# Stock Volatility and Index Fund Ownership

Shmuel Baruch*and Xiaodi Zhang†‡

**Abstract**

The established link between the rise in index fund ownership and higher individual stock volatility is a well-known pricing distortion. Using a novel testing framework, we attribute this phenomenon to existing market participants reallocating from stock-picking to index strategies, rather than to external inflows to index funds (e.g., from new unsophisticated or herd-like investors entering via index products). In essence, we find that stock returns grow noisier not because the amateurs have arrived, but because the professionals have grown quiet.

**JEL Classification:** G12, G14, G23

**Keywords:** Index fund ownership, Stock volatility, Index Volatility, Passive investing, Asset pricing distortions, Index funds

*Baruch is from Price College of Business, University of Oklahoma, Norman, OK 73019.

†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 establishes itself as an efficient investment vehicle, it attracts further participation. Such growth in index fund ownership can arise from two distinct sources: external inflows to index funds (new investors entering the market via index products, such as 401(k)s) and internal shifts to index strategies (existing participants reallocating from individual stock selection to index-centric approaches, such as factor investing).

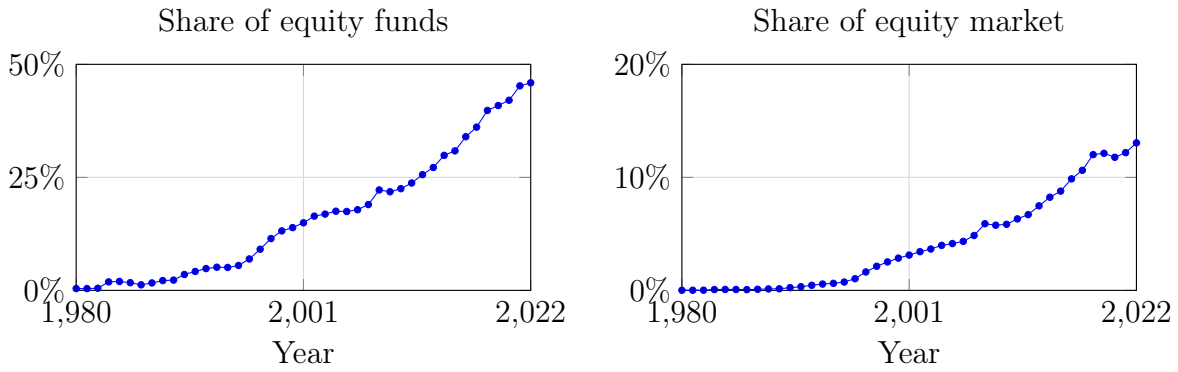



**Figure 1:** Growth in Index Fund Ownership (1980–2022). The left panel shows the share of equity funds held by index funds over time. During this time period, the share of index fund ownership grew from 0.4% in 1980 to 46% in 2022. The right panel shows index funds' share of the overall equity market. Index fund ownership grew from virtually zero to 13.05%. Fund and market capitalization data are from CRSP.

An established empirical finding, documented by Ben-David, Franzoni, and Moussawi (2018) and Coles, Heath, and Ringgenberg (2022), is the positive link between increased individual stock return volatility and growth in index fund ownership.[1] We consider this phenomenon a pricing distortion.[2] The goal of this paper is to identify which source of growth in index fund ownership drives this distortion: external inflows to index funds (new investors entering

[1] We attribute the lack of volatility trend concurrent with rising index fund ownership to the upward trend in market capitalization. As most index funds are value-weighted, index ownership is positively correlated with firm size, which is an established factor shown to have a negative cross-sectional relationship with volatility.

[2] In efficient markets, volatility reflects fundamental risks and forward-looking uncertainty. When index fund growth amplifies it artificially, without corresponding changes in company fundamentals, it misprices stocks by inflating perceived risk, potentially leading to suboptimal investment decisions. Thus, we consider the amplified volatility to be a distortion.

the market via index products) or internal shifts to index strategies (existing participants reallocating from individual stock selection to index-centric approaches).

Our methodological choice is to use fluctuations in index fund ownership as a proxy to identify the causal impacts of its overall growth trend. Stocks differ in their index memberships, generating both cross-sectional and time-series variations in index fund ownership.[3] Our empirical analysis exploits these variations and stands in contrast with the existing literature, which relies on the sporadic reconstitution of index membership.

While we confirm the positive relation between stock volatility and fluctuations in index fund ownership, we cannot directly observe whether changes in index fund ownership stem from external flows or internal shifts. Therefore, we develop a theoretical model that encompasses both sources of fluctuations and use the differential diagnosis approach to answer our research question.

Building on Treynor (1987), we develop a model that links volatility to indexing, whether through external inflows to index funds or internal shifts to index strategies. In this model, trade arises from the dispersion in traders' subjective estimates of stocks' future values. Optimistic traders buy from pessimistic ones at prices that reflect a compromise between their conflicting estimates. Consequently, prices represent the average of market participants' estimates, and dollar returns are the average deviation of those estimates from the true realized future values. As a result, the pricing distortion takes the form of a positive relationship between the dispersion of these average estimation errors and the number of index traders.

That the dispersion of the average error decreases as sample size increases (i.e., as the number of index traders contributing estimates to price discovery grows) is a robust statistical out-

[3]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.

come.[4] This sample size effect makes the established empirical finding naturally inconsistent with external inflows as a valid source for the increase in stock return volatility. However, valid economic reasoning permits assumptions that can override this statistical relationship, making the distortion consistent with external inflows to index funds.

Specifically, two sources of external inflows to the index funds are consistent with increased stock volatility. First, indexing's lower barriers to entry may attract unsophisticated investors to participate in the market and hence in the price discovery process. Here, we need to assume that fluctuations in these barriers drive fluctuations in levels of index fund ownership. It follows that each new unsophisticated trader introduces more noise. If the noise new traders bring to the market increases at a rate that exceeds the statistical benefits of increased sample size, then growth in the investor base can indeed drive the distortion. Second, new investors may engage in herding behavior, echoing public sentiment (e.g., following a media pundit) rather than contributing to price discovery. If the public sentiment is sufficiently erroneous, then again, external inflows can drive the distortion.

On the other hand, the established empirical finding is naturally consistent with internal shifts. Specifically, index traders participate only indirectly in individual stock price discovery. Thus, when traders shift from stock to index trading, the errors they contribute to individual stock prices are noisier, leading to noisier individual stock returns.

To sum up so far, our model suggests that indexing-related pricing distortions could arise both from external inflows, albeit under strong and potentially unrealistic assumptions, and from internal shifts. Given these theoretical limitations, we extend the scope of our theoretical analysis and use the model to predict the impact of the growth in index fund

[4]When we perceive traders' errors as exchangeable random variables, a standard variance calculation shows that as long as the errors are not perfectly correlated, the variance of the average decreases with the number of traders participating in trading. More generally, this is an example of a key principle: even when the standard Central Limit Theorem fails because the dispersion does not shrink to zero, that dispersion may still strictly decrease as the sample size grows.

participation on the volatility of the index itself. We find that the strong assumptions needed to justify external inflows as a source for the increase in stock volatility also imply that index volatility should not only increase with index fund market share, but should dramatically increase. In contrast, strategy reallocations may have no bearing on the pricing of the index. Now, we turn back to empirical analysis. We introduce a test examining index participation's impact on index volatility itself. We find no statistical relationship between these variables. This result is consistent with internal shifts and inconsistent with external inflows.[5]

Before we are ready to conclude that the increase in stock return volatility stems from internal shifts to index strategies, we run yet another test. If internal reallocation drives the volatility-indexing relationship, we should observe increases in stocks' idiosyncratic volatility but not in stocks' systematic volatility. We test this prediction by decomposing individual stock volatility into these two components. Consistent with internal shifts and inconsistent with external inflows, we find that an increase in index fund ownership of individual stocks increases the stocks' idiosyncratic volatility, but has no statistically significant impact on the stocks' systematic volatility.

In conclusion, although we cannot observe directly the sources of growth in index fund ownership, we provide strong evidence that the positive relationship between volatility and indexing is driven by internal shifts rather than external inflows to index funds.

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 consequences of index

[5]In Baruch and Zhang (2022), we exclusively study internal shifts. There, however, the setup is a rational expectation CAPM equilibrium. Also in that richer setup, we find that as investors withdraw from the price discovery process of individual stocks, stock prices become more volatile, while there is no impact on the index itself.

trading. For example, Sushko and Turner (2018) find that ETF trading affects the returns of the aggregate bond index, and Anadu, Kruttli, McCabe, and Osambela (2020) explore how passive investment impacts broader financial market stability. Our contribution is to provide a specific mechanism: the distinction between internal shifts and external flows that explains a key pricing distortion (stock volatility) arising from indexing..

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. By distinguishing the source of index fund growth, we provide a novel explanation for how this growth translates into specific market outcomes, complementing studies that analyze investor behavior within fund families.

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

We now develop a formal model that operationalizes the theoretical insights discussed 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* $\theta$ such that each fund share represents ownership of a portfolio containing $\theta n_k$ 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.[6]

We make several simplifying assumptions: transaction costs, bid-ask spreads, and creation/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 $i$th 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 $\alpha$ 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 $\alpha$ is common to all index traders.

[6]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.

Similarly, the $j$th 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 $\beta$ 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 $\beta$ 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, ensuring 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:

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

The system of equations is simultaneous, and the coupling arises from the presence of index traders. Even if the index traders themselves don't 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 wouldn't exist otherwise. When index traders become more numerous or active (higher $I$ or $\alpha$), the coupling between stocks strengthens, potentially causing correlated price movements even among fundamentally unrelated companies.[7]

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

[7] Barberis, Shleifer, and Wurgler (2005) show that stocks added to S&P 500 have larger beta with the S&P and they attribute their findings to market frictions or investors' sentiment. Greenwood (2007) documents stronger comovement among stocks in Nikkei index. His evidence further confirms that the extra comovement can be a result of commonality in investors' behavior. In addition, Cong, Huang, and Xu (2024) theoretically show that introduction of composite securities increases comovement of underlying assets.

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

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

Multiplying (7) by $\theta n_k$ 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 we can see that if we add up (9), we recover (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 $\mu_k$ 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 will not assume that $\mu_k$'s are statistically independent of each other, thus we allow for the model to capture market risk or more generally factors 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 $\mu$s and $\epsilon$s:

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

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

Before specifying our distributional assumptions, we outline the empirical patterns our model needs to explain: increased index trading is associated with higher volatility in constituent stocks, but there is no statistical evidence of any relationship, positive or negative, between index trading and the volatility of the index itself.

## 3.1 External inflows

It is already established in the literature that growth in index fund ownership is empirically linked to increased volatility in constituent stocks. The goal of this section is to show that if this increase is driven by external inflows to index funds, we should also observe an increase in the volatility of the index itself. While this prediction seems intuitive, developing the result analytically reveals that not only should the index volatility increase, but that this increase should be significantly larger than the rise in volatility observed in constituent stocks.

In Theorem 1, we will impose distributional assumptions that allow us to neatly deliver the result. Importantly, these distributional assumptions will not rule out the possibility

that $(\text{var}(\epsilon_{Fi}))_i$ is an increasing sequence, nor the rate at which it increases. Also, those distributional assumptions will not rule out that $\epsilon_{Fi}$ are replicas of each other, as if index traders merely echo some public sentiment, nor that this sentiment is very flawed. Thus, Theorem 1 addresses the possibility that what drives the pricing distortion is one of the two forms of growth in the investor base discussed in the introduction: unsophisticated traders introducing excessive noise and herding behavior based on erroneous public sentiment.

For any expression that depends on the index traders, we let $\Delta$ 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** (Volatility Co-movement of Index and Constituent Stocks - Part I)**.** *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* $\bar{\epsilon}_k$ *are independent with identical variance,* $\sigma^2$*. Then,*

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

The proof of Theorem 1 is in Appendix A.

The theorem shows that, under the stated assumptions, an increase in the volatility of the dollar return on stock $k$ must be accompanied by an increase in the volatility of the ETF's dollar return. Specifically, using ETFs holding 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. We get this number because $n_k^2 \ll \sum_{h=1}^{K} n_h^2$. In other words, we have

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

Therefore, if empirical data show increased volatility in $r_k$ without a corresponding and much larger increase in the volatility of the fund's dollar return, this would be inconsistent with growth in the investor base as the driver of the observed pricing distortion. In our empirical test, we not only fail to find a larger increase in the volatility of the index, but we also find no statistical evidence of any increase at all. This rules out growth in the investor base as the primary cause of the observed pricing distortion.

## 3.2 Internal Reallocation

Here we consider the possibility that the growth in index fund ownership is due to traders shifting from individual stock picking to index tracking. We will show that this internal reallocation is consistent with the empirical findings. The result hinges on three key assumptions. First, we specify a relationship between the trading intensities of stock and index traders that ensures their aggregate market impact can be compared. This assumption is contrived, but it plays the role of an anchor. If, once shifting, index traders trade at an intensity even lower (higher) than specified in the theorem, the resulting volatility of the index would be lower (respectively higher). Our only claim is that the 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 two assumptions, the resulting value of the index is unaffected by the composition of traders into index and stocks traders, and hence also the return on the index is unaffected.

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 2** (Volatility Co-movement of Index and Constituent Stocks - Part II)**.** *Assume the following conditions hold:*

*A)* ***Equivalent Trading Intensity:*** *The trading intensities $\alpha$ (for index traders) and $\beta$ (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:*

$$x_{F,I+1} = \sum_{k=1}^{K} \theta n_k x_{kJ} = \sum_{k=1}^{K} \theta n_k (\mu_k + \epsilon_{kJ}) \tag{13}$$

*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 2 is in Appendix A.

# 4 Data and Empirical Analyses

In this section, we present the empirical analyses that relate volatility to index fund ownership. This section is self contained in that a reader is not required to have carefully read

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

the model section. 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. The interpretation of the empirical findings relies on our theoretical framework.

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 should include all open-ended mutual funds, closed-end funds, and ETFs that are all called 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:

$$\text{Volatility}_{k,t} = \alpha + \beta\ \text{Index-Fund-Ownership}_{k,t-1} + \text{Controls}_{k,t-1} + \varepsilon_{k,t} \tag{14}$$

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

The results in Table 2 are consistent with the existing literature, although our method does not rely on the reconstitution of index membership but rather on the fact that there are so many indices, generating cross-sectional variation in index fund ownership (different stocks belong to different indices).

Our theory (specifically, Theorem 1) tells us that if the primary driver of the positive relation between stocks' volatility and index fund ownership is changes in the investor base, then when we regress the volatility of the index itself on the index fund ownership, we should find a similar, and even more pronounced positive relation. To that end, we will 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 asset 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 $t$

## Table 1: Summary statistics

This table provides the summary statistics of variables used in the paper. See Appendix B 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&P500** | | | | | | |
| 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&P500 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 constructed by CRSP. In particular, we use the same value for 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

**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 B 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 |

regression:

$$\text{Volatility}_{\text{index/ETF},t} = \alpha + \beta\ \text{Index-Fund-Market-Share}_{t-1} + \text{Ret}_{\text{index/ETF},t-1} + \text{Log(Policy-Uncertainty)}_{t-1} + \text{S\&P500 Dummy} + \varepsilon_{\text{index/ETF},t}, \quad (15)$$

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 (playing the role of index fixed effect). Table 1 Panel B provides the summary statistics of these index level variables.

We can see that the regression results in Table 3 show no statistical 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 3. The coefficient on

**Table 3: 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 asset 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&P500 and 0 otherwise. See Appendix B 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) |
| $\text{Ret}_{t-1}$ | −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 |

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 stands in stark contrast to 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 a dramatically larger increase in index volatility. Specifically, the theory predicts that the coefficient in the index volatility regression (Table 3) should be orders of magnitude 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. The data are therefore inconsistent with the external-inflows hypothesis, providing strong evidence against it.

To enhance our confidence in the conclusion we are drawing, we revisit the stock volatility

analysis. We decompose stocks' return volatility into systematic and idiosyncratic volatilities. We want to see if we find a positive relation between stocks' idiosyncratic stock volatility and index fund ownership, but find no relation between stocks' systematic stock volatility and index fund ownership.

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

$$r_{k,t,i} = \alpha_{k,t} + \beta_{k,t} \times r_{\text{mkt},t,i} + \varepsilon_{k,t,i}. \tag{16}$$

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 $\frac{252}{N_{k,t}-1}$ SSD as the dependent variable. In Table 4, 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 regressions' results in Table 4 not only reconcile the results in Table 2 and Table 3, but also give us great confidence in our conclusion that internal shifting is the driving force behind the link between index fund ownership and stock volatility.

# 5 Concluding Remarks

John C. Bogle, among others, argues that low-cost index investing promotes financial inclusion by allowing ordinary individuals to participate in market returns without high fees or expertise. However, this democratization may have come at a cost, as prior studies asso-

## Table 4: 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 B 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 |

ciate higher levels of indexation with greater stock volatility, a pricing distortion that could undermine market efficiency. In this paper, we employ a novel approach to show that this distortion is unlikely to be driven by ordinary individuals accessing markets via indexing.

While the data do not reveal the identity of the participants, we can, using a theoretical model of price discovery, detect their indirect effects. The intuition is straightforward: if external inflows from ordinary individuals were responsible for the increase in stock volatility, we would expect a corresponding rise in index volatility, which our tests do not detect.

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 bolsters our confidence that new entrants who participate in the price discovery of the market but not in the price discovery of individual stocks do not bear responsibility for enhanced stock volatility.

Instead, the source of the association lies in 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.

Our empirical strategy also adds methodological value. With traditional index reconstitution events losing power due to anticipation by sophisticated traders and diminished price effects Greenwood and Sammon (2025), our cross-sectional approach, exploiting natural variation in stocks' memberships across multiple indices, offers a robust alternative for causal inference on indexing's impacts, potentially serving as a tool for future research.[9]

[9]This robustness stems from several features that mitigate common endogeneity concerns. First, our identification leverages ongoing, natural variation in ownership from stocks' differing index memberships, providing a scalable source of cross-sectional and time-series differences that is not susceptible to anticipation biases as in rare reconstitution events Greenwood and Sammon (2025). Second, we use lagged index fund ownership in regressions to address reverse causality. Third, we employ firm fixed effects and year fixed effects; combined with controls (e.g., market cap, leverage), this minimizes omitted variable bias. Finally, the volatility decomposition serves as a falsification check, confirming effects only in idiosyncratic components as theory predicts, enhancing causal credibility through triangulation of evidence. While no method is immune to all biases, these elements collectively strengthen inference.

# Appendix

## A Proofs

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

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

We have

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

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

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

where $g$ does not depend on $I$. Specifically,

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

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

■

*Proof of Theorem 2.* 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:

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

We can rearrange this equation to isolate the terms that depend on the number of stock traders, $J$. The terms $\mu_h$ are common errors and are unaffected by the switch. From Part 1, $r_F$ is also invariant. Let's group all invariant terms together:

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

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.

■

# B Variable Definition

| Variable | Definition |
|---|---|
| **Stock Level** | |
| Index-Fund-Ownership | Proxy for external inflows and internal shifts, which is 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 asset 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.

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., 2025. 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 Analyst Journal 43 (3), 50–53.