

0.1 Data

The share of a company passively owned is the main variable of study, and to construct this variable, we need 3 datasets. To construct the share of passive ownership, we combine three datasets: CRSP’s Survivor-Bias-Free US Mutual Fund Database, which identifies index funds; LSEG’s S12 filings, which report quarterly mutual fund holdings; and WRDS MFLINKS, which links CRSP fund identifiers to LSEG filings.

The first of the three is CRSP’s Survivor Bias Free US Mutual Fund Database. This dataset contains the variable index fund flag, which flags funds that are either index fund based, purely index fund, or index fund enhanced. In this paper, we focus on purely index funds; however, in the robustness analysis, we allow for index fund based funds as well. The second dataset is LSEG’s S12 filings, which contain the quarterly holdings for US mutual funds. The third dataset needed is the Wharton Research Data Services’ MFLINKS dataset, which provides the codes necessary for merging the LSEG data with the CRSP data.

For each stock-quarter observation, we sum the shares held by funds flagged as index funds to get the total number of shares outstanding owned passively. The drawback of using this method is that we rely on funds flagging themselves as index funds. Funds that don’t identify themselves as passive but are in practice index funds are not recorded.

An alternative method to construct the share of passive ownership is to use the 13F filings as employed by Moltke & Sløk (2024). They compute the elasticity of demand of stocks for each fund and if the value is sufficiently close to 0, they flag the fund as a passive investor. A key limitation is that 13F filings do not distinguish between funds owned by the same filing manager. For example, Vanguard may simultaneously operate one fund that follows the index and another that actively trades, but both are aggregated under a single 13F report. Due to this limitation, we adopt the CRSP–LSEG–MFLINKS measurement technique.

In order to construct the dependent variable, we need daily stock data on prices. we use CRSP’s daily stock dataset to get daily prices, returns and volume data.

LSEG’s IBES dataset provides the earnings dates for each stock. Since exchanges are open from 9:00 am to 4:30 pm, if an earnings announcement takes place after 4:00 pm, we consider the earnings announcement date to be the next day. This is because the markets can only react to announcement after opening the next day.

We use Compustat’s quarterly fundamentals for controls and we use Wikipedia to construct data on stocks in the S&P 500 for my instrument.

My sample extends from January 1, 2021 to March 31, 2025 with a total of 17 quarters. The sample starts from 2021 to avoid affects of COVID-19. In total, there are 3502 unique stocks in the dataset.

0.2 Variables

The dependent variable is price volatility, defined as the standard deviation over a window. Since we are looking at a small event window (-30 to +15), we use three different time windows 3, 5 and 7 days. Choosing a large time period, such as 30 days, could lead to other confounding factors affecting the results, which is why we employ to smaller time periods. Figures 3, 4 and 5 provide trends in volatility around earnings dates. These figures clearly suggest that higher passive ownership is correlated with higher volatility and greater spikes in volatility post-earnings.

The main independent variable is the share of stock passively held, defined as:

$$\text{Passive} = \frac{\text{Outstanding Shares Passively Held}}{\text{Total Outstanding Shares}}$$

This variable captures the proportion of a firm’s equity passively held by mutual funds, ranging from 0 (no passive ownership) to 1 (fully passive ownership).

Since index-tracking funds must replicate the S&P 500 composition, inclusion in the index mechanically increases passive fund demand for the stock, making it a relevant instrument. The instrument variable is a dummy variable equal to 1 if a stock at time t was in the S&P 500 index and 0 otherwise.

Table 1 provides summary statistics for the variables.