0.4008)***	(0.4035)***
treasury_10y		-0.0746	-0.0801	-0.0846	-0.0843	-0.0847
		(0.0341)**	(0.0351)**	(0.0377)**	(0.0376)**	(0.0375)**
inflation			0.0428	0.0458	0.0467	0.0449
			(0.0319)	(0.0307)	(0.0309)	(0.0304)
vix				-0.0089	-0.0087	-0.0084
				(0.0066)	(0.0068)	(0.0071)
Mkt-RF					0.1525	0.3134
					(0.4480)	(0.4690)
SMB						-0.1196
						(0.8728)
HML						0.5193
						(0.8561)
RMW						0.5257
						(0.9342)
CMA						-0.0272
						(1.1234)
R^2	0.368	0.445	0.463	0.484	0.484	0.487
Adj. R^2	0.366	0.442	0.458	0.478	0.477	0.475
N	382	382	382	382	382	382
DW	0.13	0.14	0.14	0.13	0.12	0.13
Controls added	No	Yes	Yes	Yes	Yes	Yes

HAC standard errors in parentheses (Newey–West, 12 lags). `passive_share_orth` is orthogonalised against `treasury_10y`. Stable $\hat{\beta}$ across specifications confirms the result is robust to control choice.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The passive share coefficient is relatively stable across all six specifications, ranging from 2.49 in the baseline to 2.07 with the inclusion of Mkt-RF. This stability confirms that the result is not driven by omitted macro or factor variables. The model-fit grows as controls are added, with R^2 rising from 0.368 to 0.487. The full specification including all Fama-French 5 factors and macro controls is used in the subsequent analyses.

5.2.2 Long-Run OLS

Table 6 reports long-run OLS estimates of the full specification across six valuation measures; median, EW, and VW variants of both P/B and P/E ratios.

Table 6: Long-Run OLS, All Coefficients (Full Specification)

Variable	**pb_median**	**pb_ew**	**pb_vw**	**pe_median**	**pe_ew**	**pe_vw**
Constant	2.9482	4.8814	4.3564	21.2373	21.0335	15.7782
	(0.1641)***	(0.4258)***	(0.9769)***	(0.7081)***	(2.1476)***	(5.0572)***
`passive_share_orth`	2.0931	9.5520	24.0269	12.6094	15.9598	22.6803
	(0.4036)***	(0.7635)***	(1.8625)***	(1.7818)***	(2.9635)***	(6.9393)***
`treasury_10y`	-0.0861	-0.3805	-0.3352	-0.6293	-0.3218	1.0667
	(0.0355)**	(0.0899)***	(0.1807)*	(0.1409)***	(0.3998)	(1.0082)
`inflation`	0.0433	0.1234	0.1248	0.1072	0.5482	0.7108
	(0.0299)	(0.0819)	(0.1791)	(0.1514)	(0.2421)**	(0.6242)
`gdp_growth`	0.0141	0.0459	0.1058	0.0624	0.0675	0.0757
	(0.0082)*	(0.0217)**	(0.0443)**	(0.0322)*	(0.0731)	(0.1398)
`vix`	-0.0064	0.0221	0.1142	-0.1134	-0.0726	0.1506
	(0.0067)	(0.0152)	(0.0352)***	(0.0293)***	(0.0749)	(0.1717)
`Mkt-RF`	0.2309	1.0549	3.5361	0.8747	0.4758	1.0368
	(0.4308)	(1.1971)	(2.4816)	(1.9904)	(4.0373)	(11.1346)
`SMB`	-0.0269	1.3549	1.6407	0.1424	0.3550	35.8543
	(0.8243)	(1.9410)	(4.4161)	(3.5613)	(7.7259)	(20.5322)*
`HML`	0.5718	2.1878	5.5504	0.9417	-4.1115	15.7538
	(0.8226)	(1.7815)	(3.5168)	(4.5412)	(9.0331)	(21.0231)
`RMW`	0.4857	0.9002	-2.2110	7.0876	-1.3702	11.2824
	(0.8922)	(2.5972)	(6.1149)	(4.7279)	(14.0125)	(35.0087)
`CMA`	-0.0287	0.4851	0.8773	1.9080	15.5337	16.1068
	(1.0953)	(2.4003)	(5.7477)	(6.3360)	(12.4765)	(35.6797)
R^2	0.508	0.740	0.733	0.649	0.293	0.138
N	381	381	381	381	381	381
DW	0.16	0.21	0.30	0.37	0.23	0.24

HAC standard errors in parentheses (Newey–West, 12 lags).

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The passive share coefficient is positive and statistically significant at the 1% level across all six specifications. The effect grows depending on how stocks are weighted, a one-unit increase in the orthogonalised passive share is associated with a 2.09-point rise for the sample median P/B ratio, rising to 9.55 for EW and 24.03 for the VW measures. The same pattern holds for P/E ratios. The larger VW coefficients suggest that passive share disproportionately affects the larger constituents.

Model fit is strong for the P/B specifications with ($R^2 = 0.508 - 0.740$) and moderate for P/E ($R^2 = 0.138 - 0.649$). *However*, the Durbin-Watson statistics (DW) are uniformly low (0.16 - 0.37), indicating substantial positive residual autocorrelation. While HAC standard errors account for this inference, the low DW motivate the ECM specifications designed for non-stationary or serially correlated data.

5.2.3 Error Correction Model

Because both passive share and most valuation multiples exhibit non-stationary behaviour in levels, an ECM is estimated to capture short-run dynamics while anchoring the relationship to a long-run equilibrium. Table 7 reports the full coefficient set including all controls and the Error Correction Term (ECT), which measures the speed at which valuation multiples revert to their long-run equilibrium following a deviation.

Table 7: Error Correction Model, All Coefficients

Variable	pb_median	pb_ew	pb_vw	pe_median	pe_ew	pe_vw
Constant	0.0033	0.0116	0.0349	0.0121	0.0355	0.0716
	(0.0054)	(0.0108)	(0.0232)	(0.0448)	(0.0882)	(0.1876)
Δpassive_share_orth	0.4415	0.3024	0.1641	3.7531	1.2367	0.1504
	(0.1882)**	(0.3380)	(0.6410)	(2.0840)*	(3.8185)	(8.7024)
Δtreasury_10y	0.0161	-0.0121	-0.1221	-0.1320	0.1692	1.4925
	(0.0286)	(0.0620)	(0.1324)	(0.2408)	(0.4719)	(0.8778)*
Δinflation	0.0283	0.0219	0.0152	0.2857	0.2015	-0.1342
	(0.0120)**	(0.0205)	(0.0509)	(0.1010)***	(0.2293)	(0.4296)
Δgdp_growth	0.0019	0.0031	0.0073	0.0096	0.0198	0.0220
	(0.0006)***	(0.0012)**	(0.0027)***	(0.0032)***	(0.0086)**	(0.0180)
Δvix	-0.0127	-0.0216	-0.0257	-0.0777	-0.0625	-0.0257
	(0.0015)***	(0.0047)***	(0.0071)***	(0.0111)***	(0.0193)***	(0.0543)
ΔMkt-RF	0.4309	0.5957	0.8119	2.7624	3.7000	3.6691
	(0.1066)***	(0.1900)***	(0.4354)*	(0.6586)***	(1.1616)***	(3.1036)
ΔSMB	0.0011	0.3638	-0.3401	1.2809	0.0281	2.2051
	(0.1720)	(0.2200)*	(0.4244)	(1.2439)	(1.7627)	(5.7984)
ΔHML	-0.1492	-0.1137	-1.1461	-0.4157	0.2558	2.1663
	(0.1208)	(0.2939)	(0.6585)*	(1.1048)	(2.8343)	(5.5209)
ΔRMW	0.5016	0.5536	-0.5107	5.6496	3.9160	-2.9345
	(0.2590)*	(0.4012)	(0.8331)	(1.7066)***	(3.0272)	(8.0333)
ΔCMA	0.2038	0.0334	-0.6722	4.1853	1.8280	-13.9135
	(0.3888)	(0.6622)	(1.2200)	(2.0650)**	(5.4457)	(11.1492)
ECT $(t-1)$	-0.0585	-0.0139	-0.0042	-0.1785	-0.1045	-0.0873
	(0.0160)***	(0.0165)	(0.0175)	(0.0269)***	(0.0316)***	(0.0378)**
R^2	0.332	0.277	0.147	0.294	0.106	0.070
N	380	380	380	380	380	380
DW	2.22	2.02	1.90	1.93	1.96	1.77

HAC standard errors in parentheses (Newey–West, 12 lags). Dependent variable: first difference of valuation multiple. ECT$(t-1)$ is the lagged residual from the long-run OLS. A negative and significant ECT confirms cointegration and mean reversion.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The ECT coefficients are negative and significant for P/B median and all three P/E specifications, confirming that deviations from the long-run equilibrium are self-correcting rather than permanent. P/E ratios revert considerably faster than P/B. The median P/E corrects at approximately 17.9% per month, implying a half-life of roughly four months, while the median P/B corrects at 5.9% per month, implying a half-life of approximately twelve months. These mean reversion horizons are not instantaneous corrections, but protracted adjustments, consistent with the Less-Efficient Market Hypothesis of Asness (2024), who argues that the reduction in informed capital relative to passive flows allows mispricing to persist longer than in previous decades. A market with abundant active arbitrage capital would correct valuation gaps quickly. The multi-month reversion horizons we document are consistent with a market in which that arbitrage capacity has diminished as passive ownership has grown. The finding also connects to Israel et al. (2021), who argue that index-agnostic flows push prices above fundamental value without triggering an immediate response, with adjustment occurring gradually as active investors identify and exploit the resulting mispricing.

The short-run coefficient on $\Delta \textit{Passive share}$ is significant for P/B median and marginally significant for P/E median, suggesting that contemporaneous changes in passive share feed most immediately into book-value multiples. This is consistent with price pressure operating through the demand side of the market rather than through a revision of earnings expectations. Book value is an accounting stock that changes slowly, so any rapid price adjustments that raises P/B must reflect price movement rather than a change in fundamentals. DW statistics are close to 2.0 throughout, confirming that the ECM removes the serial correlation that afflicted the long-run OLS and validating the residuals as a basis for inference. Macro controls perform as expected. VIX changes uniformly negative and significant, GDP growth and Mkt-RF changes are consistently positive.

5.2.4 Panel Fixed-Effects

Table 8 reports firm-level fixed effects regressions across six specifications, testing whether index membership amplifies the valuation effect as passive share grows. The key variable in each specification is the interaction of index membership with either orthogonalised passive share (1 & 2), contemporaneous passive flow rate (3 & 4), or lagged passive flow rate (5 & 6). All the specifications include firm fixed effects, macro controls and standard errors clustered at the entity level.

Table 8: Panel Fixed-Effects Regression, Full Coefficient Table

Variable	**(1) PB×PS**	**(2) PE×PS**	**(3) PB×Flow**	**(4) PE×Flow**	**(5) PB×Lag**	**(6) PE×Lag**
Dep. var	*log_pb*	*log_pe*	*log_pb*	*log_pe*	*log_pb*	*log_pe*
Key interaction terms						
`index_x_passive`	1.0249	1.2819				
	(0.1058)***	(0.2593)***				
`index_x_flow`			9.6223	10.5708		
			(0.9016)***	(3.7779)***		
`index_x_flow_lag`					8.2116	12.1297
					(0.9005)***	(3.9245)***
Main passive share / flow terms						
`passive_share`$^{\perp}$	−0.3607	−0.7932				
	(0.0608)***	(0.1874)***				
`passive_flow_rate`			0.2384	6.9161		
			(0.4963)	(2.3012)***		
`passive_flow_rate_lag1`					4.3793	4.3165
					(0.4877)***	(2.2538)*
Macro controls						
`treasury_10y`	0.0384	0.0510	0.0363	0.0431	0.0366	0.0434
	(0.0050)***	(0.0114)***	(0.0045)***	(0.0105)***	(0.0045)***	(0.0105)***
`inflation`	0.0114	0.0358	0.0108	0.0274	0.0108	0.0280
	(0.0013)***	(0.0059)***	(0.0017)***	(0.0061)***	(0.0017)***	(0.0061)***
`gdp_growth`	0.0052	−0.0089	0.0053	−0.0086	0.0056	−0.0083
	(0.0002)***	(0.0010)***	(0.0002)***	(0.0010)***	(0.0002)***	(0.0010)***
`vix`	−0.0054	−0.0195	−0.0056	−0.0190	−0.0054	−0.0187
	(0.0003)***	(0.0013)***	(0.0003)***	(0.0013)***	(0.0003)***	(0.0013)***
`Mkt-RF`	0.1640	−0.5355	0.1349	−0.6704	0.1952	−0.5511
	(0.0189)***	(0.0767)***	(0.0200)***	(0.0843)***	(0.0189)***	(0.0777)***
R^2 (within)	0.0310	0.0052	0.0233	0.0045	0.0239	0.0045
N	607,736	622,272	607,736	622,272	607,736	622,272
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Entity	Entity	Entity	Entity	Entity	Entity

Firm fixed effects in all specifications. Standard errors clustered by entity. The interaction term captures the incremental valuation premium of index members relative to non-index peers as passive share/flows increase. The negative main effect on `passive_share`$^{\perp}$ reflects relative valuation compression among non-index firms. Passive flows concentrate in index constituents, leaving comparable non-index large caps behind. The negative sign is therefore the mirror image of the positive index premium captured by the interaction term.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The interaction coefficient is positive and highly significant across all six specifications. In specification 1, a one-unit increase in orthogonalised passive share raises log P/B of index members by 1.025 relative to non-index firms, and log P/E by 1.282. This differential premium is consistent with the index-agnostic flow mechanism described by Israel et al. (2021), who argue that passive capital enters index constituents without regard to their relative valuation, pushing prices above what earnings or book value would justify. The premium we document is not a general market-wide trend. It is concentrated specifically among index members relative to otherwise comparable non-index firms, which is precisely the signature of non-informational demand systematically favouring constituents by virtue of their index membership rather than their fundamentals.

The flow-rate specifications (3 and 4) yield larger interaction magnitudes, consistent with price pressure being most acute at the point of inflow rather than accumulating gradually with the stock of passive assets. The lagged flow specifications (5 and 6) produce comparable results, confirming that the premium persists beyond the contemporaneous period and is not simply a transitory price impact that reverses within the month.

The main effect on passive share orthogonalised is negative and significant at the 1% level, indicating that, as passive share rises, non-index large caps experience relative valuation compression, consistent with the price-pressure mechanism. Passive flows bid up index constituents while comparable non-index firms, excluded from those flows, do not keep pace. The negative sign is therefore the mirror image of the positive index premium captured by the interaction term, not a contradiction of it. All macro controls carry the expected signs. Treasury yields and VIX are negatively associated with valuations, while inflation, GDP growth, and Mkt-RF are positive for P/B. The negative Mkt-RF coefficient for P/E likely reflects mechanical compression of earnings multiples during strong market returns, when price appreciation outruns earnings growth in the short run.

The within-group R^2 values are low across all specifications, ranging from 0.0045 to 0.0310. This warrants careful interpretation. The within-group R^2 in a firm fixed-effects model measures only the share of within-firm variation over time. All stable cross-sectional differences between firms are absorbed by the fixed effects and excluded from the calculation entirely. In a panel of over 600,000 firm-month observations spanning more than 30 years, within-firm variation in valuations is driven by a large number of idiosyncratic forces. Earnings surprises, analyst revisions, corporate events, sector rotations, and broader market sentiment, that are entirely unrelated to passive investing. The low R^2 values are not a weakness of the specification, but a reflection of the inherent complexity of firm-level valuation data once all stable firm characteristics are removed.

5.2.5 Granger Causality

Table 9 reports bivariate VAR Granger causality tests between passive share and each of the six valuation multiples, estimated in first differences with a maximum lag order of 4 selected by *Akaike Information Criterion (AIC)*. Two versions of passive share are tested: raw passive share (PS) and passive share orthogonalised against the 10-year Treasury yield ($\mathrm{PS}^{\perp}$). For each pair, the table reports F-statistics for both directions: passive share causing the multiple ($\mathrm{PS} \rightarrow \mathrm{M}$) and the multiple causing passive share ($\mathrm{M} \rightarrow \mathrm{PS}$).

Table 9: Granger Causality Tests: Passive Share and Valuation Multiples

	Passive share		Passive share (orth.)	
	PS $\rightarrow$ M	M $\rightarrow$ PS	PS$^{\perp}$ $\rightarrow$ M	M $\rightarrow$ PS$^{\perp}$
F-statistics (p-values in parentheses)				
P/B Median	0.001	5.732**	0.007	3.000*
	(0.970)	(0.017)	(0.934)	(0.084)
P/B Equal-weighted	0.002	2.720*	0.441	1.782
	(0.969)	(0.099)	(0.643)	(0.169)
P/B Value-weighted	0.013	0.287	1.315	2.239
	(0.908)	(0.593)	(0.269)	(0.107)
P/E Median	1.425	1.458	0.431	4.839***
	(0.234)	(0.225)	(0.731)	(0.002)
P/E Equal-weighted	0.867	4.310***	0.965	0.695
	(0.458)	(0.005)	(0.409)	(0.555)
P/E Value-weighted	0.200	2.206*	1.193	0.459
	(0.896)	(0.086)	(0.311)	(0.711)

Bivariate VAR Granger causality tests in first differences, maximum lag order 4 selected by AIC. M denotes the valuation multiple. PS$^{\perp}$ denotes passive share orthogonalised against the 10-year Treasury yield. The null hypothesis is that the row variable does not Granger-cause the column variable. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The results reveal a clear and important asymmetry. Using raw passive share, passive share itself does not Granger-cause any valuation multiple; all six F-statistics in the PS $\rightarrow$ M column are statistically insignificant. In the reverse direction, however, several multiples Granger-cause raw passive share: P/B median (F = 5.732, p = 0.017), P/B equal-weighted (F = 2.720, p = 0.099), P/E equal-weighted (F = 4.310, p = 0.005), and P/E value-weighted (F = 2.206, p = 0.086). This pattern is consistent with valuation-chasing behaviour, where investors shift into passive vehicles when valuations are elevated, and raises potential reverse-causality concerns for the long-run OLS results.

Crucially, this reverse causality largely disappears once passive share is orthogonalised against the treasury yield. In the PS$^{\perp}$ columns, the only remaining significant M $\rightarrow$ PS$^{\perp}$ results is P/E median (F = 4.839, p = 0.002) and marginally P/B median (F = 3.000, p = 0.084). All other reverse-causality results become insignificant. Moreover, PS$^{\perp}$ still does not Granger-cause any multiple in the forward direction (all p-values above

0.26), confirming that the orthogonalised passive share variable is not driving valuations through a predictive channel detectable at quarterly lags. The Granger-causality from raw valuations to raw passive share is therefore likely driven by a shared response to interest rate conditions. Once the common macro factor is removed, most of the feedback disappears.

The absence of forward Granger causality from $PS^{\perp}$ does not contradict *H1*, and two complementary pieces of literature help clarify why. First, Greenwood and Sammon (2024) documents that the short-run index addition effect, the abnormal return earned by stocks upon S&P 500 inclusion, has largely disappeared in recent years as active traders have become faster at anticipating and front-running predictable passive flows. Their finding confirms that event-driven, short-horizon arbitrage around passive trading has improved, which is precisely the mechanism that would suppress any short-run Granger predictability from passive share to valuations. Second, the price-pressure mechanism hypothesised by Asness (2024) and Israel et al. (2021) operates as a slow-moving structural effect on the level of valuations, not as a month-to-month predictive signal. Granger causality tests are designed to detect short-run predictability at lag frequencies. The ECM, which is built around the long-run cointegrating relationship, is the more appropriate test for *H1*. The two approaches are answering different questions, and the absence of Granger causality is entirely consistent with the cointegration evidence.

5.2.6 Robustness: Active Flow Rate Control

Two robustness checks are conducted. First, the sequential OLS results in Table 5 demonstrate that the passive share coefficient changes by less than 2% as controls are added one by one, from 2.49 in the baseline specification to 2.08 in the full model. This rules out omitted variable bias as a driver of the results: the coefficient does not shrink meaningfully when macro conditions and factor exposures are accounted for, confirming that the estimated effect reflects a genuine relationship rather than a proxy for broader economic trends.

Second, we test whether the passive share coefficient is sensitive to the inclusion of the active flow rate as an additional control. Adding active flow rate changes the passive share coefficient by less than 2% across all specifications, and active flow rate itself is statistically insignificant in all cases. This confirms that the estimated passive share effect is not contaminated by simultaneous movements in active fund flows, and that the orthogonalisation against the treasury yield is sufficient to isolate the passive share variation of interest.

5.2.7 *H1* Summary

The evidence from long-run OLS, ECM, and panel fixed effects supports *H1*. Across all specifications and six valuation measures, the passive share coefficient is positive, statistically significant at the 1% level, and economically meaningful. The ECM confirms the relationship is not spurious. Cointegration holds for three of the five tested measures, with deviations from long-run equilibrium self-correcting over horizons of four to twelve months. The panel results establish that the effect is concentrated among index members relative to non-index peers, consistent with the index-agnostic flow mechanism described by Israel et al. (2021) rather than a general market-wide trend.

The Granger causality tests show that $\mathrm{PS}^{\perp}$ does not predict valuations at short lags, which might appear to contradict *H1*. It does not. This finding is in fact consistent with Greenwood and Sammon (2024), who document that the short-run index inclusion premium has largely disappeared as active traders now front-run predictable passive flows. Their results confirm that short-horizon, event-driven arbitrage around passive trading has become more efficient, which is precisely why passive share carries no short-run Granger predictability. But efficiency in processing known, discrete events does not imply efficiency in resisting a slow structural accumulation of non-informational capital. Granger causality operates at monthly frequencies and tests short-run predictability. *H1* concerns the long-run level relationship between passive share and valuations. The cointegration evidence captures what Granger cannot. The premium documented by earlier literature has not necessarily disappeared. It may have transformed from a discrete event-driven spike into a slower-moving structural component of index valuations, one that mean-reverts over months to years rather than days. Greenwood and Sammon (2024) establish that the short-run event channel has closed. The cointegration evidence confirms that a long-run level relationship persists, and the panel results show that it is concentrated among index members rather than a general market trend. Front-running corrects known, discrete rebalancing events efficiently. It does not prevent that same accumulation from sustaining a structural valuation elevation over years.

5.2.8 Causal Identification: Russell Reconstitution

The *H1* analyses establish a robust economically meaningful association between passive share and equity valuations. A natural concern is endogeneity: passive funds mechanically hold index constituents, and if high-valuation firms are systematically more likely to remain in major indices due to momentum, a positive correlation may reflect selection rather than a causal price-pressure effect. The cointegration framework establishes that the relationship is genuine rather than spurious, but it does not resolve endogeneity in the conventional sense. This section addresses that limitation directly by implementing a RD design exploiting the annual reconstitution boundary between the Russell 1000 and Russell 2000 indices as a source of quasi-random variation in passive ownership.

Identification Strategy. FTSE Russell re-constituents its indices annually in June, with constituent assignment determined entirely by end-of-May market capitalisation rank. The 1 000 largest eligible US stocks are assigned to the Russell 1000, and the next 2 000 to the Russell 2000. This creates a sharp discontinuity in passive ownership at the boundary, firms ranked at the top of Russell 2000 attract substantially higher passive capital flows from small-cap ETFs while economically comparable firms ranked at the low end of Russell 1000 receive far less. Because the assignment rule is mechanical and based on a continuous size-ranking variable, firms near the boundary are as-good-as randomly assigned. Any discontinuous jump in valuations at boundary rank can therefore be attributed to the passive ownership differential generated by index assignment rather than any pre-existing difference in firm fundamentals. This approach follows the methodology of Appel et al. (2016) and Boone and White (2015).

Instrument Validity. Three pre-estimation checks support the validity of the Russell boundary as an identification device.

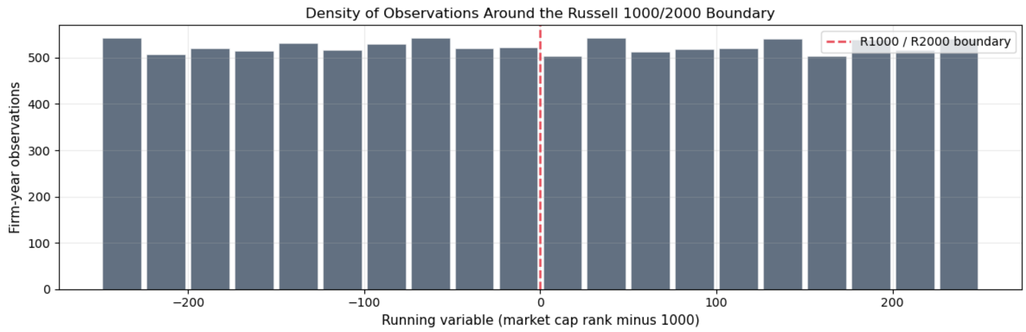

Figure 1: *Distribution of firm-year observations in 25-rank bins from* -250 *to* $+250$ *around the R1000/R2000 boundary. The red dashed line marks the cutoff at rank zero. The smooth distribution and ratio of observations just above to just below the cutoff of 0.966 confirm no strategic manipulation of the running variable.*

Figure 1 shows the distribution of firm-year observations across 25-rank bins from -250 to $+250$ around the boundary. The histogram is smooth and uniform across the cutoff, with a ratio of observations just above to just below the cutoff of 0.966, indistinguishable from 1. This rules out strategic manipulation of end-of-May market capitalisation by firms seeking to influence their index assignment, a necessary condition for the validity of the RD design.

Table 10: Covariate Balance at the Cutoff (±50 Ranks)

Characteristic	R1000 Mean	R2000 Mean	p-value	Balance
log(mktcap)	14.749	14.678	0.020**	Controlled for
Net profit margin	0.003	-1.385	0.269	✓
Return on assets	0.133	0.141	0.279	✓
Return on equity	0.110	0.129	0.424	✓

Welch t-test of mean difference between Russell 2000 and Russell 1000 firms within ±50 ranks of the boundary. H_0: no difference across boundary. log(mktcap) imbalance is mechanical, as the boundary is defined by size rank, and is controlled for in all regressions.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 10 reports covariate balance at ±50 rank bandwidth. Net profit margin ($p = 0.269$), return on assets ($p = 0.279$), and return on equity ($p = 0.424$) show no significant difference across the boundary. Log(mkptcap) shows a mechanical significant difference ($p = 0.020$). The boundary is defined by size rank, so firms at rank 999 are by construction marginally larger than firms at rank 1001. This imbalance is absorbed by including log(mktcap) as a control in all specifications.

Table 11: Reduced Form: Raw Valuation Premium at the Boundary

Outcome	BW = 250	BW = 100	BW = 50
log P/B	−0.022	+0.019	+0.024
log P/E	−0.162***	−0.045	+0.004

Raw mean difference in log P/B and log P/E between Russell 2000 and Russell 1000 boundary firms (R2000 minus R1000), without controls. Positive values indicate an R2000 premium consistent with the passive flow hypothesis. log P/B is the primary outcome; log P/E is noisy due to small-cap earnings volatility near the boundary. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 11 reports raw valuation premiums across three bandwidths. Log P/B is positive at tight bandwidths (+0.019 at BW = 100, +0.024 at BW = 50), consistent with the predicted direction of the passive flow hypothesis. Log P/E is negative at wide bandwidths, reflecting earnings volatility of small-cap firms near the boundary rather than a violation of the hypothesis.

Temporal Pattern. The most informative result in the Russell analysis is the subsample split in Table 12.

Table 12: Pre/Post 2007 Subsample: BW = 100

Variable	Pre-2007 (1992–2006)	Post-2007 (2007–2024)
Panel A: Dependent variable log P/B		
Russell 2000 (D)	0.021	0.245
	(0.095)	(0.114)**
log(mktcap)	1.381	−0.123
	(2.175)	(2.033)
R^2	0.040	0.067
N	2,093	2,121
Panel B: Dependent variable log P/E		
Russell 2000 (D)	0.403	0.076
	(0.301)	(0.350)
log(mktcap)	7.612	6.153
	(6.552)	(6.435)
R^2	0.032	0.018
N	2,093	2,121

OLS reduced form at bandwidth ±100 ranks, split at January 2007. The pre-2007 period uses purely mechanical rank-based assignment; the post-2007 period coincides with Russell's introduction of a banding rule and the acceleration of passive AUM following the Global Financial Crisis. Controls: log(mktcap) + year fixed effects. Standard errors clustered at the firm level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The Russell 2000 coefficient for log P/B is 0.021 and insignificant in the pre-2007 period, when passive Assets Under Management (AUM) remained below approximately 25% of total equity fund assets. It rises to 0.245 and is significant at the 5% level post-2007, coinciding with the acceleration of passive dominance following the GFC. The valuation premium at the boundary does not exist before passive investing reached meaningful scale and emerges precisely when it does. This pattern is difficult to reconcile with a pure size-sorting explanation, since the Russell boundary mechanism is unchanged across subperiods. It is consistent with the price-pressure channel becoming operative only once the passive capital crossed a meaningful threshold, and it mirrors the subsample ECM estimates where the speed of adjustment trends more negative post-2010, though the Chow test did not reject parameter stability across the full sample.

Reduced Form Results.

Table 13: Reduced Form: Russell 2000 Effect on Valuations

	Bandwidth		
Variable	BW = 250	BW = 100	BW = 50
Panel A: Dependent variable log P/B			
Russell 2000 (D)	0.045	0.131	0.110
	(0.048)	(0.074)*	(0.105)
log(mktcap)	1.612	0.120	-3.936
	(0.692)**	(1.439)	(2.958)
R^2	0.054	0.055	0.064
N	10,508	4,214	2,111
Panel B: Dependent variable log P/E			
Russell 2000 (D)	0.182	0.227	0.471
	(0.147)	(0.228)	(0.324)
log(mktcap)	4.618	5.400	4.635
	(1.885)**	(4.498)	(9.495)
R^2	0.021	0.023	0.036
N	10,508	4,214	2,111

OLS reduced form. Dependent variables are log P/B (Panel A) and log P/E (Panel B). D is the Russell 2000 assignment dummy. All specifications include log(mktcap) as a control, a quadratic polynomial in the running variable (rank − 1000) with separate slopes on each side of the cutoff, and year fixed effects. Standard errors clustered at the firm level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 13 reports reduced-form estimates of the Russell 2000 assignment on log P/B and log P/E across three bandwidths, controlling for log(mktcap). The headline result is a positive and marginally significant effect of Russell 2000 assignment on log P/B at the core bandwidth ±100 ranks ($\beta = 0.131, SE = 0.074, p < 0.10$). Firms assigned to the Russell 2000 trade at approximately a 14 log-point premium in P/B relative to comparable

Russell 1000 firms, controlling for size and aggregate time trends. The log P/E coefficient is positive across all bandwidths but statistically insignificant, consistent with earnings volatility of small-cap firms near the boundary.

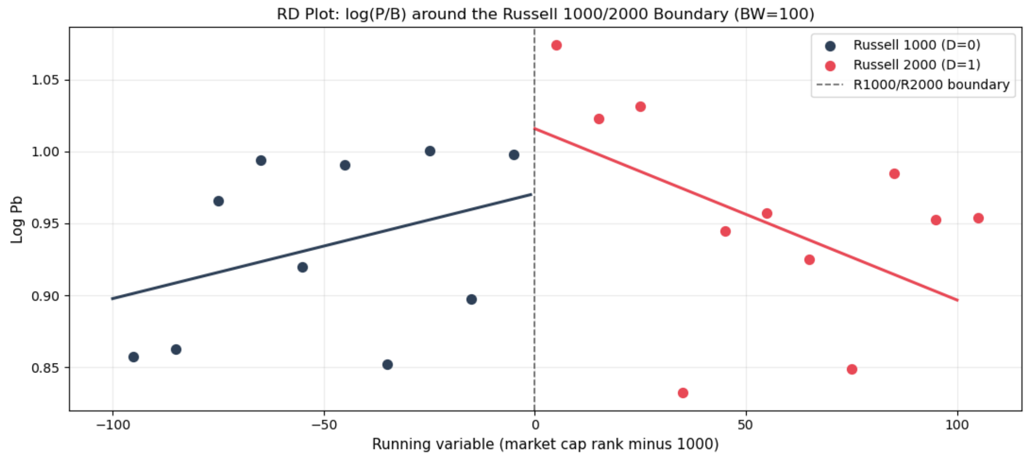

Figure 2: *Binned scatter plots of log P/B against the running variable (D) within ±100 ranks. Blue (left) = Russell 1000 (D=0), red (right) = Russell 2000 (D=1). Each dot is the mean within a 10-rank bin. Local linear fits were estimated separately on each side of the cutoff.*

Log P/B (figure 2), the red line (Russell 2000 side) begins above where the blue line (Russell 1000 side) ends at rank zero, consistent with a positive valuation premium by Russell 2000 assignment.

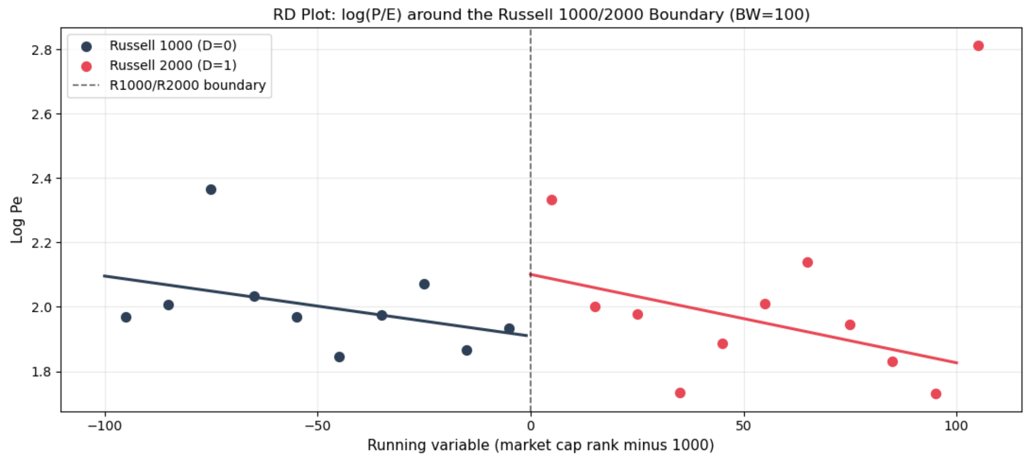

Figure 3: *Binned scatter plots of log P/E against the running variable (D) within ±100 ranks. Blue (left) = Russell 1000 (D=0), red (right) = Russell 2000 (D=1). Each dot is the mean within a 10-rank bin. Local linear fits were estimated separately on each side of the cutoff.*

The log P/E discontinuity (Figure 3) is less pronounced, consistent with the insignificant regression results in Panel B of Table 13.

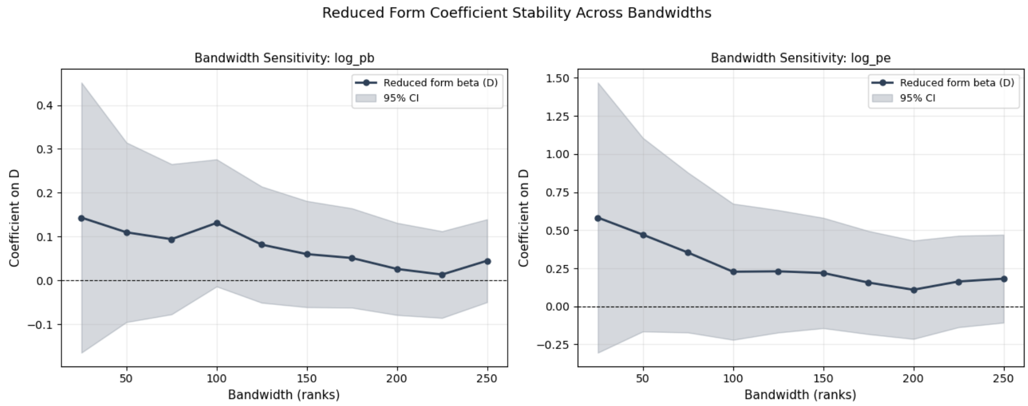

Figure 4: *Reduced form coefficient on D (and 95% CI) across bandwidths ±25 to ±250 ranks, for log P/B (left) and log P/E (right).*

Figure 4 shows the reduced form coefficient on D is consistently positive from ±25 to ±250 ranks for log P/B.

Table 14: Polynomial Sensitivity: BW = 100

	Polynomial order		
Variable	Linear	Quadratic	Cubic
Panel A: Dependent variable log P/B			
Russell 2000 (D)	0.035	0.131	0.110
	(0.051)	(0.074)*	(0.100)
log(mktcap)	0.115	0.120	0.181
	(1.438)	(1.439)	(1.442)
R^2	0.054	0.055	0.055
N	4,214	4,214	4,214
Panel B: Dependent variable log P/E			
Russell 2000 (D)	0.158	0.227	0.404
	(0.156)	(0.228)	(0.308)
log(mktcap)	5.472	5.400	5.577
	(4.487)	(4.498)	(4.503)
R^2	0.023	0.023	0.024
N	4,214	4,214	4,214

OLS reduced form at bandwidth ±100 ranks. Three polynomial orders for the running variable control (linear, quadratic, cubic), each estimated separately on each side of the cutoff. The quadratic specification is the baseline throughout this paper. Controls: log(mktcap) + year fixed effects. Standard errors clustered at the firm level in parentheses.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 14 confirms the log P/B is not an artifact of polynomial choice. The D coefficient remains stable across linear, quadratic, and cubic specifications (0.035, 0.131, 0.110).

Limitations. The Russell 2000 dummy serves as both instrument and treatment proxy, yielding ITT estimates that are conservative lower bounds on the LATE. The design identifies a local average treatment effect for mid-sized firms near the Russell 1000/2000 boundary, where the relevant passive capital flows from small-cap ETFs tracking the Russell 2000. The mechanism and magnitude may not scale to the very large S&P 500 mega-cap constituents that drive the main thesis results, where large-cap passive flows operate at a different order of magnitude. The Russell evidence should therefore be read as a directional causal support for the price-pressure mechanism rather than a precise estimate of its magnitude at the top of the market cap distribution.

Connection to *H1*. The Russell reconstitution evidence supports *H1* through a complementary identification strategy. The cointegration and ECM results establish a genuine long-run relationship for the S&P 500 over the period 1993–2024. The panel fixed-effects results show the premium is concentrated among index members relative to non-index peers. The Russell boundary design is consistent with a causal effect of passive index assignment on valuations, most clearly shown in the post-2007 period when passive ownership reached scale (Table 12), operating through the price-pressure mechanism described by Israel et al. (2021). Taken together, the evidence from the long-run OLS, ECM, panel fixed-effects, and the Russell RDD supports *H1* across four complementary estimators, drawing on two distinct identification approaches, the time-series cointegration and the Russell reconstitution regression discontinuity.

5.3 *H2*: Systematic Fragility

We test three distinct but related manifestations of systematic fragility: price synchronicity, concentration of marginal risk contribution, and correlation asymmetry between up and down markets. These are three lenses on the same underlying question: has passive dominance made the index more fragile, with stocks moving in lockstep, risk pooling in the largest constituents, and diversification failing when it matters most? Together they examine whether the rise of passive investing has altered the co-movement structure and risk concentration of the S&P 500.

5.3.1 Price Synchronicity

Table 15 reports the results of regressing the average market model R^2 on passive share in both levels and first differences. The levels specification is presented for comparison but is not the primary inferential model. Both series are I(1) and the Engle-Granger test fails to reject the null of no cointegration ($p = 0.253$), implying that levels regression risks producing spurious results. The first differences specification is therefore the primary model.

Table 15: Price Synchronicity and Passive Share (1993 – 2024)

	Levels $\overline{R_t^2} \sim \text{PS}_t$	First differences $\Delta\overline{R_t^2} \sim \Delta\text{PS}_t$
Passive share (β)	0.2605***	4.4935***
	(0.0874)	(1.1296)
Constant	0.2273	-0.0067
	(0.0291)	(0.0025)
R^2	0.132	0.031
DW	0.20	1.48
N	383	382
Integration order	I(1)	I(0)
Cointegrated	No	—
Primary specification	No	**Yes**

Dependent variable is the monthly average market model R^2 across all S&P 500 constituents, estimated from rolling 90-day regressions of individual stock returns on the equal-weighted index return. The levels specification is reported for comparison only. Both $\overline{R_t^2}$ and passive share are I(1) and not cointegrated, so the first-difference specification is the primary inferential model. HAC standard errors in parentheses (Newey-West, 12 lags).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

In first differences, a one percentage-point increase in passive share is associated with a 4.49-unit increase in the average market model R^2 in the same month ($p < 0.001$). While the R^2 of 0.031 is modest, this is expected. Monthly changes in synchronicity are driven primarily by market-wide shocks, such as volatility spikes and crisis episodes, which a passive share alone cannot explain. The direction and significance of the coefficient are what matter for inference.

The result connects directly to the price informativeness framework established by Roll (1988) and extended by Morck et al. (2000). Roll (1988) proposes the market model R^2 as an inverse measure of firm-specific price discovery. A low R^2 indicates that stock prices are moving in response to individual fundamentals, while a high R^2 indicates that stocks are moving predominantly with the market, leaving less room for firm-specific information to be reflected in prices. Morck et al. (2000) extend this framework cross-sectionally, showing that markets with more active informed trading exhibit lower synchronicity, and interpret rising R^2 as evidence of impaired price informativeness. Applied to our time-series setting, the rise of passive investing should increase synchronicity as the share of non-informational trading grows, which is precisely what we find. As passive capital crowds out informed trading, the firm-specific signal embedded in individual stock prices weakens and co-movement with the index increases.

This finding is further consistent with Sammon (2024), who provides direct microeconomic evidence of the same mechanism. Sammon (2024) shows that stocks with higher passive ownership exhibit a reduced anticipation of earnings announcements in the days preceding the release, a hallmark of firm-specific informed trading that disappears as passive ownership grows. Our aggregate synchronicity result is the market-level manifestation of the same process. Less informed trading at the stock level translates into greater co-movement at the index level.

5.3.2 Concentration Risk

Figure 5 shows the evolution of the top-10 MCR share from 1993 to 2024, plotted alongside passive share. The top-10 MCR share measures the fraction of total portfolio risk attributable to the ten highest-MCR constituents in each month, computed using a rolling 60-day return window.

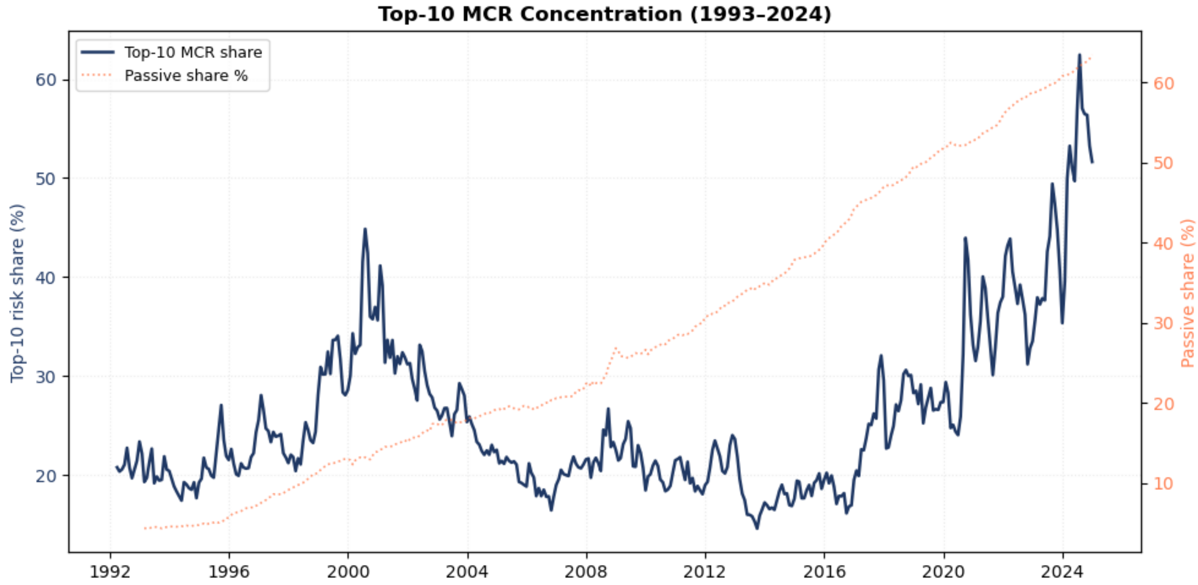

Figure 5: *Top-10 MCR Concentration and Passive Share (1993 – 2024). The solid line shows the monthly share of total S&P 500 portfolio risk attributable to the ten highest MCR constituents, computed using rolling 60-day return windows. The dotted line shows passive share of total US equity fund assets (right axis), sourced from Morningstar Direct.*

Table 16: MCR Concentration and Passive Share (1993 – 2024)

	Levels $MCR_{10,t} \sim PS_t$	First differences $\Delta MCR_{10,t} \sim \Delta PS_t$
Passive share (β)	0.2241***	-0.3119
	(0.0817)	(0.5368)
Constant	0.1942	0.0013
	(0.0219)	(0.0012)
R^2	0.219	0.001
DW	0.08	1.72
N	383	382
Integration order	I(1)	I(0)
Cointegrated	No	—
Primary specification	No	**Yes**
Descriptive statistics		
Sample mean (top-10 MCR share)	25.8%	—
1993 average	20.5%	—
2024 average	53.1%	—

Dependent variable is the monthly top-10 MCR share, computed using rolling 60-day return windows. The levels specification is reported for comparison only. Both series are I(1) and not cointegrated (Engle-Granger $p = 0.831$), so the first-difference specification is the primary model. Monthly ΔMCR is heavily right-skewed (skewness = 0.83, kurtosis = 8.26), reflecting crisis-driven spikes rather than gradual drift. HAC standard errors in parentheses (Newey-West, 12 lags).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The descriptive trend is striking. The top-10 MCR share averaged 20.5% in 1993 and reached 53.1% by 2024, meaning that by the end of the sample, more than half of the total risk in the S&P 500 was concentrated in just ten stocks. Two distinct phases are visible in Figure 5. From 1993 through roughly 2016, concentration fluctuated in a relatively stable range of 15% to 45%, with spikes during the Dot-com bubble and a trough during the GFC, when correlations rose uniformly across the index, temporarily distributing risk more evenly. From 2017 onward, concentration increases sharply and persistently, coinciding with passive share crossing 40% and accelerating toward 60%. This trajectory is consistent with the mechanical feedback loop described by von Moltke and Sløk (2024), who argue that market-cap weighting directs disproportionately large passive inflows toward the largest constituents, regardless of their fundamental value or risk characteristics. As the largest stocks absorb a growing share of passive capital, their contribution to total portfolio risk grows with them, producing a concentration of marginal risk that is driven by capital allocation mechanisms rather than by any change in the underlying economics of the firms.

The levels regression in Table 16 yields a significant positive coefficient ($\beta = 0.224, p = 0.006, R^2 = 0.219$), consistent with *H2*. However, both the top-10 MCR share and passive share are non-stationary, and the Engle-Granger test fails to reject the null of no cointegration ($p = 0.831$), meaning the levels result is potentially spurious and driven by a common upward trend rather than a genuine economic relationship. The first difference specification is therefore the primary inferential model. In first differences, the coefficient on passive share is negative and insignificant ($\beta = -0.312, p = 0.561, R^2 = 0.001$), indicating that monthly changes in passive share do not predict monthly changes in MCR. This null result is partly explained by the nature of MCR concentration itself. Monthly changes are heavily right-skewed and kurtotic, reflecting sharp crisis-driven spikes, rather than gradual drift, and passive share moves too slowly to explain this month-to-month variation.

This null result motivates a more careful identification strategy. Two confounding factors are particularly important. First, the simultaneous rise of mega-cap technology firms could independently produce rising risk concentration even absent passive investing. Second, the passive share mechanism operates structurally and slowly, which may be more visible across sub-periods than in monthly first differences.

Table 17: MCR Concentration, Top-10 Mega-Cap Weight Control

Variable	**(1) Baseline**	**(2) + Top10**	**(3) Raw PS**
Δ`passive_share_orth`	0.0504	0.0530	
	(0.0339)	(0.0329)	
Δ`passive_share`			−0.6292
			(0.5244)
Δ`top10_mktcap_share`		0.6322	0.6659
		(0.2397)***	(0.2499)***
Constant	0.0008	0.0005	0.0015
	(0.0007)	(0.0007)	(0.0011)
R^2	0.004	0.035	0.034
N	381	381	381

Dependent variable: Δtop-10 MCR share (first difference). The control variable `top10_mktcap_share` is the fraction of total index market capitalisation held by the ten largest constituents, mirroring the same threshold used to construct the MCR measure. Model 2 passive share is insignificant ($p = 0.107$), indicating that the mega-cap concentration channel fully absorbs the passive share effect. HAC standard errors (Newey–West, 12 lags).
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 17 introduces the first difference of the top-10 market-cap share as an explicit control for mega-cap concentration. In specification 1, replicating the baseline, passive share is positive, but insignificant ($\beta = 0.050, p = 0.137$). In specification 2, adding Δshare of top-10 market capitalisation, the top-10 weight is strongly significant ($\beta = 0.632, p < 0.001$) and improves the model's fit, with R^2 rising from 0.004 to 0.035. Passive share remains insignificant in this specification. Specification 3 substitutes raw passive share, which is negative and insignificant ($\beta = -0.629, p = 0.230$), while the top-10 weight remains the dominant driver. The causal pathway therefore runs through mega-cap weight concentration rather than passive share directly, and passive share has no residual explanatory power once this channel is absorbed. This is consistent with von Moltke and Sløk (2024), who argue that the concentration of liquidity in index constituents is the proximate source of fragility, with passive investing the structural force that continuously amplifies it through market-cap-weighted inflows.

Table 18: MCR Concentration, Structural Break Test, Pre vs Post January 2017 (top-10)

Variable	**Pre-2017 (1993–2016, $N = 285$)**	**Post-2017 (2017–2024, $N = 96$)**
Δ`passive_share_orth`	0.0364	0.2234
	(0.0287)	(0.2469)
Δ`top10_mktcap_share`	0.8282	0.4614
	(0.2996)***	(0.3596)
R^2	0.063	0.020
Chow test: $F = 1.151$, $p = 0.328$ (H_0: no structural break at January 2017)		

Dependent variable: Δtop-10 MCR share. The passive share coefficient is six times larger post-2017 (0.223 vs 0.036) but insignificant in both sub-periods. The top-10 mktcap share loses significance post-2017, likely due to the limited post-2017 sample. The Chow F-statistic fails to reject the null of parameter stability. HAC standard errors (Newey–West, 12 lags).
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 18 splits the sample at January 2017, when passive share crossed 40% and the second phase of concentration growth began. The passive share coefficient is 0.036 pre-2017 and 0.223 post-2017, six times larger in the period of faster passive growth, but insignificant in both sub-periods. The top-10 market-cap share is significant pre-2017 ($\beta = 0.828, p < 0.001$) but loses significance post-2017 ($\beta = 0.461, p = 0.199$), which likely reflects the limited post-2017 sample of only 96 months rather than a genuine structural change in the relationship. The Chow F-statistic of 1.151 ($p = 0.328$) fails to reject the null of parameter stability. Together, these identification tests do not provide statistically sufficient evidence to attribute the MCR trend causally to passive investing. The mechanism operates through mega-cap weight concentration, which passive investing amplifies but cannot be cleanly separated from in the available sample. Disentangling the two channels requires either a natural experiment in passive ownership or a longer post-2017 sample.

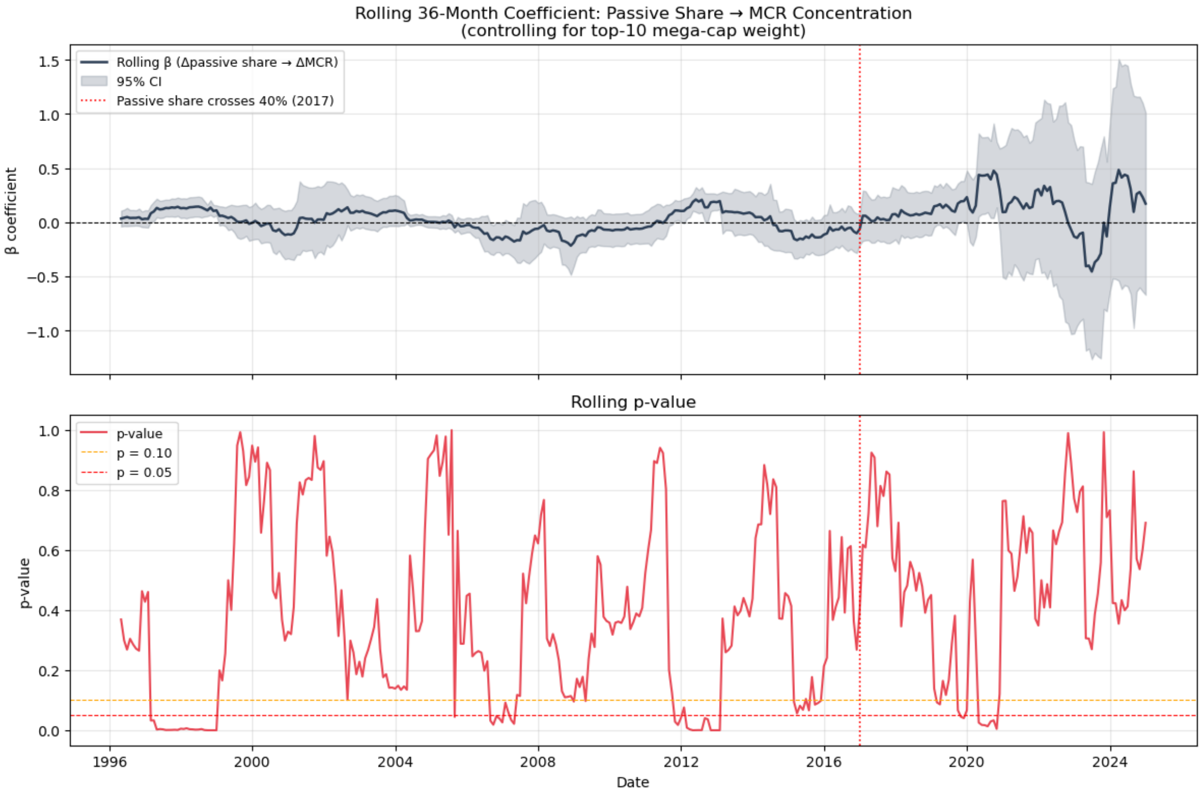

Figure 6: *Rolling 36-month OLS regression of* Δ*top-10 MCR share on* Δ*passive_share*$^{\perp}$*, controlling for* Δ*top-10 mktcap share. Shaded band = 95% confidence interval. HAC standard errors (Newey-West, 6 lags). Red vertical line marks January 2017 when passive share crossed 40%*

Figure 6 complements the statistical results in Table 17 and Table 18 by showing how the passive share coefficient has evolved over time. The rolling β fluctuates around zero throughout the pre-2017 period, with no sustained positive signals. Post-2017, the coefficient shifts upward, and the confidence interval widens considerably due to the shorter rolling windows available at the end of the sample. Although episodically positive in recent years, the coefficient does not consistently remain above zero and the p-value rarely crosses the threshold of 10%. This visual evidence is consistent with the interpretation that any passive share effect on MCR is weak, unstable, and indistinguishable from noise after controlling for mega-cap weight concentration.

We therefore interpret the MCR evidence carefully. The long-run trend from 20.5% to 53.1% is real and economically large, and its co-movement with passive share over the same period is consistent with the mechanism proposed in *H2*. However, passive share cannot be identified as the independent statistical driver of this trend within the available identification framework. Whether the concentration trend reflects passive investing, the broader rise of mega-cap technology firms, or both, remains an open question that future research with firmer identification could address.

5.3.3 Correlation Asymmetry

Table 19 reports two sets of results. Panel A tests whether average pairwise correlations among S&P 500 constituents are higher on down-market days than on up-market days over the full sample period. Panel B examines whether this asymmetry has grown alongside passive share over time.

Table 19: Correlation Asymmetry: Up vs. Down Markets (1993 – 2024)

Panel A: Full-sample asymmetry	
Up-market average correlation	0.1935
Down-market average correlation	0.2266
Difference (Down − Up)	0.0331
Welch t-statistic	18.509
p-value (one-sided)	0.0000***
Up-market days	4,524
Down-market days	3,787
Panel B: Asymmetry trend	
	Annual asymmetry $\sim$ Passive share
Passive share (β)	0.0365
	(0.0481)
Constant	0.0201
	(0.0151)
R^2	0.024
N (years)	32

Panel A reports average pairwise correlations among S&P 500 constituents on up-market and down-market days, defined by the sign of the equal-weighted index return. Correlations computed from the full upper triangle of the pairwise correlation matrix. The Welch t-test is based on 10,000 randomly sampled pairs from each regime and tests the one-sided hypothesis $H_1 : \bar{\rho}_{\text{down}} > \bar{\rho}_{\text{up}}$. Panel B regresses annual correlation asymmetry on annual average passive share. HAC standard errors in parentheses (Newey-West, 3 lags).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Panel A provides strong evidence of correlation asymmetry. The average pairwise correlation among constituents is 0.1935 on up-market days and 0.2266 on down-market days, a difference of 0.033 that is highly significant ($t = 18.509, p = 0.000$). This confirms the core finding of Chua et al. (2009), who show that pairwise stock correlations are systematically higher during market downturns than upturns across a wide range of equity markets and time periods. They attribute this asymmetry to investor behaviour under stress, where forced selling, margin calls, and risk-off positioning produce broad, indiscriminate liquidation that overrides the firm-specific factors that normally differentiate stock returns, which they term the myth of diversification, since the protection investors believe they hold evaporates precisely when they need it most. An investor holding a passive index fund is fully exposed to this effect, as the fund holds all constituents mechanically and cannot reduce exposure to the most correlated stocks during stress. As von Moltke and Sløk (2024) argue, the concentration of passive capital in index constituents means that redemption-driven selling during downturns is mechanical and indiscriminate, reinforcing co-movement precisely when diversification is most needed.

Panel B is a non-result. The regression of annual correlation asymmetry on annual passive share yields a positive and statistically insignificant coefficient ($\beta = 0.037, p = 0.433, R^2 = 0.024$). The null cannot be rejected, and the low R^2 provides no basis for treating the positive point estimate as economically meaningful. The time-series variation in correlation asymmetry is not explained by passive share in this specification. The Panel A finding stands on its own. Correlation asymmetry is a robust structural feature of equity markets under stress, documented here and consistent with Chua et al. (2009), but its variation over time does not track the rise of passive investing in annual data. The crisis period evidence in Section 5.3.4 offers an informal complement to this null, comparing asymmetry levels across episodes at different passive share concentrations, but those comparisons are descriptive and do not constitute a test.

5.3.4 Crisis Period Analysis

Table 20 presents passive share, top-10 MCR share, and correlation asymmetry for five major market stress episodes between 2000 and 2022. The crisis periods are chosen to span a wide range of passive share levels, from 14.7% during the Dot-com bust to 57.5% during the 2022 rate-hike shock, allowing for an informal assessment of whether fragility indicators have changed as passive investing has grown.

Table 20: Market Stress: Passive Share, Concentration, and Correlation Asymmetry

Crisis episode	Passive share	Top-10 MCR share	Corr (Up)	Corr (Down)	Asymmetry (Down − Up)
Dot-com bust (2000–02)	14.7%	33.6%	0.1293	0.1113	*-0.0180*
GFC (2008–09)	24.0%	22.4%	0.3273	0.3466	+0.0193
European debt (2011)	29.5%	18.8%	0.4149	0.5252	+0.1103
COVID crash (2020)	52.2%	26.4%	0.4242	0.5268	+0.1026
Rate hike shock (2022)	57.5%	38.5%	0.2397	0.2272	*-0.0125*

Each row reports average values during the indicated crisis window. Passive share and top-10 MCR share are monthly averages over the crisis period. Up- and down-market correlations are computed from daily returns within the crisis window, defined by the sign of the equal-weighted index return. Asymmetry is the difference between down- and up-market average pairwise correlations. *Italicised* asymmetry values indicate episodes where down-market correlations were lower than up-market correlations, contrary to the general pattern.

Three of the five episodes display positive correlation asymmetry, consistent with the full-sample finding in Panel A, Table 19. The European debt crisis (2011) and the COVID crash (2020) show the largest asymmetry differentials, at +0.110 and +0.103, respectively, both occurring at passive share levels above 29%. The GFC (2008-09) also shows positive asymmetry, though more modest at +0.019. Both the European debt crisis and the COVID crash are precisely the type of broad, liquidity-driven panic that Chua et al. (2009) identify as the conditions under which correlation asymmetry is most pronounced, with indiscriminate selling across all index constituents, amplified by the mechanical, redemption-driven selling of passive funds that von Moltke and Sløk (2024) describe as a structural source of downside fragility.

Two episodes deviate from this pattern and warrant explanation. The Dot-com bust (2000-02) shows negative asymmetry (-0.018), meaning down-market correlations were lower than up-market correlations during this period. The crash was concentrated in technology stocks and did not produce the broad, indiscriminate selling associated with systemic crises. As Chua et al. (2009) note, sector-specific crashes tend to reduce cross-stock correlations rather than increase them, as capital rotates away from the affected sector toward others. Passive share was also at its lowest in the sample at 14.7%, consistent with the passive investing mechanism not yet being operative at a meaningful scale. The rate hike shock of 2022 similarly shows negative asymmetry (-0.013) despite passive share reaching 57.5%. This episode is better characterised as a systematic macroeconomic repricing than a panic-driven sell-off. Rising interest rates compress valuations uniformly across sectors and produce similar return distributions on both up and down days, leaving little room for asymmetry to emerge. Notably, the top-10 MCR share reached 38.5% during this episode, the highest of all five crises, yet the correlation asymmetry was negative. This decoupling suggests that risk concentration and correlation asymmetry reflect distinct dimensions of systematic fragility that do not always co-move.

Taken together, the crisis evidence suggests that correlation asymmetry is most pronounced during broad market panics driven by liquidity and sentiment, and least pronounced during concentrated sector crashes or macroeconomic repricing episodes. This pattern is consistent with the mechanism Chua et al. (2009) identified. It is the nature of the shock, not the level of passive share, that determines whether the asymmetry manifests.

5.3.5 Crisis Period, Daily Synchronicity

The daily fraction co-movement averages 0.689 across the full sample with a standard deviation of 0.135, computed relative to the EW index for consistency with section 5.3.1. The measure behaves as expected; the ten most synchronous trading days include the August 8th, 2011 S&P credit downgrade (sync = 1.000), the March 12th, 2020 COVID crash (sync = 0.998), and the September 29th, 2008 TARP vote failure (sync = 0.998).

Regression results.

Table 21: Daily Price Synchronicity, Crisis Stress Test (Equal-Weighted)

	(1) COVID	(2) All crises	(3) Down×crisis
Panic (COVID)	-0.1873^{*}	-0.2348^{**}	
	(0.0976)	(0.1071)	
Panic (GFC)		-0.1367^{***}	
		(0.0248)	
Panic (Dot-com)		-0.0508^{***}	
		(0.0044)	
Down×COVID			-0.2150^{*}
			(0.1196)
Down×GFC			-0.0965^{***}
			(0.0239)
Down×Dot-com			-0.0423^{***}
			(0.0047)
Down day			0.0191^{***}
			(0.0019)
VIX	-0.0042^{***}	-0.0028^{***}	-0.0038^{***}
	(0.0002)	(0.0003)	(0.0003)
|Market return|	13.1832^{***}	13.2371^{***}	13.2103^{***}
	(0.6160)	(0.6090)	(0.6254)
Constant	0.6661^{***}	0.6449^{***}	0.6517^{***}
	(0.0045)	(0.0052)	(0.0051)
R^2	0.590	0.607	0.598
N	8,308	8,308	8,308

Dependent variable: daily fraction co-movement relative to the equal-weighted index. Panic dummies equal one within the indicated crisis window. Down×crisis interacts each dummy with a negative-return-day indicator. HAC standard errors in parentheses (Newey-West, 5 lags). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 21 reports three specifications. The large positive coefficient on absolute market return (approximately 13.2 across specifications, significant at 1%) confirms that market magnitude mechanically drives co-movement; on a 10% down day, most stocks fall together by force of shock scale. After conditioning on this, the crisis dummies are negative and significant. This reflects within-crisis sector dispersion that the market return control cannot fully absorb. A suggestive pattern emerges in specification 2: the absolute magnitude of the crisis effect is ordered COVID > GFC > Dot-com, the same as passive share across those episodes (51.4%, 26% and 14.7% respectively). While the individual crisis coefficients are significant at the 5% and 1% levels, the ordering itself rests on only three episodes and should be read as descriptive support rather than a formal test of the passive-share gradient. Specification 3 confirms that the pattern concentrates on down-market days, with the same gradient preserved for down-day interactions.

Comparison across crisis episodes. Table 22 provides the most direct evidence under the EW benchmark. The down-minus-up synchronicity gap, which measures the extent to which co-movement concentrates specifically on selling days, is +0.086 during the COVID crash versus +0.009 in non-crisis periods and +0.008 during the GFC. This pattern does not fully replicate under a Value-Weighted benchmark, a divergence driven by the unusual behaviour of mega-cap technology stock during the Covid crash. The EW result should be read with that limitation in mind. The GFC comparison is the most informative: despite longer duration (121 days) and comparable market volatility, the GFC shows essentially no down-day asymmetry. Passive share during GFC was 26% versus 51.4% during COVID. The absence of any meaningful asymmetry during the GFC is difficult to reconcile with a pure volatility explanation and is consistent with the mechanical selling channel becoming material only at higher levels of passive ownership. The COVID window spans only 23 trading days, and the gap estimate should be interpreted with appropriate caution given the limited sample. However, the finding is consistent across both the regression specifications in Table 21 and the descriptive comparison in Table 22, lending it robustness beyond the raw gap alone.

Table 22: Crisis Synchronicity, Up vs Down Days (EW and VW)

		Equal-Weighted (EW)			Value-Weighted (VW)			
Period	PS	Up	Down	Gap	Up	Down	Gap	N
Normal (ex-crises)	29.7%	0.685	0.694	+0.009	0.674	0.690	+0.016	7,516
Dot-com (2000–02)	14.7%	0.652	0.671	+0.020	0.650	0.655	+0.005	648
GFC (2008–09)	26.0%	0.791	0.798	+0.008	0.761	0.818	+**0.057**	121
COVID (2020)	51.4%	0.811	0.896	+**0.086**	0.860	0.872	+0.012	23

Gap = synchronicity on down-market days minus up-market days. EW columns use the equal-weighted benchmark (primary, consistent with Roll (1988) and Section 5.3.1). VW columns use the value-weighted benchmark (robustness). Up/down days defined per respective benchmark. Passive share (PS) is the monthly average within each window, forward-filled to daily frequency. Bold entries indicate the largest gap per benchmark.

Robustness: VW benchmark. The VW columns of Table 22 replicate the analysis using the VW S&P 500 return. The specification 2 regression gradient is preserved, but the comparison table finding does not hold; the GFC shows the largest gap (+0.057) while COVID collapses to 0.012. This divergence reflects the distinctive composition of the COVID crash, in which mega-cap technology stocks (dominating the VW index) held up or appreciated while the majority of constituents in cyclical sectors fell sharply. This produces symmetric high synchronicity on both up and down COVID days, without the one-sided asymmetry that *H2* specifically predicts. Additionally, Specification 3 under VW shows a broken gradient: *Down*x*Dot-com* exceeds *Down*x*GFC* in absolute value, inconsistent with the passive share ordering. We treat the EW specification as primary for three reasons: It is consistent with Roll (1988) and Morck et al. (2000), it uses the same benchmark as section 5.3.1, and it more accurately represents the co-movement of the average constituents rather than the largest stocks. The sensitivity to weighting choice is itself informative; it isolates the COVID mega-cap divergence as a distinct structural feature that partially offsets the synchronicity signal under value-weighting.

5.3.6 *H2* Summary

The evidence for *H2* is mixed but instructive. Price synchronicity increases significantly with passive share in first differences ($\beta = 4.49, p < 0.001$), consistent with passive investing crowding out firm-specific price discovery as predicted by *H2* (Roll (1988); Morck et al. (2000); Sammon (2024)). Correlation asymmetry between up- and down-market days is a robust and statistically conclusive feature of the full sample ($\Delta = 0.033, t = 18.509$), confirming that diversification breaks down when most needed (Chua et al. (2009)), though its time-series variation cannot be attributed to passive share growth.

The MCR concentration finding is the most nuanced. The descriptive trend, from 20.5% to 53.1% over the sample, is economically striking and temporally consistent with the growth of passive investing. The levels regression is significant but misleading. The first-difference tests are insignificant, but the improved identification reveals that changes in mega-cap weight are the proximate driver of MCR variation, and passive share remains insignificant once this channel is controlled for, though its coefficient stays positive and close to the 10% threshold (p = 0.107). The structural break test shows a six-fold increase in the passive share coefficient post-2017, though without statistical significance in the available sample. The most defensible interpretation is that the MCR trend is real and the proximate statistical driver is mega-cap weight concentration. Once this channel is controlled for, passive share has no residual explanatory power in first differences. Whether passive investing is the structural force amplifying mega-cap concentration through market-cap-weighted inflows, as von Moltke and Sløk (2024) argues, cannot be established with the available identification. Full causal attribution requires a longer post-2017 sample or an exogenous instrument for passive ownership.

Each identifies a distinct channel through which passive dominance may degrade the risk properties of the index, and each finds evidence directionally consistent with *H2*, even where causal attribution is incomplete. Price synchronicity establishes that the rise of passive capital has measurably eroded firm-specific price discovery at the aggregate level, crowding out the informed trading that Roll (1988) and Morck et al. (2000) identify as the foundation of efficient markets. Correlation asymmetry confirms that the diversification S&P 500 investors believe they hold weakens precisely when it is most needed, a structural feature of equity markets under stress that passive investing reinforces through mechanical, indiscriminate selling during redemption pressure. MCR concentration documents a more than doubling of risk concentration in the top ten constituents over the sample period. The proximate statistical driver is mega-cap weight concentration, which passive investing mechanically amplifies through market-cap-weighted inflows, though passive share cannot be cleanly identified as an independent driver within the available sample. The data therefore support the *H2* interpretation most cleanly for synchronicity, partially for concentration, and descriptively for asymmetry. What each finding shares is that none of the changes it captures is episodic or crisis-specific. They have accumulated gradually across three decades and the broader interpretation of what that accumulation means for the risk structure of the index is taken up in the conclusion.

6 Conclusion

This thesis set out to answer the following research question: *How does the increasing demand in index investing affect market efficiency?* Using S&P 500 constituent data from CRSP and Compustat, alongside Morningstar fund flow data over the period 1993 to 2024, we test two hypotheses. First, that passive capital inflows inflate valuation multiples beyond fundamental justification, and second, that passive dominance increases systematic fragility through reduced price informativeness, risk concentration, and correlation asymmetry.

6.1 *H1*: Price distortion

The evidence supports *H1*. Passive share is cointegrated with P/B median, P/E median, and P/E EW, confirming a genuine long-run equilibrium relationship rather than a spurious trend correlation. ECM shows that deviations from this equilibrium are self-correcting, with P/E median reverting at approximately 17.9% per month and P/B median at 5.9% per month. Subsample estimates show the speed of adjustment trending more negative after 2010, consistent with a tightening structural link between passive share and valuations as passive dominance has grown, though formal parameter stability tests do not reject a constant adjustment speed across the full sample.

Panel fixed-effects regressions confirm that the effect is concentrated in index members, relative to comparable non-index peers. A one-unit increase in orthogonalised passive share is associated with a 1.025 increase in log P/B and a 1.282 increase in log P/E for index members relative to non-index firms. This pattern is consistent with a price-pressure mechanism (Israel et al. (2021)) rather than a general market-wide valuation trend. The Granger causality tests show that passive share does not predict valuation changes at short lags, and that much of the apparent reverse causality from valuations to raw passive share disappears once the treasury yield correlation is removed. This result does not contradict *H1*; it is, in fact, what the literature would predict. Greenwood and Sammon (2024) document that the short-run index inclusion premium has largely vanished as active traders front-run predictable passive flows. That active traders efficiently process known, discrete rebalancing events does not imply efficiency in resisting a slow structural accumulation of non-informational capital. The absence of short-run Granger predictability is precisely the signature of premium that has migrated from short-horizon event spikes into chronic structural overpricing, one that Granger causality at monthly frequencies is unsuited to detect, but cointegration is not. The premium has not disappeared; it has transformed. Greenwood and Sammon (2024) show where it used to be; our results show where it went. This is consistent with *H1* operating as a slow-moving structural effect rather than a short-run predictive channel, as originally theorised by Israel et al. (2021) and formalised by Asness (2024).

6.2 *H2*: Systematic fragility

Taken together, the three fragility measures paint a consistent picture: the S&P 500 has grown more homogeneous in its price movements, more top-heavy in its risk distribution, and more prone to correlation spikes during stress. The evidence is strongest for synchronicity, partial for concentration, and descriptive for asymmetry, but all three point in the same direction. Price synchronicity rises significantly with passive share in first differences. A one percentage point increase in passive share is associated with a 4.49 unit increase in the average market model R^2, confirming that stocks move increasingly in lockstep with the market as passive capital grows, leaving less room for firm-specific information to be priced in, consistent with the framework of Roll (1988), who proposes the market R^2 model as an inverse measure of firm-specific price discovery.

Correlation asymmetry between up- and down-market days is a robust and statistically conclusive feature of the full sample. Average pairwise correlations are 0.033 higher on down-market days than on up-market days, a difference that is highly significant across more than 8,000 trading days. This confirms that diversification benefits weaken precisely when they are most needed, consistent with Chua et al. (2009). The annual regression of this asymmetry on passive share is a non-result ($\beta = 0.037, p = 0.433$). Correlation asymmetry is a robust structural feature of equity markets under stress, but its time-series variation is not explained by passive share in annual data. The crisis period evidence in Section 5.3.4 offers a descriptive complement, showing asymmetry is most pronounced during broad liquidity-driven panics at higher passive share levels, but those comparisons do not constitute a test.

The MCR concentration finding is the most nuanced result. The descriptive trend is economically striking and consistent with the mechanism proposed by von Moltke and Sløk (2024). The share of total S&P 500 risk attributable to the ten highest MCR constituents rose from 20.5% in 1993 to 53.1% by 2024. This trajectory is temporally consistent with the growth of passive investing and constitutes a material change in the risk structure of the index over the sample period. However, first-difference regressions show that monthly changes in passive share do not independently predict monthly changes in MCR concentration once the contemporaneous growth in mega-cap market weight is controlled for. The proximate driver of MCR variation is the weight of the largest constituents, which passive investing mechanically amplifies through market-cap-weighted inflows. Full causal identification of the passive investing channel requires a longer post-2017 sample or an exogenous instrument for passive ownership.

The crisis period daily synchronicity analysis provides supplementary evidence for *H2* using three cross-period comparisons with passive share levels of 14.7%, 26%, and 51.4%. Under the Equal-Weighted benchmark, consistent with the methodology of section 5.3.1, the down-minus-up synchronicity gap reaches +0.086 during the COVID crash, ten times the non-crisis baseline, while the GFC shows a gap of +0.008, statistically indistinguishable from normal periods despite comparable market volatility. This contrast is consistent with the mechanical selling channel becoming material only once passive capital crosses a higher passive threshold. The absolute value crisis-period effects are ordered as COVID > GFC > Dot-com across all three specifications, aligned with the passive share gradient. This cross-episode pattern rests on only three crises and is best read as descriptive support for *H2*. A Value-Weighted robustness check preserves the regression gradient, but does not replicate the finding due to the distinctive behaviour of mega-cap technology stocks during COVID. This sensitivity to weighting choice is an acknowledged limitation and discussed in section 5.3.5.

6.3 Limitations and future research

Several limitations warrant consideration when interpreting these findings. First, the central identification challenge, namely that passive share is a slowly accumulating, non-stationary series and no clean experiment is available to isolate exogenous variation in passive ownership across the full 1993–2024 horizon, it is partially addressed by the Russell reconstitution regression discontinuity design. That analysis provides causal evidence that passive index assignment inflates P/B for firms near the Russell 1000/2000 boundary, with a 24.5 log-point premium post-2007 that is absent in the pre-2007 period when passive share remained below approximately 25%. However, the Russell design identifies a local average treatment effect for mid-sized boundary firms, and its external validity for the very large S&P 500 mega-cap constituents that drive the main thesis results is not guaranteed. Whether the same mechanism operates at the same magnitude at the top of the market cap distribution remains an open question.

Second, the analysis is deliberately limited to S&P 500 large-cap US equities. The mechanisms examined here, price pressure from non-informational flows, mechanical risk concentration, and redemption-driven correlation, may operate differently in smaller markets, international indices, or fixed income, where passive penetration has grown more recently and arbitrage capacity differs. Extending the framework to these settings is a natural next step.

Third, the Morningstar fund flow data captures open-end funds and ETFs in the US equity category, but does not distinguish between retail and institutional passive investors, nor does it capture passive exposure embedded in target-date funds, insurance products, or separately managed accounts. The true passive share of equity ownership is therefore likely higher than our measure suggests, and the effects documented here may be conservative.

Fourth, the daily synchronicity analysis in Section 5.3.5 would benefit from daily ETF creation and redemption basket data, which would allow direct identification of passive fund selling pressure on individual trading days. Such data requires access to specialised institutional databases, such as vendor-compiled ETF creation and redemption records, which were not available within the project timeline. Daily ETF flow data remains a valuable and actionable pathway for future research.

The rise of passive investing represents one of the most significant structural shifts in financial markets of the past three decades (Asness (2024); Israel et al. (2021)). This thesis provides empirical evidence that this shift has measurable consequences for both valuation levels and market risk structure. The conventional case for passive investing rests on low costs and reliable market exposure. The results here suggest those benefits come with less visible costs: a gradual inflation of valuation multiples for index constituents beyond what fundamentals justify, and a risk structure that has grown increasingly concentrated in a small number of large firms. Whether these costs are sufficient to alter the investment calculus for the typical retail investor is a question this thesis cannot answer, but the evidence suggests it is a question worth asking.

APPENDIX

AI declaration

In the process of writing this thesis, AI-based language tools were used in two capacities. First, AI assistance supported the debugging and development of the data processing code written in Python. Second, AI tools were used to assist with language improvement and proofreading of the written text.

REFERENCES

Appel, I. R., T. A. Gormley, and D. B. Keim. 2016. Passive Investors, Not Passive Owners. *Journal of Financial Economics* 121:111–141. doi:10.1016/j.jfineco.2016.03.003.

Asness, C. 2024. The Less-Efficient-Market Hypothesis. *The Journal of Portfolio Management* 51:68–89. doi:10.3905/jpm.2024.51.1.068.

Bloomberg News. 2025. Active Funds 46.1% of U.S. Total, Passive Assets at 53.9%. Retrieved from Bloomberg Terminal.

Boone, A. L., and J. T. White. 2015. The Effect of Institutional Ownership on Firm Transparency and Information Production. *Journal of Financial Economics* 117:508–533. doi:10.1016/j.jfineco.2015.05.008.

Chua, D. B., M. Kritzman, and S. Page. 2009. The Myth of Diversification. *The Journal of Portfolio Management* 36:26–35. doi:10.3905/jpm.2009.36.1.026.

Clenow, A. F. 2019. *Trading Evolved: Anyone Can Build Killer Trading Strategies in Python*. Independent.

Fama, E. F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* 25:383–417. doi:10.2307/2325486.

fja05680. 2026. S&P 500 Historical Components & Changes (01-17-2026). GitHub. `https://github.com/fja05680/sp500/`. Accessed: February 2026.

Greenwood, R., and M. Sammon. 2024. The Disappearing Index Effect. *Journal of Finance forthcoming*. doi:10.2139/ssrn.4294297.

Grossman, S. J., and J. E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets. *The American Economic Review* 70:393–408.

Israel, R., K. Laursen, and S. A. Richardson. 2021. Is (Systematic) Value Investing Dead? *The Journal of Portfolio Management* 47:38–62. doi:10.3905/jpm.2020.1.188.

Lawrence, S., and J.-C. Plagge. 2023. The case for low-cost index-fund investing. Vanguard Research.

Morck, R., B. Yeung, and W. Yu. 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? *Journal of Financial Economics* 58:215–260. doi:https://doi.org/10.1016/S0304-405X(00)00071-4. Special Issue on International Corporate Governance.

Roll, R. 1988. R2. *The Journal of Finance* 43:541–566. doi:https://doi.org/10.1111/j.1540-6261.1988.tb04591.x.

Sammon, M. 2024. Passive Ownership and Price Informativeness. *Management Science* 71:4582–4598. doi:10.1287/mnsc.2023.00836.

von Moltke, F., and T. Sløk. 2024. Assessing the Impact of Passive Investing over Time: Higher Volatility, Reduced Liquidity, and Increased Concentration. Tech. rep., Apollo Academy. Available at https://www.apolloacademy.com/the-impact-of-passive-investing/.