Stock Volatility and Index Fund Ownership

Shmuel Baruch* and Xiaodi Zhang†‡

Abstract

Low-cost index funds have broadened market participation, raising the concern that democratized access may make markets more fragile. We test this concern using the documented positive relation between index fund ownership and individual stock volatility. If that relation is driven by new investors entering through index products, their flows should increase the systematic component of stock volatility and broad-index volatility. We find little evidence of either. The relation between index ownership and volatility is concentrated in idiosyncratic stock volatility, and index-fund market share is not significantly related to S&P 500 or DJIA volatility. The evidence rules against a volatility footprint from broadened index-fund participation and is more consistent with internal reallocation by existing investors from individual-stock strategies toward index-centric strategies.

JEL Classification: G12, G14, G23

Keywords: Index fund ownership, Stock volatility, Index Volatility, Passive investing, Index funds

*Baruch is from Price College of Business, University of Oklahoma, Norman, OK 73019. E-mail: baruch@ou.edu

†Zhang is from the School of Finance, Shanghai University of Finance and Economics, Shanghai 200433, China. E-mail: zhang.xiaodi@sufe.edu.cn.

‡The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1 Introduction

As the index fund industry has grown into a widely used investment vehicle, it has attracted substantial participation.[1] Such activity can arise from two distinct sources: external flows to index funds (investors entering and exiting the market via index products, such as 401(k)s) and internal shifts to index strategies (existing participants reallocating between individual stock selection and index-centric approaches, such as factor investing).

An established empirical finding, documented by Ben-David, Franzoni, and Moussawi (2018) and Coles, Heath, and Ringgenberg (2022), is the positive link between individual stock return volatility and index fund ownership. What remains unsettled is whether this relation reflects external flows that expand the set of investors participating through index products, or instead reflects a change in how a given investor base participates in price discovery. This paper focuses on the testable implication of the external-flow explanation.

Our diagnostic logic follows from a basic economic observation about how index-linked demand is implemented. Changes in demand for index exposure are ultimately expressed through trading (and rebalancing) in the index basket. Consequently, index-linked trading pressure is concentrated at the aggregate (index) level. Thus, if the stock-level volatility effect is primarily caused by external flow shocks that inject noise through index products, that disturbance should not be confined to idiosyncratic stock movements: it should, first and foremost, load on the systematic component of constituent returns and, in the aggregate, on index volatility itself.

By contrast, internal reallocation has a different footprint. When sophisticated traders shift back and forth between direct stock selection and index-centric positions, they continue to participate in the price discovery of the market portfolio: explicitly when they trade index

[1] Recent evidence in Chinco and Sammon (2024) further shows that passive ownership can extend beyond registered index funds through internal indexing and related index-linked strategies.

instruments, and indirectly when they trade the underlying constituents. What changes is their direct participation in the price discovery of individual stocks. With fewer (or less intensive) stock-specific traders, stock-level prices can become noisier relative to fundamentals even if market-wide price discovery remains anchored. In this case, a rise in indexing can show up primarily as higher idiosyncratic volatility of constituents, with little or no corresponding change in either systematic volatility or index volatility.

This logic motivates a diagnostic empirical strategy: we examine not only whether index fund ownership predicts total stock volatility, but also *which component* of volatility moves (systematic vs. idiosyncratic) and whether the volatility of the aggregate index moves at all. These implications require no knife-edge assumptions about functional forms or distributions; they come from the institutional fact that indexing aggregates trades and from the economic distinction between (i) shocks that enter through the index basket and (ii) changes in stock-level participation in price discovery.

Our empirical strategy exploits the rich cross-sectional and time-series variation in index fund ownership generated by overlapping index memberships and heterogeneous index demand, rather than relying on infrequent index reconstitutions.[2] Using a large panel of U.S. stocks from 1980–2022, we first confirm the established positive relation between index fund ownership and individual-stock volatility.

We then decompose each stock's annual volatility into systematic and idiosyncratic components using a market-model regression and the corresponding sums-of-squares decomposition. The key diagnostic result is that the positive relation with indexing is concentrated in idiosyncratic volatility: higher lagged index fund ownership is positively related to subsequent idiosyncratic volatility, while we do not detect a statistically significant relation with systematic volatility.

[2]For example, Apple and Nike both belong to the price-weighted DJIA and the value-weighted S&P 500, but only Apple belongs to the QQQ. Money flowing in and out of the QQQ thus creates differences between the two stocks, which, after controlling for other explanatory variables, enable us to isolate the impact of changes in index fund ownership on stock volatility.

Finally, we test the aggregate implication directly. If external flows through index funds were the primary source of the stock-level volatility effect, we would expect a corresponding positive relation between index-fund market share and the volatility of broad indices. In annual time-series regressions for the S&P 500 and the DJIA (and their flagship ETFs), we find no such relationship.[3] Taken together, these results are difficult to reconcile with external flows as the primary driver of the stock-level volatility effect.

To keep the paper self-contained and to connect the above diagnostic to a transparent price-discovery mechanism, we present a simple model in the spirit of Treynor (1987). The model shows that, because demand shocks enter through the index basket, external flows should be reflected not only in constituent-stock volatility but also in systematic and index-level volatility. Moreover, the model predicts that, for a broad value-weighted index such as the S&P 500, an external-flow effect large enough to increase the variance of a median constituent should imply a much larger increase in the variance of the index fund. In our S&P 500 calibration, the model-implied index-fund variance increase is several orders of magnitude larger than the corresponding increase for the median constituent. Empirically, the predicted footprint is absent: index-fund market share is not significantly related to broad-index volatility; the estimated relation is far smaller than the model-implied benchmark from the S&P 500 calibration; and index ownership is not significantly related to systematic stock volatility.

The model clarifies what our tests can and cannot identify. They do not reveal the identities of marginal investors, but they do test whether the stock-level volatility relation is accompanied by the systematic and index-level volatility responses implied by external flows. The absence of statistically significant responses rules against the external-flow interpretation. At the

[3]Barberis, Shleifer, and Wurgler (2005) document a positive link between indexing and stock comovement. An increase in comovement, however, does not necessarily imply higher market-level volatility. Return comovement (R^2 or pairwise correlation) reflects the relative share of systematic variance versus idiosyncratic noise. As shown in Baruch and Zhang (2022), when passive investing reduces stock-specific price discovery without altering market-wide information aggregation, stock comovement increases while aggregate market volatility remains unchanged.

same time, the evidence is consistent with an internal-reallocation mechanism.

Our paper contributes to the broad literature on the consequences of index investing. For a comprehensive survey of this field, we refer the reader to Allard, Farkas, Pedio, Rubio, and Tonks (2025). Our work is specifically situated at the intersection of two key research areas.

First, we add to the literature that investigates the macroeconomic and market-wide consequences of index trading. For example, Sushko and Turner (2018) show that ETF trading can affect the returns of the aggregate bond index, and Anadu, Kruttli, McCabe, and Osambela (2020) examine how passive investment may affect broader financial market stability. Our contribution is to test whether the documented stock-level volatility relation has the footprint of external flows into indexed products. This distinction matters because external flows should affect both constituent stocks and the aggregate index, whereas internal shifts can increase stock-level volatility without producing a comparable increase in index-level volatility, a possibility formalized in the rational-expectations model of Baruch and Zhang (2022). The empirical pattern we document therefore speaks not only to whether indexing matters, but also to how it matters.

Second, our study contributes to the literature on equity mutual fund flows. While this literature often focuses on flow-performance sensitivity (see Dannhauser and Pontiff (2025) for a comprehensive study and Hortaçsu and Syverson (2004) for a focused analysis of S&P 500 funds), our findings offer a new perspective on what fund-flow explanations should imply in prices. We do not need to observe the identity of the marginal investor to test the external-flow volatility story, because that story has an observable implication: index-flow pressure should move the systematic component of constituent volatility.

2 A Model of the Price Discovery Process in the Presence of Indexers

We now develop a simple model that formalizes the diagnostic logic described in the introduction. Our approach extends the model of Treynor (1987) by incorporating both index and stock traders, which allows us to derive testable predictions. We consider an index consisting of K stocks, indexed by $k \in \{1, 2, \ldots, K\}$. Each constituent stock k trades at a current price p_k per share, with a future price v_k per share.

The index is defined as

$$\text{Index} = \frac{1}{\text{divisor}} \sum_{k=1}^{K} n_k p_k,$$

where the divisor is some constant. If n_k is the number of shares outstanding, then the index is value-weighted (e.g., S&P 500). If all the n_k's are equal (for example, $n_k = 1$), then the index is price-weighted (e.g., the DJIA).

An index fund that tracks the index issues shares. Each share trades at a current price p_F, with a future price v_F, and is backed by a portfolio of stocks. Specifically, there exists a *portfolio scaling parameter* θ such that each fund share represents ownership of a portfolio containing θn_k shares of stock k. Accordingly, the Net Asset Value (NAV) of the fund is given by

$$\text{NAV} = \theta \sum_{k=1}^{K} n_k p_k.$$

This setup ensures that the percentage return on the fund NAV equals the percentage change in the index. A reasonable choice for the portfolio scaling parameter is the reciprocal of the index's divisor, since in that case the fund's NAV equals the index.[4]

We make several simplifying assumptions: transaction costs, bid-ask spreads, and cre-

[4]The portfolio scaling parameter is not a term used in the industry. SPY's scaling parameter is 10% of the reciprocal of the divisor of the S&P 500, leading to a NAV that is exactly 10% of the index level. DIA's scaling parameter is 1% of the reciprocal of the divisor of the DJIA, leading to a NAV that is exactly 1% of the index level it tracks.

ation/redemption fees are absent, and fractional shares can be traded.

2.1 The Traders

There are I index traders, indexed by $i \in \{1, 2, \ldots, I\}$, who trade only shares of the index fund. There are also J stock traders, indexed by $j \in \{1, 2, \ldots, J\}$, who trade only shares of individual stocks.

Traders trade in proportion to what they perceive as asset misvaluation. Specifically, we assume that the ith index trader trades $\alpha(v_{Fi} - p_F)$ shares of the index fund, where v_{Fi} is the trader's subjective estimation of the future NAV of the fund share, and α is some positive constant, capturing the trading intensity. Thus, whenever an index trader perceives the fund to be undervalued, he buys, and whenever the trader perceives it to be overvalued, he sells fund shares. For simplicity, we assume that α is common to all index traders.

Similarly, the jth stock trader trades $\beta(v_{kj} - p_k)$ shares of stock k, where v_{kj} is the trader's subjective estimation of the future price of stock k, and β is the trading intensity. Thus, when a stock trader perceives a stock to be undervalued, he buys, and when he perceives it to be overvalued, he sells. Again, for simplicity, we assume that β is common to all stock traders and all stocks.

2.2 The price discovery process

Index funds exist in two forms: mutual funds and exchange-traded funds (ETFs). A mutual fund receives demand in the form of market orders (the demand is expressed in dollars or shares), and the fund managers determine the price and number of shares created or redeemed at NAV after the trading in the underlying assets has taken place. A static model like ours is not suitable to study price discovery by means of market orders. We therefore assume the

fund is organized as an ETF, and we explicitly model the authorized participants. These authorized participants are arbitrageurs who make profits from a discrepancy between the fund share price and its NAV and, in doing so, ensure that any difference between the price and the NAV is negligible.

Authorized participants may occasionally hold inventory for strategic purposes. However, for our analysis, we assume that their position is zero after the trade concludes, as they are neither investors nor speculators. Furthermore, to focus on how the price discovery transmits sentiments, we assume the authorized participants earn no arbitrage profit, and thus the fund shares trade at par with their NAV.

In our static model, there are three legs to the price-discovery process and they occur concurrently. For pedagogical reasons, we describe the process sequentially.

First, the share price of the fund, p_F, satisfies the market clearing condition:

$$0 = AP_F + \sum_{i=1}^{I} \alpha(v_{Fi} - p_F), \tag{1}$$

where AP_F represents the trading activity of authorized participants. The clearing condition states that the trade of the authorized participants offsets the index traders' order imbalance. In all likelihood, $AP_F \neq 0$. If it is positive, then the authorized participants redeem the AP_F shares from the fund issuer and receive $\theta n_k \times AP_F$ shares of each of the K stocks. They then sell those stock shares. If, on the other hand, AP_F is negative, the reverse happens and the authorized participants buy $\theta n_k \times AP_F$ shares of each of the K assets, and deliver this basket to the ETF issuer in exchange for AP_F shares of the fund. Either way, after the creation/redemption process is complete, the authorized participants hold zero shares of the index fund.

Second, the share prices of the underlying stocks satisfy the K clearing conditions

$$0 = -\theta n_k \times AP_F + \sum_{j=1}^{J} \beta(v_{kj} - p_k), \quad k = 1 \dots K. \tag{2}$$

The clearing conditions state that the trade of the authorized participants offsets the stock traders' order imbalance. Note that we put a minus sign in front of AP_F, because here the authorized participants undo the position they took in their ETF trading.

Finally, in the third leg, the current share price equals its NAV:

$$p_F = \theta n_1 p_1 + \theta n_2 p_2 + \ldots + \theta n_K p_K. \tag{3}$$

Inserting (3) into the clearing condition for the ETF shares, (1), solving for AP_F and then inserting the resulting AP_F into the underlying stocks clearing condition, (2), gives us a system of K equations for the equilibrium pricing of the K stocks:

$$0 = \theta n_k \sum_{i=1}^{I} \alpha \left(v_{Fi} - (\theta n_1 p_1 + \theta n_2 p_2 + \ldots + \theta n_K p_K) \right) + \sum_{j=1}^{J} \beta (v_{kj} - p_k), \quad k = 1 \ldots K \tag{4}$$

The system of equations is simultaneous, and the coupling arises from the presence of index traders. Even if the index traders themselves do not participate in stock trading, the authorized participants' trade conveys the index traders' sentiment. This system of equations reveals how index trading creates price linkages between seemingly unrelated stocks. The first term represents the index traders' influence on stock k's price, transmitted through authorized participants' arbitrage activities. The second term captures direct stock-specific sentiment. In equilibrium, each stock's price reflects both direct trading based on its own market sentiment and indirect pressure from index-level sentiment. This creates a mechanism of price comovement among index constituents that would not exist otherwise. When index traders become more numerous or active (higher I or α), the coupling between stocks strengthens, potentially causing correlated price movements even among fundamentally unrelated companies.[5]

[5]Barberis, Shleifer, and Wurgler (2005) show that stocks added to the S&P 500 have larger betas with the S&P 500, and they attribute their findings to market frictions or investor sentiment. Greenwood (2008) documents stronger comovement among stocks in the Nikkei index. This evidence further confirms that the extra comovement can result from commonality in investors' behavior. In addition, Cong, Huang, and Xu (2024) theoretically show that the introduction of composite securities increases comovement of underlying assets.

The model we have described so far is of subjective beliefs. Traders agree to disagree about pricing, meaning that every trader understands that the counterparty to the trade holds opposing views on the stock's value.

We refer to the difference between the actual future price and the stock traders' estimates as traders' errors. Specifically, for each of the J stock-traders and each of the K stocks, we define the trader's error as the deviation:

$$\underbrace{v_k}_{\text{future price}} - \underbrace{v_{kj}}_{\substack{\text{stock trader's} \\ \text{estimation of} \\ \text{future price}}} =: \underbrace{x_{kj}}_{\text{trader's error}}, \quad j = 1, \ldots, J, \; k = 1, \ldots, K. \tag{5}$$

Similarly, we refer to the difference between the unobserved actual NAV and the index traders' estimates as traders' error. Thus, for each of the I index-traders, we write

$$\underbrace{\sum_{h=1}^{K} \theta n_h v_h}_{\text{future NAV}} - \underbrace{v_{Fi}}_{\substack{\text{index trader's} \\ \text{estimation of} \\ \text{future NAV}}} =: \underbrace{x_{Fi}}_{\text{error}}, \; i = 1, \ldots, I. \tag{6}$$

We denote by

$$\bar{x}_k := \frac{1}{J} \sum_{j=1}^{J} x_{kj}$$

and

$$\bar{x}_F := \frac{1}{I} \sum_{i=1}^{I} x_{Fi}.$$

Let the dollar returns on the stocks be

$$r_k = v_k - p_k$$

and let the dollar return on the fund's NAV be

$$r_F = \sum_{h=1}^{K} \theta n_h r_h = \sum_{h=1}^{K} \theta n_h (v_h - p_h).$$

Next, we re-express the equilibrium pricing condition, (4), in terms of the dollar returns and errors rather than the asset prices and perceptions of future prices, which simplifies to

$$0 = \theta n_k \alpha I \left(r_F - \bar{x}_F \right) + \beta J \left(r_k - \bar{x}_k \right), \quad k = 1 \dots K \tag{7}$$

Multiplying (7) by θn_k and summing over k, we get a single linear equation in r_F and its solution is

$$r_F = \underbrace{\frac{I\alpha \sum_{k=1}^{K} \theta^2 n_k^2}{I\alpha \sum_{k=1}^{K} \theta^2 n_k^2 + J\beta}}_{=: a} \times \bar{x}_F + \underbrace{\frac{J\beta}{I\alpha \sum_{k=1}^{K} \theta^2 n_k^2 + J\beta}}_{= 1-a} \times \sum_{k=1}^{K} \theta n_k \bar{x}_k \tag{8}$$

That is, the dollar return on the fund's NAV is a weighted average of traders' errors.

After plugging r_F back into (7), we get a system of K linear equations in r_k that is no longer simultaneous, and therefore the solution of the system is straightforward. We have

$$\theta n_k r_k = \theta n_k \bar{x}_k + \frac{n_k^2}{\sum_{h=1}^{K} n_h^2} a \left(\bar{x}_F - \sum_{h=1}^{K} \theta n_h \bar{x}_h \right) \tag{9}$$

and adding up (9) recovers (8).

Equation (7) reveals the mechanism through which index trading affects individual stock returns. Each stock's return reflects both direct trading pressure from stock-specific sentiment ($\bar{x}_k$) and indirect pressure transmitted through index traders via the authorized participants' arbitrage activities. The term a (defined in (8)) captures the relative influence of index versus stock trading.

3 Index Trading and Volatility of Returns

We now adopt the perspective of an econometrician analyzing data, treating the traders' errors in valuations as random variables. This means it becomes meaningful to use the model

to compute the standard deviation (volatility) of the returns and relate it to the number of indexers in the economy.

We start by decomposing the traders' errors as

$$x_{kj} = \mu_k + \epsilon_{kj} \qquad \text{and} \qquad x_{Fi} = \sum_{k=1}^{K} \theta n_k \mu_k + \epsilon_{Fi} \tag{10}$$

On its own, this decomposition is vacuous. Nevertheless, it motivates the distributional assumptions we are about to impose. We interpret μ_k as representing common valuation errors affecting stock k that stem from two sources: exogenous information shocks (e.g., news events) and systematic biases in trader sentiment (e.g., media-driven optimism or pessimism that affects all traders). In particular, we do not assume that the μ_k's are statistically independent of each other, which allows the model to capture market risk or, more generally, factor risk.

We denote

$$\bar{\epsilon}_k := \frac{1}{J} \sum_{j=1}^{J} \epsilon_{kj}$$

and

$$\bar{\epsilon}_F := \frac{1}{I} \sum_{i=1}^{I} \epsilon_{Fi}.$$

We rewrite the dollar returns, (8) and (9), in terms of μs and ϵs:

$$r_F = \sum_{k=1}^{K} \theta n_k \mu_k + a\bar{\epsilon}_F + (1-a) \sum_{k=1}^{K} \theta n_k \bar{\epsilon}_k \tag{11}$$

$$\theta n_k r_k = \theta n_k (\mu_k + \bar{\epsilon}_k) + \frac{n_k^2}{\sum_{h=1}^{K} n_h^2} a \left(\bar{\epsilon}_F - \sum_{h=1}^{K} \theta n_h \bar{\epsilon}_h \right) \tag{12}$$

Before specifying our distributional assumptions, we connect the model to the empirical diagnostic. The literature has already established the stock-level relation: higher index fund ownership is associated with higher volatility in constituent stocks. The external-inflows

explanation implies more. If this stock-level pattern is driven by external inflows to index funds, the same channel should also increase the volatility of the index, and hence of the index fund, itself.

The reason is mechanical in our setup. Index-fund order imbalance is accommodated by authorized participants via creation/redemption, which is implemented by trading the underlying basket. As a result, any external-flow-driven distortion that shows up in the dollar returns of constituents is transmitted to, and aggregated in, the index-fund dollar return. For broad value-weighted indices, this aggregation also has a sharp magnitude implication: the implied increase in index-fund volatility is typically larger than the increase in the volatility of a median constituent.

Theorem 1 formalizes this implication under distributional assumptions chosen for tractability. These assumptions are meant to isolate the external-flow footprint, not to define external inflows narrowly. In particular, they do not preclude (i) the possibility that marginal index-fund participants become noisier with entry, nor (ii) the possibility that index-fund participants herd on a common sentiment. The theorem shows how index-fund volatility and constituent-stock volatility must change together under the external-inflows channel.

For any expression that depends on the index traders, we let Δ denote the forward difference operator. For example,

$$\Delta \operatorname{var}(\bar{\epsilon}_F) := \operatorname{var}\left(\frac{1}{I+1}\sum_{i=1}^{I+1}\epsilon_{Fi}\right) - \operatorname{var}\left(\frac{1}{I}\sum_{i=1}^{I}\epsilon_{Fi}\right).$$

We have the following.

Theorem 1. *Assume* $\operatorname{cov}(\bar{\epsilon}_F, \bar{\epsilon}_k) = 0$, $\operatorname{cov}(\bar{\epsilon}_F, \mu_k) = 0$, *and* $\operatorname{cov}(\bar{\epsilon}_k, \mu_h) = 0$ *for every* $h, k \in \{1, \ldots, K\}$.

Further assume that the random variables $\{\bar{\epsilon}_k\}_{k=1}^{K}$ *are independent and have common variance*

σ^2. *Then,*

$$\Delta \operatorname{var}(\theta n_k r_k) = \left(\frac{n_k^2}{\sum_{h=1}^{K} n_h^2} \right)^2 \Delta \operatorname{var}(r_F) \tag{13}$$

The proof of Theorem 1 is in Appendix B.

While Theorem 1 is derived in terms of dollar returns to preserve the linear tractability of the equilibrium mapping, empirical researchers conventionally test variances of percentage returns. Because dollar variances are not scale-invariant, one might be concerned that the dramatic difference between the volatility response of the index and the constituent stocks is merely an artifact of scaling. We translate the variance of dollar returns into the variance of percentage returns. From the perspective of the econometrician, the equilibrium price is a random variable reflecting fluctuating market sentiments. Therefore, the percentage return, $R = r/p$, is the ratio of two random variables.[6] Following the long-standing tradition in asset pricing of using first-order Taylor approximations to linearize returns around their unconditional means, akin to the return decomposition framework pioneered by Campbell and Shiller (1988), we map the unconditional variance of dollar returns to percentage returns.

Corollary 2 (Translation to Percentage Returns)**.** *Let $R_k = r_k/p_k$ and $R_F = r_F/p_F$ denote the percentage returns of the stock and the fund, respectively. Let $\bar{p}_k$ and $\bar{p}_F$ denote their unconditional mean prices. The increase in the percentage-return variance of the constituent stock is related to the increase in the percentage-return variance of the index by:*

$$\bar{w}_k^2 \Delta \operatorname{var}(R_k) \approx \left(\frac{n_k^2}{\sum_{h=1}^{K} n_h^2} \right)^2 \Delta \operatorname{var}(R_F) \tag{14}$$

where $\bar{w}_k = \frac{\theta n_k \bar{p}_k}{\bar{p}_F}$ is the unconditional expected weight of stock k in the index fund.

The proof is a straightforward application of the Delta method to the variances on each side of the identity in Equation (13) (i.e., a first-order Taylor expansion around the respective unconditional means $\bar{r}$ and $\bar{p}$) and is therefore omitted.

[6] Nowhere in our framework do we assume normality. Consequently, we are not forced to assign a price support that includes zero, ensuring the unconditional variance of the percentage return remains well-defined.

The analysis shows that under the external-inflows channel, an increase in constituent-stock volatility should be accompanied by an increase in index volatility, with the latter generally larger in magnitude. Specifically, using ETF holdings data from September 11, 2025, the theorem implies that the increase in the variance of SPY is 3,627 times the median increase in the variance of stocks in the index. This number arises because $n_k^2 \ll \sum_{h=1}^{K} n_h^2$.[7] In other words, we have

$$\Delta \operatorname{var}(r_k) > 0 \implies \Delta \operatorname{var}(r_k) \ll \Delta \operatorname{var}(r_F).$$

Corollary 2 shows that the dollar-return restriction in Theorem 1 survives the translation to percentage returns. Rearranging Equation (14) gives

$$\Delta \operatorname{var}(R_F) \approx \bar{w}_k^2 \left(\frac{\sum_{h=1}^{K} n_h^2}{n_k^2} \right)^2 \Delta \operatorname{var}(R_k). \tag{15}$$

Thus, the relevant stock-to-index variance multiplier is not any single factor in Equation (14), but the entire coefficient on $\Delta \operatorname{var}(R_k)$ after the equation is solved for $\Delta \operatorname{var}(R_F)$. For a broad value-weighted index, this coefficient is large for the median constituent, and more generally for most constituents: the aggregation term $\left(\sum_h n_h^2/n_k^2\right)^2$ dominates the attenuation from the squared portfolio weight $\bar{w}_k^2$. Hence, under the external-inflows channel, a positive increase in a constituent's percentage-return variance implies a positive, and typically much larger, increase in the index fund's percentage-return variance. For a price-weighted index, by contrast, the same coefficient is of order one for a representative constituent, so the model predicts a proportional rather than magnified increase in index-fund variance.

Therefore, if empirical data show increased volatility in R_k without a corresponding positive increase in the fund's percentage-return volatility, this evidence is difficult to reconcile with

[7] A closely related external-flows formulation treats new index-fund participation as exogenous, price-inelastic noise-trader demand for ETF shares. Such demand is accommodated through authorized participants' creation/redemption activity. In this variant, the flow shock loads directly on the index fund's dollar return, while its impact on each constituent stock's dollar return is scaled by the stock's index weight in the creation/redemption basket (proportional to n_k). As a result, the induced increase in constituent-stock dollar-return variance is accompanied by a (typically much larger, for broad value-weighted indices) increase in index dollar-return variance. Hence, the qualitative prediction behind Theorem 1 is unchanged if we model external flow as pure inelastic noise demand.

external index-flow pressure as the driver of the observed pattern. In our empirical tests, we find no statistical evidence of a proportional increase in DJIA volatility, nor of the larger increase in S&P 500 volatility predicted by the value-weighted case. More broadly, we find no evidence that index ownership is associated with higher market-level volatility. This makes the external-inflows hypothesis harder to reconcile with the data under our model's assumptions.

The evidence therefore rules against the volatility footprint implied by the external-flow explanation. At the same time, the observed pattern of higher stock-level volatility without higher systematic or index-level volatility can arise if the link between index-fund ownership and volatility primarily reflects existing investors substituting away from stock-specific price discovery toward index-centric strategies. This is the internal-reallocation mechanism studied in Baruch and Zhang (2022). To keep the paper self-contained, Appendix A records a simple partial-equilibrium benchmark for this consistency result, but the empirical test itself does not require us to identify reallocating investors directly.

4 Data and Empirical Analyses

This section presents the empirical analysis and interprets the results through the model's predictions. That said, we cannot directly observe whether variation in index fund ownership reflects changes in the investor base or strategic reallocation by existing investors; we can only regress volatility on the level of index fund ownership.

Our sample period spans from 1980 to 2022, giving us 43 years of data. We choose 1980 as our beginning year because Thomson Reuters S12 started to record mutual fund ownership data at the end of 1979. We collect stock data from CRSP and related accounting data from the CRSP-Compustat merged database. We estimate the regressions at the annual frequency.

To construct an annual proxy of index fund ownership for each constituent stock, we rely on the CRSP Mutual Fund and the Thomson Reuters S12 databases. From CRSP mutual fund data, we identify equity funds by restricting the CRSP objective code to "EDC" or "EDY". To select equity index funds, we either let the index fund flag be "D" or follow Appel, Gormley, and Keim (2016) by searching fund names that contain any of the following strings: {Index, Idx, Indx, Ind_ (where _ indicates a space), Russell, S & P, S and P, S&P, SandP, DOW, Dow, DJ, MSCI, Bloomberg, KBW, NASDAQ, NYSE, STOXX, FTSE, Wilshire, Morningstar, 100, 400, 500, 600, 900, 1000, 1500, 2000, or 5000}. This definition is designed to include open-end mutual funds, closed-end funds, and ETFs identified as index funds. Next, for each index fund, we use Thomson Reuters S12 to locate its quarterly stock ownership. With these data, we construct quarterly index fund ownership for each stock k by summing all shares held by index funds and dividing by shares outstanding in the corresponding quarter. Then our measure of index fund ownership of stock k in year t, Index-Fund-Ownership$_{k,t}$, is the average of quarterly ownership in year t.

To construct an annual measure of volatility, we use CRSP daily returns to compute the sum of squared deviations (SSD) for each stock k in each year t. Our measure of volatility is the annualized SSD, i.e. $\frac{252}{N_{k,t}-1}\,\text{SSD}_{k,t}$, where $N_{k,t}$ is the number of daily returns we have for stock k in year t. This measure is proportional to the conventional variance and therefore provides a legitimate proxy for volatility.

We run, at the annual frequency, the following regression:

$$\text{Volatility}_{k,t} = \alpha + \beta\ \text{Index-Fund-Ownership}_{k,t-1} + \text{Controls}_{k,t-1} + \varepsilon_{k,t} \tag{16}$$

where the control variables are lagged and include the lagged market capitalization (large-cap stocks typically have lower volatility), the firm's lagged leverage (higher leverage amplifies equity volatility, as per Modigliani-Miller Proposition 2), the lagged annual return (volatility increases following a negative return), and the lagged market-to-book value of equity (high

values suggest growth, and therefore are associated with higher volatility). We also control for lagged volatility and, following Ben-David, Franzoni, and Moussawi (2018), the lagged reciprocal of the price. Finally, we control for year and firm fixed effects. Table 1 Panel A provides the summary statistics of these stock-level variables; Appendix C provides detailed definitions.

We do not control for total institutional ownership because the relevant variation is the ownership share held through index funds rather than the total institutional presence in a stock. Conditioning on total institutional ownership would therefore absorb part of the shift toward indexed ownership that the paper is designed to study.

The results in Table 2 are consistent with the existing literature. Unlike reconstitution-based designs, our method exploits cross-sectional variation in index fund ownership generated by the multiplicity of overlapping indices: different stocks belong to different index sets.

To test whether external flows drive this baseline result, we decompose the stocks' return volatility into systematic and idiosyncratic volatilities. If external index-flow pressure drives the volatility-indexing relationship, we should observe a positive relation between stocks' systematic volatility and index fund ownership, not only a relation with idiosyncratic volatility.

To estimate the annual systematic and idiosyncratic volatilities, we do the following. We use CRSP daily stock returns and CRSP daily returns of the market portfolio, r_{mkt}. In each year and for each stock, we compute the ordinary least squares regression of the market model

$$r_{k,t,d} = \alpha_{k,t} + \beta_{k,t} \times r_{\text{mkt},t,d} + \varepsilon_{k,t,d}, \tag{17}$$

where d stands for a trading day in year t. We use the identity

$$\text{SSD} = \text{SSE} + \text{SSR}$$

for the decomposition, where SSE is the explained sum of squares and SSR is the residual sum of squares. Specifically, as for the stock's volatility regression in Table 2, we use $\frac{252}{N_{k,t}-1}\,\text{SSD}$

Table 1: Summary statistics

This table provides the summary statistics of variables used in the paper. See Appendix C for detailed variable definitions.

Panel A: Summary statistics of stock sample

Variable	Mean	P25	Median	P75	Std	Number
Index-Fund-Ownership	0.0329	0.0014	0.0081	0.0442	0.0508	156,207
Volatility	0.3763	0.0846	0.1832	0.4221	0.5415	156,548
Idiosyncratic-Volt	0.3367	0.0664	0.1509	0.3674	0.5190	156,544
Systematic-Volt	0.0400	0.0048	0.0162	0.0420	0.0661	156,544
MktCap($Million)	3618.02	120.42	465.92	1949.31	10149.17	155,892
Log(MktCap)	6.2263	4.7910	6.1440	7.5752	2.0322	155,892
Leverage	0.2142	0.0100	0.1416	0.3473	0.2290	152,955
1/Price	0.1928	0.0296	0.0557	0.1277	0.6492	156,350
Ret	0.1242	−0.2097	0.0606	0.3384	0.5793	156,275
Mkt-to-Book	3.2442	1.1984	1.9097	3.3735	4.4905	147,320

Panel B: Summary statistics of index sample

Variable	Mean	P25	Median	P75	Std	Number
Index-Fund-Market-Share	0.0414	0.0024	0.0312	0.0670	0.0421	43
Policy-Uncertainty	120.14	90.60	112.94	142.40	44.84	43
Log(Policy-Uncertainty)	4.7363	4.5065	4.7269	4.9586	0.3126	43
S&P 500						
Volatility	0.0325	0.0129	0.0216	0.0333	0.0324	43
Ret	0.0998	−0.00003	0.1278	0.2577	0.1628	43
S&P 500 ETF: SPY						
Volatility	0.0357	0.0124	0.0245	0.0495	0.0353	30
Ret	0.1100	0.0127	0.1428	0.2633	0.1809	30
DJIA						
Volatility	0.0308	0.0118	0.0215	0.0348	0.0307	43
Ret	0.0991	−0.0061	0.1185	0.2234	0.1423	43
DJIA ETF: DIA						
Volatility	0.0366	0.0135	0.0258	0.0473	0.0369	25
Ret	0.0931	0.0012	0.0981	0.2083	0.1523	25

as the dependent variable. In Table 3, for the stock's systematic volatility, we use $\frac{252}{N_{k,t}-2}$ SSE as the dependent variable, and for the stock's idiosyncratic volatility, we use $\frac{252}{N_{k,t}-2}$ SSR as the dependent variable.

The regression results in Table 3 are difficult to reconcile with the external-flow explanation. Index fund ownership is strongly related to idiosyncratic volatility, but not to systematic volatility. This pattern is consistent with internal reallocation, but our main conclusion is

Table 2: Volatility and Index Investment of Constituent Stocks

This table presents the regression results for the stock-level relationship between annual volatility and index investment. The dependent variable Volatility is the annualized sum of squared deviations of stock daily returns. Index-Fund-Ownership is the average index ownership of a stock in a year. MktCap is the market value of equity. Leverage is the long-term debt divided by the sum of long-term debt and market value of equity. 1/Price is the inverse of the stock price. Ret is the annual stock return. Mkt-to-Book is the market-to-book ratio. See Appendix C for detailed definitions. All regressions include firm and year fixed effects. Standard errors are clustered at both firm and year levels and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable: $\text{Volatility}_{k,t}$
$\text{Index-Fund-Ownership}_{k,t-1}$	0.3759^{***}
	(0.00)
$\log(\text{MktCap})_{k,t-1}$	-0.0342^{***}
	(0.00)
$\text{Leverage}_{k,t-1}$	0.2178^{***}
	(0.00)
$1/\text{Price}_{k,t-1}$	0.2092^{***}
	(0.00)
$\text{Ret}_{k,t-1}$	-0.0454^{**}
	(0.01)
$\text{Mkt-to-Book}_{k,t-1}$	0.0069^{***}
	(0.00)
$\text{Volatility}_{k,t-1}$	0.2479^{***}
	(0.00)
Firm FE	Yes
Year FE	Yes
Clustered SE	Yes
Observations	126,265
R^2	0.6578

the sharper rejection of the external-flow volatility footprint.

Having established that the pattern is concentrated in the idiosyncratic component, we conduct a final macro-level falsification test on the index itself. Our theory (specifically, Theorem 1) tells us that if the primary driver of the positive relation between stocks' volatility and index fund ownership is external index-flow pressure, then when we regress the volatility of the index itself on the index fund market share, we should find a similar positive relation. To that end, we test the relation with two indices, the S&P 500 and the DJIA. As a proxy for annual index fund market share, we use the assets under management of all equity index funds at the end of year t divided by the total market capitalization (i.e. sum of all equity values) at the end of year t as constructed by CRSP. In particular, we use the same value for

Table 3: Idiosyncratic and Systematic Volatility with Index Investment

This table presents the regression results for the stock-level relationship between index investment and either idiosyncratic or systematic volatility. Idiosyncratic-Volt is the annualized residual sum of squares (SSR) from the market model regression for each stock. Systematic-Volt is the corresponding annualized explained sum of squares. Index-Fund-Ownership is the average index ownership of a stock in a year. MktCap is the market value of equity. Leverage is the long-term debt divided by the sum of long-term debt and market value of equity. 1/Price is the inverse of a stock price. Ret is the annual stock return. Mkt-to-Book is the market value of equity divided by book value of equity. See Appendix C for detailed definitions. All regressions include firm and year fixed effects. Standard errors are clustered at both firm and year levels and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable:	
	(1) Idiosyncratic-Volt$_{k,t}$	(2) Systematic-Volt$_{k,t}$
Index-Fund-Ownership$_{k,t-1}$	0.2889^{***}	0.0606
	(0.00)	(0.34)
log(MktCap)$_{k,t-1}$	-0.0457^{***}	0.0078^{***}
	(0.00)	(0.00)
Leverage$_{k,t-1}$	0.1857^{***}	0.0247^{***}
	(0.00)	(0.00)
1/Price$_{k,t-1}$	0.2200^{***}	-0.0008
	(0.00)	(0.52)
Ret$_{k,t-1}$	-0.0499^{***}	0.0069^{**}
	(0.00)	(0.04)
Mkt-to-Book$_{k,t-1}$	0.0060^{***}	0.0007^{**}
	(0.00)	(0.03)
Idiosyncratic-Volt$_{k,t-1}$	0.2152^{***}	0.0134^{***}
	(0.00)	(0.00)
Systematic-Volt$_{k,t-1}$	0.1689	0.3139^{***}
	(0.20)	(0.00)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Clustered SE	Yes	Yes
Observations	126,263	126,263
R^2	0.6566	0.6420

both indices.

As a measure of annual volatility, we use the same measure we used for stocks, namely the annualized SSD of the respective index. We run, at the annual frequency, the following regression:

$$\text{Volatility}_{\text{index/ETF},t} = \alpha + \beta\ \text{Index-Fund-Market-Share}_{t-1} + \text{Ret}_{\text{index/ETF},t-1} + \text{Log(Policy-Uncertainty)}_{t-1} + \text{S\&P 500 Dummy} + \varepsilon_{\text{index/ETF},t}, \quad (18)$$

where the lagged measure of policy uncertainty is as constructed by Baker, Bloom, and Davis (2016). We also include a dummy variable for the index, which serves as an index fixed effect. Table 1 Panel B provides the summary statistics of these index-level variables.

Table 4: Index Volatility and Index Fund Market Share

This table presents the regression results for the index-level relationship between annual volatility and index investment. The dependent variable Volatility is the annualized sum of squared deviations of daily returns of an index or its ETF. Index-Fund-Market-Share is the assets under management of all index funds divided by the market capitalization of equities at the end of a year. Ret is the annual return of an index or its ETF. Policy-Uncertainty is the economic policy uncertainty index. S&P500 Dummy is equal to 1 if the index is S&P 500 and 0 otherwise. See Appendix C for detailed definitions. Standard errors are clustered at the year level and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable:	
	(1) $\text{Volatility}_{\text{index},t}$	(2) $\text{Volatility}_{\text{ETF},t}$
$\text{Index-Fund-Market-Share}_{t-1}$	0.1582	0.1408
	(0.43)	(0.58)
Ret_{t-1}	−0.0095	−0.0163
	(0.75)	(0.63)
$\text{Log(Policy-Uncertainty)}_{t-1}$	−0.0079	−0.0138
	(0.78)	(0.66)
S&P500 Dummy	0.0018	0.0013
	(0.17)	(0.55)
Clustered SE	Yes	Yes
Observations	86	53
R^2	0.0291	0.0148

Table 4 shows no statistically significant relation between index fund market share and the volatility of the index. To explore this null result further, we examine the economic significance of the estimate in Column (1) of Table 4. The coefficient on lagged index fund market share is 0.1582 with a p-value of 0.43. This implies a 95% confidence interval of approximately $[-0.24, 0.56]$. This interval not only contains negative values, but its upper bound of 0.56 is economically modest. This finding contrasts with the prediction from our model's external-inflows scenario (Theorem 1). If external inflows were the driver, any increase in stock volatility must be accompanied by an increase in index volatility. Specifically, the theory predicts that the coefficient in the index volatility regression (Table 4) should be materially larger than the coefficient in the stock volatility regression (Table 2). Our empirical results show the opposite: the point estimate of 0.1582 is not only statistically

indistinguishable from zero, but it is actually smaller than the highly significant coefficient of 0.3759 we find for individual stocks. This evidence makes the external-inflows hypothesis harder to reconcile with the data under our model's assumptions.

5 Concluding Remarks

Prior studies document a positive association between indexation and stock volatility. This paper asks whether that pattern is driven by external flows into index products, and finds little evidence for that explanation.

While the data do not reveal the identity of the marginal participants, a price-discovery model allows us to test the indirect footprint of the external-flow explanation. External inflows into index products should primarily operate through index-level trading pressure and therefore should leave traces in systematic or market-level volatility, not only in idiosyncratic stock-level volatility. The empirical pattern we document is different: the association between index ownership and volatility is concentrated at the constituent level, while we find little evidence of a corresponding market-level response. This pattern is difficult to reconcile with external inflows from new investors as the primary driver of the stock-level volatility effect.

As a falsification test, we decompose stock volatility into its systematic and idiosyncratic components, and find that the association between higher levels of indexation and greater stock volatility is confined to the idiosyncratic part. This result is difficult to reconcile with the view that new entrants who participate through index products are the primary driver of the enhanced stock volatility.

Taken together, the evidence rules against a simple external-flow volatility story and is consistent with existing investors shifting to index-centric strategies. By doing so, they reduce their direct participation, contributing only indirectly to the price discovery of individual

stocks through authorized participants. This insight demonstrates that growth in index fund ownership is not monolithic: its consequences hinge on the underlying mechanism.

Greenwood and Sammon (2025) show that traditional index reconstitution events are losing power due to anticipation and diminished price effects. Our cross-sectional approach offers an alternative empirical design. Relative to reconstitution-based designs, this approach exploits broader cross-sectional and time-series variation in index exposure to study indexing's effects.

Appendix

A Internal Reallocation Benchmark

The main text uses the model to derive the systematic-volatility implication of external index-flow pressure. This appendix records a simple internal-reallocation benchmark showing that the opposite footprint can arise naturally when existing traders shift from individual stock picking to index tracking. The result is not needed to reject the external-flow volatility explanation. It simply shows that the empirical pattern we document, higher stock-level volatility without higher systematic or index-level volatility, is consistent with reduced stock-level price discovery.

To establish this consistency, our result hinges on three key assumptions. First, we specify a baseline relationship between the trading intensities of stock and index traders that ensures their aggregate market impact can be compared. These assumptions are stylized, but they allow the model to isolate the internal-reallocation channel and derive a sharp empirical prediction. Furthermore, this simplification is grounded in equilibrium theory. Baruch and Zhang (2022) study a rational expectations equilibrium with endogenous trading intensities and also find that internal reallocation leaves the price of the market portfolio unchanged. Here, in our partial equilibrium model, the assumption acts as a theoretical anchor. If, upon shifting, index traders trade at an intensity even lower (higher) than specified in the theorem, the resulting volatility of the index would be correspondingly lower (respectively higher). Our objective is simply to establish that internal reallocation is consistent with finding no relation between index volatility and index fund market share.

Second, we assume that when traders switch to indexing, they form their view on the overall market by aggregating the views they previously held on individual stocks. This seems reasonable. Under these first two assumptions, the aggregate perception of the index's future

value remains completely unchanged during the transition. As a result, both the equilibrium dollar return and the equilibrium price of the index are perfectly invariant to the internal shift.

Our third assumption is that trader-specific errors are uncorrelated across traders. This simply means that the component of a trader's error that is not explained by common market factors is unique to that trader. This assumption allows us to cleanly derive the result that as there are fewer stock traders contributing to the price discovery of individual stocks, the greater the volatility of the stock.

Formally, we have the following.

Theorem A.1. *Assume the following conditions hold:*

A) ***Equivalent Trading Intensity:*** *The trading intensities α (for index traders) and β (for stock traders) satisfy the relation $\alpha \sum_{h=1}^{K} \theta^2 n_h^2 = \beta$. When a trader switches, they adjust their intensity accordingly.*

B) ***Consistent Views:*** *A trader switching from stock picking to index tracking forms their new view on the fund's future value by aggregating their prior views on the constituent stocks:*

$$x_{F,I+1} = \sum_{k=1}^{K} \theta n_k x_{kJ} = \sum_{k=1}^{K} \theta n_k (\mu_k + \epsilon_{kJ}) \tag{A.1}$$

C) ***Trader-specific Errors:*** *For each stock trader j, let $z_{kj} = \theta n_k \epsilon_{kj} - \frac{n_k^2}{\sum_h n_h^2} \sum_h \theta n_h \epsilon_{hj}$. Then, $\text{var}\left(\frac{1}{J} \sum_{j=1}^{J} z_{kj}\right)$ is decreasing with J.*[8]

Then, a switch from stock trading to index trading increases individual stock volatility while having no impact on index volatility.

The proof of Theorem A.1 is in Appendix B.

[8]This assumption holds if the trader-specific errors, ϵ_{kj} are i.i.d.

B Proofs

Proof of Theorem 1. To prove the theorem, we only need to show that

$$\operatorname{var}(\theta n_k r_k) = \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 \operatorname{var}(r_F) + \text{ some terms that do not depend on index traders.} \tag{B.1}$$

We have

$$\operatorname{var}(r_F) \underset{(11)}{=} a^2 \operatorname{var}(\bar{\epsilon}_F) + \theta^2(a^2 - 2a)\sum_{h=1}^K n_h^2\sigma^2 + f, \tag{B.2}$$

where f absorbs all the terms that do not depend on I. Specifically,

$$f = \operatorname{var}\left(\sum_{k=1}^K \theta n_k \mu_k\right) + \operatorname{var}\left(\sum_{k=1}^K \theta n_k \bar{\epsilon}_k\right).$$

We also have

$$\begin{aligned}
\operatorname{var}(\theta n_k r_k) &\underset{(12)}{=} \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 a^2 \left(\operatorname{var}(\bar{\epsilon}_F) + \theta^2 \sum_{h=1}^K n_h^2\sigma^2\right) - 2\theta^2 a \frac{n_k^2}{\sum_{h=1}^K n_h^2} n_k^2\sigma^2 + g \\
&= \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 \left(a^2 \operatorname{var}(\bar{\epsilon}_F) + a^2\theta^2 \sum_{h=1}^K n_h^2\sigma^2 - 2a\theta^2 \sum_{h=1}^K n_h^2\sigma^2\right) + g \\
&= \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 \left(a^2 \operatorname{var}(\bar{\epsilon}_F) + (a^2 - 2a)\theta^2 \sum_{h=1}^K n_h^2\sigma^2\right) + g \\
&\underset{(B.2)}{=} \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 (\operatorname{var}(r_F) - f) + g \\
&= \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 \operatorname{var}(r_F) + \underbrace{g - \left(\frac{n_k^2}{\sum_{h=1}^K n_h^2}\right)^2 f}_{\text{does not depend on } I},
\end{aligned}$$

where g does not depend on I. Specifically,

$$g = \operatorname{var}(\theta n_k \mu_k) + \operatorname{var}(\theta n_k \bar{\epsilon}_k).$$

Thus, we have shown (B.1), and this concludes the proof.

■

Proof of Theorem A.1. We prove the theorem in two parts. First, we show that the switch has no impact on the fund's dollar return, r_F, and hence its volatility. Second, we show that the switch increases the volatility of individual stocks' dollar returns, r_k.

Part 1: Invariance of Index Return (r_F)

The first assumption on equivalent trading intensity simplifies the weight a in equation (8) to $a = I/(I+J)$. This means the fund's dollar return, r_F, becomes a simple average of the perceived future NAV across all $N = I + J$ market participants. The second assumption on consistent views ensures that when a trader switches strategies, their aggregate perception of the fund's future value does not change. Since the total number of traders is constant and the set of all traders' perceptions is unchanged, their average, r_F, is identical before and after the switch. Consequently, its variance, $\text{var}(r_F)$, is also unchanged.

Part 2: Increase in Stock Volatility ($\text{var}(r_k)$)

The dollar return for an individual stock k is given by:

$$\theta n_k r_k = \theta n_k(\mu_k + \bar{\epsilon}_k) + \frac{n_k^2}{\sum_{h=1}^K n_h^2}\left(r_F - \sum_{h=1}^K \theta n_h(\mu_h + \bar{\epsilon}_h)\right)$$

Rearranging this equation isolates the terms that depend on the number of stock traders, J. The terms μ_h are common errors and are unaffected by the switch. From Part 1, r_F is also invariant. Grouping all invariant terms together gives:

$$\theta n_k r_k - \underbrace{\left(\theta n_k \mu_k + \frac{n_k^2}{\sum_{h=1}^K n_h^2}\left(r_F - \sum_{h=1}^K \theta n_h \mu_h\right)\right)}_{\text{invariant terms}} = \theta n_k \bar{\epsilon}_k - \frac{n_k^2}{\sum_{h=1}^K n_h^2}\sum_{h=1}^K \theta n_h \bar{\epsilon}_h$$

By the theorem assumption, the variance of the right hand side is increasing as J decreases. It follows that the variance of $\theta n_k r_k$ must also increase as J decreases. This concludes the proof.

■

C Variable Definition

Variable	Definition
Stock Level	
Index-Fund-Ownership	The average of quarterly index ownership in a year of the individual stock. Quarterly index ownership is defined as shares held by equity index funds from Thomson Reuters S12 divided by common shares outstanding from CRSP at the end of each quarter.
Volatility	The annualized sum of squared deviations (SSD) of daily returns, $\frac{252}{N-1}$SSD, where N is the number of a stock's daily returns from CRSP in a year.
Idiosyncratic-Volt	The annualized residual sum of squares (SSR) from the market model regression using daily returns in a year from CRSP: $r = \alpha + \beta \times r_{\text{mkt}} + \varepsilon$, $\frac{252}{N-2}$SSR, where N is the number of a stock's daily returns r in a year and r_{mkt} is the daily value-weighted return of the market.
Systematic-Volt	The annualized explained sum of squares (SSE) from the market model regression using daily returns in a year from CRSP: $r = \alpha + \beta \times r_{\text{mkt}} + \varepsilon$, $\frac{252}{N-2}$SSE, where N is the number of a stock's daily returns r in a year and r_{mkt} is the daily value-weighted return of the market.
MktCap	Market value of equity (Compustat #199×#25) at the end of a year.
Leverage	Long-term debt (#9) divided by the sum of long-term debt (#9) and market value of equity (#199×#25) at the end of a year.
1/Price	The inverse of stock price (#199) at the end of a year.
Ret	The annual return constructed using monthly returns of the individual stock from CRSP in a year.
Mkt-to-Book	Market value of equity divided by book value of equity (#60) at the end of a year.
Index Level	
Index-Fund-Market-Share	The assets under management of all equity index funds at the end of year t divided by the total market capitalization. The total market capitalization is the sum of all equity values at the end of year t constructed by CRSP.
Volatility	The annualized sum of squared deviations constructed using daily returns of an index or its ETF, where the daily return of an index is defined as the percentage change of its price level and the daily return of an index ETF is from CRSP.
Ret	The annual return of an index constructed using its price levels or that of an index ETF constructed using monthly returns from CRSP.
Policy-Uncertainty	A normalized index of the volume of news articles discussing economic policy uncertainty constructed by Baker, Bloom, and Davis (2016). See www.PolicyUncertainty.com for details.

References

Allard, A., Farkas, M., Pedio, M., Rubio, S., and Tonks, I., 2025. A survey of the consequences of passive investment funds for financial markets. Working paper, University of Bristol.

Anadu, K., Kruttli, M., McCabe, P., and Osambela, E., 2020. The shift from active to passive investing: Risks to financial stability? Financial Analysts Journal 76 (4), 23–39.

Appel, I., Gormley, T., and Keim, D., 2016. Passive investors, not passive owners. Journal of Financial Economics 121 (1), 111–141.

Baker, S., Bloom, N., and Davis, S., 2016. Measuring Economic Policy Uncertainty. The Quarterly Journal of Economics 131 (4), 1593–1636.

Barberis, N., Shleifer, A., and Wurgler, J., 2005. Comovement. Journal of Financial Economics 75 (2), 283–317.

Baruch, S. and Zhang, X., 2022. The distortion in prices due to passive investing. Management Science 68 (8), 6219–6234.

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

Campbell, J.Y. and Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1 (3), 195–228.

Chinco, A. and Sammon, M., 2024. The passive-ownership share is double what you think it is. Journal of Financial Economics 157: 103860.

Coles, J., Heath, D., and Ringgenberg, M., 2022. On index investing. Journal of Financial Economics 145 (3), 665–683.

Cong, L., Huang, S., and Xu, D., 2024. The rise of factor investing: "passive" security design and market implications. Technical report, National Bureau of Economic Research.

Dannhauser, C. and Pontiff, J., 2025. Flow. Working paper, Villanova University.

Greenwood, R., 2008. Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights. The Review of Financial Studies 21 (3), 1153–1186.

Greenwood, R. and Sammon, M., 2025. The Disappearing Index Effect. Journal of Finance 80 (2), 657–698.

Hortaçsu, A. and Syverson, C., 2004. Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. Quarterly Journal of Economics 119 (2), 403–456.

Sushko, V. and Turner, G., 2018. The implications of passive investing for securities markets. BIS Quarterly Review, March.

Treynor, J., 1987. Market Efficiency and the Bean Jar Experiment. Financial Analysts Journal 43 (3), 50–53.

Vijh, A., 1994. S&P 500 trading strategies and stock betas. Review of Financial Studies 7, 215–251.