Replication: Passive Investors, Not Passive Owners

Dexin Guo

Feb 28 2025

Overview

This write-up has four sections,

- First, the definitions and sample constructions of the replications in the replications, which analysis the possible reasons leads to inconsistent results compared with the original research.
- Second, discussions about the questions compared with Health et al.(2024)
- Third, the replications results from Table 1 to Table 8, which has been displayed with necessary information for comparison with the corresponding Table in the paper.
- Fourth display of an alternative way to discuss the function of active/passive share in corporate governance[1]

Definition and Differences

Following Appel et al.(2016),we define the key variables used in our replication as shown in Table 1. It is worth noted that, we don't include *Indicator for hedge fund activism* where Appel et al. required from Alon Brav. And the results in this replications are different from Appel et al. due to the following possible reasons:

Table 1: Variable definitions.

Variable name	Source	Definition
R2000	Russell investments	Indicator equal to one if firm is in the Russell 2000
Mutual fund ownership %	Thomson Reuters S12 files	% of shares outstanding held by mutual funds in September of year t
Passive %	Thomson Reuters S12 files	% of shares outstanding held in September of year t by passively managed funds
Active %	Thomson Reuters S12 files	% of shares outstanding held in September of year t by actively managed funds
Unclassified %	Thomson Reuters S12 files	% of shares outstanding held in September of year t by unclassified funds
Independent director %	Riskmetrics (Directors)	% of board seats held by directors classified as independent by Riskmetrics
Poison pill removal	Shark Repellent (FactSet)	Indicator equal to one if poison pill is withdrawn or allowed to expire at time t
Greater ability to call spec. meet.	Riskmetrics (Governance)	Indicator equal to one if shareholders better able to call a special meeting at time t
Indicator for dual class shares	Riskmetrics (Governance)	Indicator equal to one if a firm has dual class shares at time t
Mngt. proposal support %	Riskmetrics (Voting Results)	Percentage of "Yes" votes for management proposals
Shareholder gov. prop. support %	Riskmetrics (Voting Results)	Percentage of "Yes" votes for shareholder governance proposals
ROA	Compustat	Net income (ni) / total assets (at)

Sample Construction

The key challenge in this replication lies in the imputed values of mutual fund share holdings before May 2004, as funds were only required to report their holdings twice a year. Notably, we observed that some report dates fell in January. Consequently, if we strictly follow Appel et al. (2016)'s method of using the prior month's observations from CRSP to impute missing values, we would lose the sample for January 1998. To address this issue, rather than relying solely on prior-month data from CRSP for imputation, we adopt the following approach:

- First, we obtain data from CRSP, starting from 1998.
- Second, we impute missing observations using the nearest available data points. The original method assumes that missing observations align with those from the previous month. However, from a time-series perspective, the closest available observations are more likely to accurately reflect the missing values. Therefore this leads to more observations in *Total mutual fund ownership %* in our dataset.(see Table 1 for more details)

[1] Petajisto, A. (2013). Active Share and Mutual Fund Performance. Financial Analysts Journal, 69(4), 73–93. https://doi.org/10.2469/faj.v69.n4.7

Furthermore, potential inconsistencies in regression results, particularly in the heterogeneous analysis of mutual fund types, may stem from differences in how fund types are identified. Specifically, our code for `passive flag` is as following:

```
gen passive_flag = 0
gen fund_name = lower(fundname)
replace passive_flag = 1 if regexm(fund_name, "index|idx|indx|ind_|russell|s & p|s and
p|s&p|sandp|sp|dow|dj|msci|bloomberg|kbw|nasdaq
|nyse|stoxx|ftse|wilshire|morningstar|100|400|500|600|900|1000|1500|2000|5000")
```

where `fund_name` is collected from `Mflink` in `CRSP`, and `regexm`,the regular expression matching, acts as more effective ways for matching compared with traditional loop method. Besides, we use `CRSP Monthly Mutual Fund Summary` to merge with the dataset to identify `unclassified mutual fund`.

Selection Rules

More importantly, a stricter rule is implemented to calculate heterogeneous mutual fund ownership, we replace `passive_pct=. / active_pct=. /unclassified_pct=.` when

- the sum of these variables is larger than the `total mutual fund ownership`
- the shares of passive/active/unclassified is larger than the `SHROUT` in `CRSP`
- any observations with `missing(passive_pct)` or `missing(active_pct)` or `missing(unclassified_pct)` will be marked as missing heterogeneous ownership.[2]

Therefore, this is the potential reasons leads to less observations in heterogeneous mutual fund ownership

Discussion about the Disagreement with Heath et al(2024)

First, the main challenge in causal inference when studying the effects of index fund investment is the **endogeneity of fund holdings** by active and passive funds. Specifically, we can model governance as:

$$governance = f(index\ fund) + \epsilon$$

where missing firm characteristics (ϵ) are correlated with ownership and influence the dependent variable. Additionally, different firm policies may attract different types of investors, leading to **reverse causality**. Heath et al. (2024) address these challenges with the following advantages:

- The use of **firm-by-year fixed effects** is made possible by leveraging firm-fund-agenda level data, which accounts for time-varying firm characteristics.
- The **Russell index reconstitution in the post-2006 period** serves as a quasi-exogenous source of variation in fund holdings, allowing for the implementation of a **difference-in-differences (DID) model** that absorbs both unobserved firm heterogeneity and time-varying aggregate shocks.

This approach focuses on **increasing data dimensionality** (by including agenda-level information) and leveraging an **exogenous shock** for identification.

On the other hand, Appel et al. (2016) focus on constructing a comparable control group to estimate counterfactual outcomes, thus addressing omitted variable bias. They use **Russell 2000 inclusion as an instrument** for passive fund ownership. However, this **RDD-based** strategy may introduce the following issues (not only in Appel et al. (2016)):

- **Lack of Publicly Available Assignment Criteria**: Russell does not publicly release the **unadjusted May rankings**, which are the actual determinants of index assignment.
- **Alternative Rankings Create Bias**: Using **June rankings** (float-adjusted market capitalization) as a proxy is problematic because these rankings already incorporate index assignment, introducing endogeneity. Alternatively, imputing **unadjusted May rankings** using CRSP and Compustat data introduces measurement errors and biases.

In summary, the fundamental challenge with the **endogeneity of firm sorting for index inclusion** is that **the counterfactual is always unobserved**. We can never fully determine what would have happened had certain firms been included or excluded from the index. Two primary approaches to mitigate these concerns, as discussed in the literature, are:

- **RDD-based strategies** (e.g., Appel et al. (2016)) focus on **matching comparable observations** to infer causal effects.
- **DID-based strategies** leverage **exogenous shocks** to identify causal effects.
- **Expanding data dimensions and decomposition**—though not strictly parallel to the above approaches—focuses on analyzing firm-specific events and agendas, capturing firm-time fixed effects. This remains a compelling tool in causal inference.

[2]We use this strict rules because before collapse to generate the sum of each kind of mutual fund hold for each company, we have replaced the missing value of unclassified with 0, allowing for some firms that have no unclassified mutual fund holding their shares at certain points.

Replication Table in Appel et al.(2006)

Table 2: **Summary Statistics**

	N	Mean	Median	St.Dev
Total mutual fund ownership %	5636	32.596	33.185	3.334
Passive ownership %	2924	3.161	1.316	5.377
Active ownership %	2924	18.933	19.019	8.254
Unclassified ownership %	2924	1.986	1.323	2.070
Independent director %	4580	0.660	0.667	0.176
Poison pill removal	5636	0.549	1.000	0.498
Greater ability to call special meeting	5636	0.416	0.000	0.493
Indicator for dual class shares	5636	0.118	0.000	0.322
Mngt. proposal support %	1901	0.978	0.989	0.082
Shareholder gov. proposal support %	287	0.286	0.194	0.234
ROA	5636	0.007	0.010	0.043

Note: This table reports summary statistics of our key variables for our main sample: firms in the 250 bandwidth around the cutoff between the Russell 1000 and 2000 indexes from 1998–2006. Definitions for all variables are provided in Table A.1. Accounting variables are winsorized at the 1% level, and we delete observations where either mutual fund ownership is missing or total mutual fund holdings exceed a stock's market capitalization.

Table 3: **Impact of index assignment on mutual fund ownership**

Dependent variable =	*Percent of firm's common shares held by:*			
	All mutual funds	Passive	Active	Unclassified
	(1)	(2)	(3)	(4)
R2000	-0.260	-2.623	-2.365	-0.870*
	(0.330)	(1.835)	(1.819)	(0.507)
Bandwidth	250	250	250	250
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
# of firms	1,336	401	401	401
Observations	2,924	2,924	2,924	2,924
R-squared	0.13	0.11	0.24	0.16

Note: This table reports estimates of a regression of mutual fund holdings on an indicator for membership in the Russell 2000 index plus additional controls. Specifically, we estimate:

$$Ownership\%_{it} = \eta + \lambda R2000_{it} + \sum_{n=1}^{N} \chi_n((Ln(Mktcap_{it}))^n + \sigma Ln(Float)_{it} + \delta_t + u_{it}$$

where $R2000_{it}$ is a dummy variable equal to one if stock i is in the Russell 2000 index at end of June in year t, $Mktcap_{it}$ is the CRSP market value of equity of stock i measured at May 31 in year t, N is the polynomial order we use to control for $Ln(Mktcap_{it})$, $Float_{it}$ is the float-adjusted market value of equity (provided by Russell) at June 30 in year t, and δ_t are year fixed effects.

Mutual fund classifications are defined in the text. Standard errors are clustered at the firm level and reported in parentheses. The symbols * and *** indicate significance at the 10% and 1% levels, respectively.

Table 4: **Regression of Passive Ownership Scaled by Sample Standard Deviation**

Dependent variable = Passive % scaled by its sample standard deviation			
	(1)	(2)	(3)
R2000	-0.490	-0.469	-0.453
	(0.336)	(0.339)	(0.348)
Bandwidth	250	250	250
Polynomial order, N	1	2	3
Float control	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
# of firms	401	401	401
Observations	2,924	2,924	2,924
R-squared	0.24	0.24	0.24

Note: This table reports estimates of our first-stage regression of passive ownership onto an indicator for membership in the Russell 2000 index plus additional controls. Specifically, we estimate:

$$Passive\%_{it} = \eta + \lambda R2000_{it} + \sum_{n=1}^{N} \chi_n((Ln(Mktcap_{it}))^n + \sigma Ln(Float)_{it} + \delta_t + u_{it}$$

where $R2000_{it}$ is a dummy variable equal to one if stock i is in the Russell 2000 index at end of June in year t, $Mktcap_{it}$ is the CRSP market value of equity of stock i measured at May 31 in year t, $Float_{it}$ is the float-adjusted market value of equity (provided by Russell) at June 30 in year t, and δ_t are year fixed effects.

The model is estimated over the 1998–2006 period using a bandwidth of 250 firms around the Russell 1000/2000 threshold, and polynomial order controls for $Ln(Mktcap)$ of $N = 1, 2, 3$. Standard errors, ε, are clustered at the firm level and reported in parentheses. The symbol *** indicates significance at the 1% level.

Table 5: **Regression of Passive Ownership on Independent Director %**

	Dependent variable = Independent director %		
	(1)	(2)	(3)
Passive%	-0.007	-0.008	-0.008
	(0.012)	(0.012)	(0.012)
Bandwidth	250	250	250
Polynomial order, N	1	2	3
Float control	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
# of firms	81	81	81
Observations	312	312	312
R-squared	0.08	0.10	0.10

Note: This table reports estimates of our instrumental variable estimation used to identify the effect of ownership by passive investors on board independence. Specifically, we estimate:

$$Y_{it} = \alpha + \beta Passive\%_{it} + \sum_{n=1}^{N} \theta_n(Ln(Mktcap_{it}))^n + \gamma Ln(Float)_{it} + \delta_t + \varepsilon_{it}$$

where Y_{it} is the percentage of independent directors on the board of firm i in year t (from Riskmetrics) scaled by its sample standard deviation. $Passive\%_{it}$ is the percentage of shares outstanding owned by passively managed funds (as defined in the text) for stock i at the end of September in year t, scaled by its sample standard deviation.

$Mktcap_{it}$ is the CRSP market value of equity of stock i measured at May 31 in year t, $Float_{it}$ is the float-adjusted market value of equity (provided by Russell) at June 30 in year t, and δ_t are year fixed effects.

We instrument $Passive\%$ in the above estimation using $R2000_{it}$, an indicator equal to one if firm i is part of the Russell 2000 index in year t.

The data consist of firms in the two Russell indexes for which we obtain holdings data from Thomson Reuters Mutual Fund Holdings Database and which we match with data from the monthly CRSP file.

The model is estimated over the 1998–2006 period using 250 firms around the Russell 1000/2000 threshold, and polynomial order controls for $Ln(Mktcap)$ of $N = 1, 2$, and 3. Standard errors, ε, are clustered at the firm level and reported in parentheses. *** indicates significance at the 1% level.

Table 6: **Regression of Passive Ownership on Independent Director % (Split by Sample Years)**

	Dependent variable = Independent director %					
	Sample years = 1998–2002			**Sample years = 2003–2006**		
	(1)	(2)	(3)	(4)	(5)	(6)
Passive%	0.002	0.001	0.001	-0.056***	-0.058***	-0.060***
	(0.013)	(0.013)	(0.013)	(0.018)	(0.018)	(0.020)
Bandwidth	250	250	250	250	250	250
Polynomial order, N	1	2	3	1	2	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# of firms	62	62	62	29	29	29
Observations	255	255	255	57	57	57
R-squared	0.06	0.08	0.08	0.33	0.34	0.35

Note: This table reports estimates of the second-stage regression of our instrumental variable estimation used to identify the effect of passive investors on the percentage of independent board directors before and after the 2002 change in exchange-listing requirements regarding board independence.

The estimation is the same as in Table 4, except we now separately estimate the model over the 1998–2002 and 2003–2006 time periods using a bandwidth of 250 firms around the Russell 1000/2000 threshold, and polynomial order controls for $Ln(Mktcap)$ of $N = 1, 2$, and 3.

Both the dependent variable and $Passive\%$ are scaled by their sample standard deviations. Standard errors, ε, are clustered at the firm level and reported in parentheses. The symbols * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$ indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7: **Regression of Passive Ownership on Poison Pill Removal and Special Meeting Ability**

	Dependent variable = Takeover Defenses					
	Poison pill removal			**Greater ability to call special meeting**		
	(1)	(2)	(3)	(4)	(5)	(6)
Passive%	-0.030	-0.031	-0.030	0.010	0.012	0.011
	(0.031)	(0.031)	(0.031)	(0.027)	(0.026)	(0.026)
Bandwidth	250	250	250	250	250	250
Polynomial order, N	1	2	3	1	2	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# of firms	401	401	401	401	401	401
Observations	2924	2924	2924	2924	2924	2924
R-squared	0.04	0.04	0.05	0.02	0.04	0.04

Note: This table reports estimates of our instrumental variable estimation used to identify the effect of institutional ownership by passive investors on takeover defense outcomes. Specifically, we estimate:

$$Y_{it} = \alpha + \beta Passive\%_{it} + \sum_{n=1}^{N} \theta_n (Ln(Mktcap_{it}))^n + \gamma Ln(Float)_{it} + \delta_t + \varepsilon_{it}$$

where Y_{it} is the governance variable for firm i in year t scaled by its sample standard deviation. $Passive\%_{it}$ is the percentage of shares outstanding owned by passively managed mutual funds for stock i at the end of September in year t, scaled by its sample standard deviation.

$Mktcap_{it}$ is the CRSP market value of equity measured at May 31 in year t, and $Float_{it}$ is the float-adjusted market value of equity provided by Russell at June 30 in year t. The governance variables investigated include poison pill removals (PPILL) and greater ability to call special meetings (LSPMT), sourced from Shark Repellent and Riskmetrics.

We instrument $Passive\%$ using $R2000_{it}$, an indicator equal to one if firm i is part of the Russell 2000 index in year t. The data consists of firms in the Russell 1000/2000 indexes for which we obtain mutual fund holdings data from Thomson Reuters and CRSP. The model is estimated over the 1998–2006 period using a bandwidth of 250 firms, and polynomial order controls for $Ln(Mktcap)$ of $N = 1, 2, 3$.

Standard errors, ε, are clustered at the firm level and reported in parentheses. The symbols *** indicate significance at the 1% levels.

Table 8: **Regression of Passive Ownership on Dual Class Shares Indicator**

	Dependent variable = Indicator for dual class shares		
	(1)	(2)	(3)
Passive%	0.045*	0.044*	0.045*
	(0.026)	(0.026)	(0.025)
Bandwidth	250	250	250
Polynomial order, N	1	2	3
Float control	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
# of firms	401	401	401
Observations	2924	2924	2924
R-squared	0.04	0.04	0.06

Note: This table reports estimates of our instrumental variable estimation used to identify the effect of passive investors on the likelihood of dual class shares. Specifically, we estimate:

$$Y_{it} = \alpha + \beta Passive\%_{it} + \sum_{n=1}^{N} \theta_n (Ln(Mktcap_{it}))^n + \gamma Ln(Float)_{it} + \delta_t + \varepsilon_{it}$$

where Y_{it} is an indicator equal to one if firm i has dual class shares in year t according to Riskmetrics, scaled by its sample standard deviation. $Passive\%_{it}$ is the percentage of shares owned by passively managed mutual funds, scaled by its sample standard deviation.

$Mktcap_{it}$ is the CRSP market value of equity at May 31 in year t, and $Float_{it}$ is the float-adjusted market value of equity (provided by Russell) at June 30 in year t. We instrument $Passive\%$ using $R2000_{it}$, an indicator equal to one if firm i is part of the Russell 2000 index in year t.

The data consists of firms in the Russell 1000/2000 indexes, with mutual fund holdings obtained from Thomson Reuters and CRSP. The model is estimated for 1998–2006 using a bandwidth of 250 firms around the Russell 1000/2000 threshold, and polynomial order controls for $Ln(Mktcap)$ of $N = 1, 2, 3$.

Standard errors, ε, are clustered at the firm level and reported in parentheses. *** indicates significance at the 1% level.

Discussion about the Alternative Discussion of Active Share

Petajisto(2018) decomposes the all-equity mutual funds based on the degree of active management, which inspires that instead of discussing the on the fund level (passive/active/unclassified), how the active share/passive share/unclassified share of the fund affects the dependent variables might help to further discuss the effect of these holdings on the firm performance and corporate governance, these SAS data are available on Petajisto's personal website `https://www.petajisto.net/data.html` and a brief summary of key variables are as following:

Table 9: Descriptive Statistics

	Obs.	**Mean**	**Median**	**Std. Dev.**	**Min**	**Max**
activeshare	81,158	0.766	0.827	0.226	0.000	1.000
activeshare_min	81,158	0.748	0.806	0.224	0.000	1.000