CFR (draft), XXXX, XX: 1–43

Russell Index Reconstitutions, Institutional Investors, and Corporate Social Responsibility

Simon Glossner

[1]*Catholic University of Eichstätt-Ingolstadt*

ABSTRACT

My paper discusses four empirical approaches of the Russell 1000/2000 index reconstitutions to identify the effects of institutional investors on firm outcomes. Unbiased empirical approaches suggest that between 1998 and 2006, firms ranked at the top of the Russell 2000 had at most a 2 percentage points higher ownership of *passive investors* than firms ranked at the bottom of the Russell 1000. There is no significant difference in *total institutional ownership* around the threshold. Thus, the quasi-experiment can only identify the effects of passive investors. I also find that passive investors have no significant effect on corporate social responsibility (CSR).

Keywords: institutional investors, passive mutual funds, regression discontinuity, Russell indexes, corporate social responsibility

JEL Codes: G23, G30, M14

Corresponding address: Catholic University of Eichstätt-Ingolstadt, Chair of Finance and Banking, Auf der Schanz 49, 85049 Ingolstadt, Germany. E-mail: simon.glossner@ku-eichstaett.de. I am deeply grateful to my advisor, Thomas Mählmann, for his guidance and invaluable comments. I also

ISSN XXXX-XXXX; DOI XXXXXXXX

1 Introduction

Do institutional investors affect firm outcomes? This question is particularly relevant given that institutional investors are by far the largest shareholders of public firms in the United States. On the one hand, institutional investors may reinforce managerial myopia by frequent trading and fragmented ownership (Shleifer and Vishny, 1990; Porter, 1992). On the other hand, well-informed institutional blockholders may reduce agency problems inside firms by monitoring managers (Shleifer and Vishny, 1986; Monks, 1995) or they may not actually care, especially indexers.

Empirically, however, it is difficult to identify the effects of institutional holdings because of endogeneity. Beginning with Mullins (2014), Boone and White (2015), and Crane *et al.* (2016), researchers have developed a quasi-experiment that exploits a discontinuity in the Russell indexes, which is thought to influence institutional holdings. The Russell 1000 index contains the thousand largest U.S. firms (in terms of market capitalization), whereas the Russell 2000 index consists of the next two thousand largest U.S. firms. The Russell indexes are rebuilt every year at the end of June. During the annual index reconstitution, Russell assigns a firm to one of its indexes based on the firm's unadjusted market capitalization measured at the end of May.

The basic idea of the Russell quasi-experiment is to compare firms ranked at the bottom of the Russell 1000 with firms ranked at the top of the Russell 2000. The assumption is that only subtle differences in the firms' end-of-May market capitalizations decide whether a firm is assigned to the bottom of the Russell 1000 or to the top of the Russell 2000 during the annual index reconstitution. Firms on either side of the threshold can therefore be considered "as if" randomly assigned. However, firms at the top of the Russell 2000 have higher institutional ownership than firms at the bottom of the Russell 1000.[1] As a result, by comparing firms around the index threshold, it seems to be possible to identify the effects of institutional investors on firm outcomes.

thank Ivo Welch (the editor), two anonymous referees, Wei Wei, Alex Young, Yen-Cheng Chang, Silvina Rubio, Sebastian Calonico, and Drew Dimmery for numerous helpful comments. In addition, I am very grateful to Russell Investments for providing Russell index data. I gratefully acknowledge financial support from the PhD scholarship of the Friedrich-Naumann-Stiftung für die Freiheit. I declare no conflict of interest.

[1] One reason for this is a discontinuity in index weights around the threshold. Another is the relative frequency of Russell 2000 indexers to Russell 1000 indexers.

Table 1 provides an overview of 24 studies of the effects of institutional investors on a wide range of firm outcomes (e. g. corporate governance, corporate social responsibility, or payout policy). Although the basic idea of the quasi-experiment is simple, researchers use four different empirical approaches to model the quasi-experiment.[2] Strangely, researchers have come to different conclusions about the difference in institutional ownership between the firms that lie close to either side of the threshold. For example, Mullins (2014) claims that firms at the top of the Russell 2000 have 10% *lower* institutional ownership, Boone and White (2015) claim that Russell 2000 firms have 34% *higher* institutional ownership, and Chang *et al.* (2015) claim no difference—all with the same data.

This raises some questions: First, why do researchers find different estimates on the difference in institutional ownership between the firms that lie close to the threshold? Second, what are appropriate empirical approaches to model the Russell quasi-experiment? Third, is the Russell quasi-experiment suitable to identify the effects of institutional investors on firm outcomes?

My paper provides a detailed discussion of the four most common empirical approaches of the Russell quasi-experiment. My main finding is that researchers should use estimates of Russell's *unadjusted end-of-May rankings* to control for the distance between the observations and the index threshold (and also to select the bandwidth around the threshold). Russell's *float-adjusted end-of-June rankings* should not be used because they introduce a severe selection bias. Russell uses the June ranks to determine the index weights of the firms within an index. The problem with the June ranks is that Russell gives firms intentionally lower June ranks if their market capitalizations are not free-floating. By nature, a firm has lower institutional ownership (measured by dividing by all shares) if a larger number of its shares are not available to the public.

Unfortunately, Russell keeps its unadjusted May rankings proprietary. Researchers thus have to estimate them based on data from CRSP or Compustat, which will result in misclassifications. Good estimates of Russell's true May ranks[3] reduce weak instrument or selection bias problems. Sim-

[2] The most common approaches are the sharp regression discontinuity based on float-adjusted June ranks, the fuzzy regression discontinuity based on unadjusted CRSP/Compustat May market caps, the instrumental variable approach by Appel *et al.* (2016), and the instrumental variable approach based on index switchers.

[3] Appendix A describes how I recreate Russell's unadjusted end-of-May ranks.

ulations suggest that the fuzzy regression discontinuity approach, which uses predicted index membership as an instrumental variable for actual index assignment, is most robust against false-positive findings.

My paper shows that between 1998 and 2006, firms at the top of the Russell 2000 had a statistically significant 0.4–0.6 percentage points higher ownership of *passive mutual funds* than firms at the bottom of the Russell 1000. They also show a statistically weak 0.5–1.9 percentage points higher ownership of *quasi-indexers*, who are institutional investors that hold a diversified portfolio of stocks for a long time period. In contrast to many previous studies, I do not find significant differences in *total institutional ownership* between firms around the threshold. These findings support the use of the Russell quasi-experiment to identify the effects of passive investors (primarily index funds) on firm outcomes, but not the effects of overall institutional investors.

I then use the Russell quasi-experiment to identify the effect of passive investors on corporate social responsibility (CSR). Three papers have attempted to learn this effect but have contradicting results: Rubio and Vazquez (2018) claim that firms at the top of the Russell 2000 have a 12–13 percentage points higher institutional ownership than firms at the bottom of the Russell 1000 and that their higher institutional ownership leads to significantly *less* CSR between 2004 and 2006. Chen *et al.* (2018) present a similar difference in institutional ownership but claim that institutional investors cause significantly *more* CSR between 2003 and 2006. And Hou and Zhang (2017) find no significant difference in total institutional ownership between firms around the threshold. Instead, the authors claim that firms at the top of the Russell 2000 have a 1 percentage point higher ownership of passive mutual funds. This higher ownership of passive mutual funds leads to significantly *less* CSR between 2003 and 2006. My aim is to replicate these studies and discuss their different findings.

My paper measures CSR with scores from KLD. KLD rates every firm in the Russell indexes since 2003. Like Hou and Zhang (2017), I find that between 2003 and 2006, firms at the top of the Russell 2000 had a 0.6–0.9 percentage points higher ownership of passive mutual funds than firms at the bottom of the Russell 1000. There is no significant difference in total institutional ownership. However, in contrast to Hou and Zhang (2017), I do not find that higher ownership by passive mutual funds leads to significantly less CSR. Neither an unbiased fuzzy regression discontinuity approach nor an instrumental variable approach suggests that passive mutual funds significantly affect CSR.

My paper contributes to the literature in three ways. First, it provides an overview and evaluation of the established empirical approaches of the Russell quasi-experiment. Second, it points out that between 1998 and 2006, the difference in institutional ownership between the firms that lie close to either side of the Russell 1000/2000 threshold was at most 2 percentage points, with the bulk of the difference stemming from passive investors. Previous studies often find higher numbers because they use Russell's float-adjusted end-of-June ranks in their approaches, which introduce a severe selection bias.[4] Third, it corrects the record concerning the relationship between institutional investors and CSR (e. g. Dimson *et al.*, 2015; Barko *et al.*, 2018; Dyck *et al.*, 2019). Unlike Hou and Zhang (2017), Chen *et al.* (2018), and Rubio and Vazquez (2018), I find that passive investors have no significant effect on CSR.

My paper is related to Wei and Young (2018). They also explain why the Russell quasi-experiment is biased when using Russell's float-adjusted end-of-June ranks. Wei and Young (2018) show that an unbiased approach of the Russell quasi-experiment yields no significant difference in institutional ownership around the threshold. My paper shows no significant difference in *total institutional ownership*, but a significant difference in ownership of *passive mutual funds*.

Another related paper is Appel *et al.* (2018). Similar to my paper, they provide a discussion of different empirical approaches of the Russell quasi-experiment. They argue that the instrumental variable estimation by Appel *et al.* (2016) is the preferred specification because this approach overcomes the shortcomings of previously used approaches. My paper critically discusses the approach by Appel *et al.* (2016) and recommends two modifications (see Section 4.3).

The paper proceeds as follows. Section 2 outlines the basics of the Russell quasi-experiment. Section 3 describes the data sample. Section 4 explains the different empirical approaches of the quasi-experiment, and Section 5 discusses these approaches. Section 6 identifies the effect of passive investors on CSR. Finally, Section 7 concludes.

[4] My paper is not the first paper that raises concerns about using Russell's float-adjusted June rankings. Although several previous papers (e. g. Mullins, 2014; Chang *et al.*, 2015; Appel *et al.*, 2016) have raised similar concerns, recent literature continues to use float-adjusted June rankings.

2 The Russell 1000/2000 Index Reconstitution

A correlation between institutional ownership and a firm variable may not indicate causality. Institutional investors may drive the firm outcome or institutional investors may self-select into firms that show certain characteristics. In addition, the direction of the effect is often ambiguous.

Some recent papers try to exploit a (supposed) discontinuity in institutional ownership at the Russell 1000/2000 index threshold. This section explains the Russell quasi-experiment. Subsection 2.1 provides a description of how the Russell indexes are rebuilt during the annual index reconstitution. Subsection 2.2 describes the basic idea of the Russell quasi-experiment. Finally, Subsection 2.3 points out severe empirical problems associated with this quasi-experiment.

2.1 Description of the Russell Index Reconstitution

Russell Investments provides the Russell indexes. The Russell 1000 index consists of the thousand largest firms (in terms of market capitalization) and the Russell 2000 index comprises the next two thousand largest firms. Russell indexes are value-weighted indexes.

At the end of each June, Russell Investments rebuilds the Russell indexes. Russell first ranks all U.S. exchange-traded firms according to their total (non float-adjusted) market capitalizations measured on the last trading day in May. Then, Russell assigns firms ranked from 1 to 1,000 to the Russell 1000 index and firms ranked from 1,001 to 3,000 to the Russell 2000 index. Finally, Russell undertakes the actual index reconstitution on the last Friday in June.

To determine index weights, Russell does not use the firms' end-of-May market capitalizations. Instead, Russell uses float-adjusted market capitalizations (the "free float") measured at the end of June, immediately prior to the index reconstitution. Russell float-adjusts the market capitalizations to take account for outstanding shares that are not available for purchase. Examples of this are cross-ownership stakes held by another member of a Russell index or large ownership stakes held by corporates, private individuals, or the government. The aim of these adjustments is to make indexes more liquid and investable. Russell keeps its rules of how it adjusts the firms' market capitalizations proprietary.

2.2 Exploiting the Discontinuity in Institutional Ownership

The basic idea of the Russell quasi-experiment is to compare firms ranked at the bottom of the Russell 1000 with firms ranked at the top of the Russell 2000. The assumption is that firms ranked near the threshold cannot control which index they are assigned to.

Russell indexes are value-weighted indexes, implying that firms at the top of an index receive higher index weights than firms at the bottom of an index.[5] The top graphs of Figure 1 show this discontinuity in index weights around the threshold. The smallest stocks in the Russell 1000 index have lower index weights than the largest stocks in the Russell 2000 index, although these stocks have similar market capitalizations. The bottom graphs of Figure 1 show the market capitalizations of the firms that lie close to either side of the threshold.

The discontinuity in index weights is hypothesized to create a discontinuity in institutional ownership (i.e. firms at the top of the Russell 2000 have higher institutional ownership) because many institutional investors are benchmarked against the Russell indexes. Because these investors have an incentive to hold index stocks, they may purchase relatively more from stocks ranked at the top of a value-weighted index than from stocks ranked at the bottom of a value-weighted index. Firms ranked at the top of the Russell 2000 indeed seem to have higher institutional ownership than firms ranked at the bottom of the Russell 1000. Researchers (e.g. Boone and White, 2015; Crane *et al.*, 2016) often present a graph similar to the top left one of Figure 2, which shows a large discontinuity in institutional ownership around the threshold. Unfortunately, the firms are sorted according to their float-adjusted June rankings. This turns out to be problematic because float-adjusted June ranks render the quasi-experiment invalid, as the next subsection will show.

A clean identification of the effect of institutional investors on firm outcomes requires that firms ranked near the threshold are "as if" randomly assigned to either side of the threshold. Index membership is determined by firms' market capitalizations on the last trading day in May. Whether a firm is ranked at the place 999 or at 1,001 depends on small differences in the firms' market capitalizations on the specific day in May. (Although a firm can influence its own market capitalization, it has no control on

[5] Appel *et al.* (2018) clarify that the expected difference in institutional ownership also depends on the relative levels of money tracking the two indexes.

the market capitalizations of other index firms.) The end-of-May market capitalization is the only factor that determines whether a firm is placed in the Russell 1000 or the Russell 2000. As a result, firms on either side of the threshold can be compared because these firms are "as if" randomly assigned. This, however, holds only if the firms are ranked according to the unadjusted May rankings.

2.3 Empirical Problems of the Russell Quasi-Experiment

The most important problem is that Russell does not use the unadjusted end-of-May market capitalizations to determine the firms' index weights, but relies on float-adjusted market capitalizations measured at the end of June. Any research which uses Russell's float-adjusted end-of-June rankings has two problems:

The first problem is that stock prices may change in the month between May and June. If, for example, a stock is ranked at the place 990 on the last trading day in May, it should be placed at the very bottom of the Russell 1000 index. However, the stock may experience a large increase in its market capitalization in June. If this happens, Russell will put the stock not at the place 990, but at a higher rank, for example, at 800. This violates the random assignment assumption because the stock's performance in June affects its end-of-June rank (and index weight). The June performance affects only the rankings within an index.

The second problem is more serious. Russell float-adjusts the market capitalizations when determining the index weights. Although these adjustments do not result in a change in actual index assignment, they change the firms' ranks within an index. With this procedure, Russell gives firms intentionally lower index weights if they have less free float.

The bottom left graph of Figure 2 shows the firms' percentage free float as of total market capitalization measured immediately after the index reconstitution at the end of June. A higher ratio indicates that more market capitalization is free-floating (i.e. that relatively more shares are available for purchase to institutions). The 100 firms at the bottom of the Russell 1000 have about a 50% free float, while the 100 firms at the top of the Russell 2000 have a near 100% free float.

Comparing illiquid firms with lots of unavailable shares to liquid firms with much free float violates the random assignment assumption. Firms with relatively high free float can and typically will have higher institutional ownership (measured against all shares) than firms with relatively low free float. The top left graph of Figure 2 confirms that free-floating firms at the top of the Russell 2000 index have significantly higher institutional ownership in September than non-floating firms at the bottom of the Russell 1000 index.

Russell's "endogenous" free-float adjustments are the primary reason behind the large difference in institutional ownership around the threshold. In fact, the discontinuities in both institutional ownership and free float disappear when I plot these variables against unadjusted May ranks (right graphs of Figure 2). To avoid a severe selection bias, one should thus not use Russell's float-adjusted June ranks in the quasi-experiment. Several previous papers have raised similar concerns (e. g. Mullins, 2014; Chang *et al.*, 2015; Appel *et al.*, 2016). In line with this literature, Wei and Young (2018) provide an in-detail explanation of why using the float-adjusted June ranks is problematic. As we will see, researchers have come up with different solutions for this problem.

3 Data and Descriptive Statistics

My paper uses the Russell 1000 and Russell 2000 index constituents from 1998 to 2006 to examine the Russell quasi-experiment.[6] I match the Russell data with accounting data from Compustat, institutional investor data from Thomson Reuters Spectrum 13f, mutual funds data from Thomson Reuters s12, and stock returns from CRSP. The final sample consists of 26,616 firm-year observations from 1998 to 2006, with 5,588 unique firms. Table 2 provides descriptive statistics for the data sample.

I create three institutional ownership variables in September of each year, two months after the annual index reconstitution of the Russell indexes. The first variable is *total institutional ownership*, which is defined as the percentage of outstanding shares held by institutional investors. The second variable is *ownership by quasi-index investors*. According to Bushee (1998), quasi-index investors are long-term institutional investors who hold highly diversified portfolios. Quasi-indexers can be active or pas-

[6] I do not use observations after 2006 because, in 2007, Russell changed its methodology for rebuilding its indexes. Specifically, Russell introduced a policy called "banding" to ensure that stocks do only switch between the indexes if they cross the threshold by more than just a few ranks.

sive investors. The third variable is *ownership by passive (index-holding) mutual funds*. To identify passive funds, I use the same procedure as Appel *et al.* (2016). Specifically, I classify a mutual fund as a passive fund when the fund is labeled as an index fund by the CRSP Mutual Funds Database *or* when the name of the fund includes a certain string.[7]

Table 3 reports the dollar amount of passive assets benchmarked to the Russell 1000 and Russell 2000. The amount of benchmarked passive assets is about 2 to 3.5 times bigger for the Russell 1000. For example, in 2006, $151.9 billion were passively tracking the Russell 1000 and $43.0 billion were passively tracking the Russell 2000.

4 Empirical Approaches of the Russell Quasi-Experiment

This section explains the four most common empirical approaches of the Russell quasi-experiment: the sharp regression discontinuity (based on Russell's float-adjusted June rankings), the fuzzy regression discontinuity (based on CRSP/Compustat May ranks), the instrumental variable approach by Appel *et al.* (2016), and the instrumental variable approach based on index switchers.

4.1 Sharp Regression Discontinuity Based on Float-Adjusted June Rankings

Boone and White (2015) and others (Lu, 2013; Wong and Yi, 2017; Khan *et al.*, 2017; Fang, 2018a; Fang, 2018b; Lin *et al.*, 2018; Chen *et al.*, 2019) use variants of the same approach. They exploit the discontinuity at the Russell indexes by running a "sharp" regression discontinuity (RD) design (Lee and Lemieux, 2010). In a sharp RD design, the probability of treatment jumps from 0 to 1 at the cutoff. Formally, the sharp RD design based on June rankings is specified by a panel regression

$$Y_{i,t} = \alpha + \tau \mathrm{R2000}_{i,t} + \delta \mathrm{Rank}_{i,t}^{Jun} + \gamma \mathrm{R2000}_{i,t} \mathrm{Rank}_{i,t}^{Jun} + \nu_t + \epsilon_{i,t}, \quad (1)$$

where $Y_{i,t}$ is the outcome variable of firm i at time t, $\mathrm{R2000}_{i,t}$ is a dummy that is one if firm i is a member of the Russell 2000 after the index reconstitution in year t and zero if firm i belongs to the Russell 1000, $\mathrm{Rank}_{i,t}^{Jun}$

[7]Following Appel *et al.* (2016), I use these strings: "Index", "Idx", "Indx", "Ind ", "Russell", "S & P", "S and P", "S&P", "SandP", "SP", "DOW", "Dow", "DJ", "MSCI", "Bloomberg", "KBW", "NASDAQ", "NYSE", "STOXX", "FTSE", "Wilshire", "Morningstar", "100", "400", "500", "600", "900", "1000", "1500", "2000", and "5000".

is the rank of firm i at time t, v_t are year dummies, and $\epsilon_{i,t}$ is the error term. I construct variable $\text{Rank}_{i,t}^{Jun}$ by ranking firms according to their float-adjusted June weights, which are provided by Russell. The ranks are centered around the cutoff, i. e. firms in the Russell 1000 have negative ranks and firms in the Russell 2000 have positive ranks. Time dummies are included because the indexes are reconstructed every year. I estimate equation 1 only on those observations that lie close to the threshold, which is a non-parametric method of estimating the RD.[8]

Table 4 presents the results of a sharp RD that is specified by equation 1. The outcome variable is institutional ownership. Coefficient τ of variable $\text{R2000}_{i,t}$ measures the difference in institutional ownership between the firms that lie around the threshold. The estimates show that firms at the top of the Russell 2000 have a *18.4–32.6* percentage points higher total institutional ownership than firms at the bottom of the Russell 1000. Columns 5 to 8 show that the bulk of the difference in institutional ownership stems from quasi-index investors.

A 18.4–32.6 percentage points difference in institutional ownership is absurd given the total amount of passive assets benchmarked to the two Russell indexes (Table 3). For example, in 2006, $151.9 billion were passively tracking the Russell 1000 (about 1.1 percent of the index's total market cap) and $43.0 billion were passively tracking the Russell 2000 (about 2.9 percent of the index's total market cap). These estimates suggest an institutional ownership difference of 1.8 to 2.9 percentage points. In fact, the primary reason behind the observed difference in institutional ownership is Russell's float-adjustments: Firms at the bottom of the Russell 1000 have an up to 80 percentage points lower free float than firms at the top of the Russell 2000 (bottom left graph of Figure 2). As shown previously, there are no discontinuities in institutional ownership and free float with unadjusted May ranks (top right graph of Figure 2).

The problems of this approach also apply to studies that use the sharp RD based on float-adjusted June rankings as the first-stage regression in an instrumental variable estimation (Crane *et al.*, 2016; Bird and Karolyi,

[8]Boone and White (2015) use a slightly modified version of the sharp RD to estimate the effect of institutional ownership on firm transparency and information production. Specifically, they implement the sharp RD by running a local polynomial regression (instead of a local linear regression). This approach has the advantage that it does not assume linearity in the ranking variable. In unreported tests, I find similar results no matter whether the sharp RD is implemented by a local linear or a local polynomial regression.

2016; Bird and Karolyi, 2017; Chen *et al.*, 2017; Chen *et al.*, 2018). These studies use variants of the following first-stage regression for total institutional ownership or quasi-index investors: $\text{IO}_{i,t} = \alpha + \tau \text{R2000}_{i,t} + \sum_n \delta_n (\text{Rank}_{i,t}^{Jun})^n + \sum_n \gamma_n \text{R2000}_{i,t} (\text{Rank}_{i,t}^{Jun})^n + \nu_t + \epsilon_{i,t}$. Table IA1 of the Internet Appendix shows that this first-stage regression produces equally large differences in institutional ownership (12.0–26.5 percentage points) around the threshold as the sharp RD.

4.2 *Fuzzy Regression Discontinuity Based on CRSP/Compustat May Rankings*

An unbiased alternative to the previous approach would be a sharp RD that uses Russell's end-of-May ranks, which are effected neither by stock price changes in June nor by free-float adjustments. Unfortunately, Russell keeps their end-of-May ranks proprietary.

Chang *et al.* (2015) and Wei and Young (2018) thus use estimated CRSP/Compustat end-of-May ranks in a "fuzzy" RD design. In a fuzzy RD design, the probability of treatment increases when the observation crosses the threshold, but it does not jump from 0 to 1. Formally, the fuzzy RD design is specified by

$$\begin{aligned} \text{R2000}_{i,t} = \alpha_0 + \tau_0 \text{PredictR2000}_{i,t} + \delta_0 \text{Rank}_{i,t}^{May} \\ + \gamma_0 \text{PredictR2000}_{i,t} \text{Rank}_{i,t}^{May} + \nu_t + u_{i,t} \end{aligned} \tag{2a}$$

$$\begin{aligned} Y_{i,t} = \alpha_1 + \tau_1 \widehat{\text{R2000}_{i,t}} + \delta_1 \text{Rank}_{i,t}^{May} \\ + \gamma_1 \widehat{\text{R2000}_{i,t}} \text{Rank}_{i,t}^{May} + \nu_t + \epsilon_{i,t}, \end{aligned} \tag{2b}$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 after the annual index reconstitution in June of year t, $\text{Rank}_{i,t}^{May}$ is the end-of-May rank of firm i at year t, $\text{PredictR2000}_{i,t}$ is a dummy indicating whether $\text{Rank}_{i,t}^{May}$ predicts Russell 2000 membership, ν_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t}$ are the error terms. Variable $\text{Rank}_{i,t}^{May}$ is centered around the cutoff.

The researcher can estimate the end-of-May ranks based on firms' publicly available end-of-May market capitalizations. The fuzzy RD design takes into account that the researcher is not able to perfectly estimate Russell's May ranks (Wei and Young, 2018). Misclassifications happen especially near the threshold. For example, a firm with an estimated May rank of 1030 should be assigned to the Russell 2000. If, however, Russell indeed ranks the firm at place 970, then Russell will assign the firm to the Russell 1000 during the index reconstitution. Small measurement errors in the ranks will therefore result in misclassifications.

To recreate Russell's end-of-May market capitalizations, I multiply stock prices from CRSP by outstanding shares from Compustat.[9] Two analyses indicate that combining data from CRSP and Compustat provides a better estimate of Russell's true May ranks than using data only from CRSP. First, I find that the CRSP/Compustat May ranks result in 143 misclassifications in my sample, whereas the CRSP May ranks result in 732 misclassifications. Second, I estimate the first-stage regression of the fuzzy RD approach for both ways of constructing the end-of-May ranks. Table IA2 of the Internet Appendix shows that the CRSP/Compustat May ranks are a better predictor of actual index assignment than the CRSP May ranks.

Table 5 presents the results of the fuzzy RD design. Panel A reports the estimates of the first-stage regressions. It shows that the estimated end-of-May ranks have a probability of treatment, revealed by coefficient τ_0, of between 72% and 90%. As these values are close to 100%, they indicate that the procedure misclassifies only a few firms as a member of the Russell 2000 when the firms actually belong to the Russell 1000. Large t-statistics on coefficient τ_0 and huge F-statistics alleviate concerns about weak instruments. In fact, the instrumental variables explain almost all of the variation in variable $\text{R2000}_{i,t}$ (as indicated by an adjusted R^2 of between 83 and 94 percent).

Panel B reports the estimates of the second-stage regressions. It shows that the difference in ownership of quasi-indexers, indicated by coefficient τ_1, is insignificant in most regressions. Firms at the top of the Russell 2000 thus are not associated with a significantly higher institutional ownership than firms at the bottom of the Russell 1000. However, firms at the top of the Russell 2000 have a statistically significant 0.5–0.6 percentage points higher ownership of passive mutual funds.

---

[9] I take outstanding shares from Compustat because Compustat provides the number of outstanding shares on a firm-level. CRSP, by contrast, provides the number of outstanding shares on a stock-class level. Some firms have stock classes that are not available to the public and therefore are not reported in CRSP. Using outstanding shares from Compustat likely resembles Russell's approach because Russell states that it takes most of its firm information from annual or quarterly reports. Appendix A contains a detailed description of how I calculate the firms' end-of-May ranks.

An often-claimed argument against this approach is that the estimates are biased because of misclassified firms (e. g. Appel *et al.*, 2016; Crane *et al.*, 2016). Misclassifications are argued to result in a severe bias because they happen especially near the threshold, where precise estimates are most important. However, two reasons explain why this argument does likely not hold. First, as explained previously, misclassifications do not happen often. Second, if misclassifications were a severe problem, then we should find only small differences in index weights near the threshold. The reason for this is that misclassifications do not only affect the difference in institutional ownership, but also the variable that drives this difference, the difference in index weights. Panel B of Table 5, however, shows that firms at the top of the Russell 2000 have significantly higher index weights than firms at the bottom of the Russell 1000. The difference in index weights is large as it is slightly lower than one standard deviation of the sample firms' index weights.

4.3 Instrumental Variable Approach by Appel, Gormley, and Keim (2016)

Appel *et al.* (2016) and others (Hou and Zhang, 2017; Rubio and Vazquez, 2018; Appel *et al.*, 2019) exploit the discontinuity at the Russell indexes by running an instrumental variable (IV) approach that does not control for any firm ranks. Formally, it is specified by

$$IO_{i,t} = \alpha_0 + \tau_0 \text{R2000}_{i,t} + \sum_n l_n (\text{Mktcap}_{i,t})^n + \rho_0 \text{Float}_{i,t} + \nu_t + u_{i,t} \quad (3a)$$

$$Y_{i,t} = \alpha_1 + \tau_1 \widehat{IO_{i,t}} + \sum_n \lambda_n (\text{Mktcap}_{i,t})^n + \rho_1 \text{Float}_{i,t} + \nu_t + \epsilon_{i,t}, \quad (3b)$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 index after the index reconstitution in year t, $\text{Mktcap}_{i,t}$ is the logarithm of the market capitalization of firm i measured at the end of May of year t, $\text{Float}_{i,t}$ is the logarithm of the float-adjusted market capitalization measured at the end of June of year t, ν_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t}$ are the error terms. Appel *et al.* (2016) take the end-of-May market capitalizations from CRSP and obtain the float-adjusted market capitalizations from Russell. The regressions are estimated only on those observations that lie close to the threshold. Specifically, Appel *et al.* (2016) use a bandwidth of ±250 observations around the index threshold. The authors select the firms within the bandwidth based on Russell's float-adjusted June ranks.

The IV approach by Appel *et al.* (2016) was invented to overcome the problems of the sharp and fuzzy RD designs. In particular, the authors put forward two advantages of their approach. First, they argue that the IV approach is more robust against Russell's free-float adjustments than the sharp RD design because the IV approach does not rely on the end-of-June rankings to control for the distance between the observations and the threshold. Second, the authors claim that the IV approach is more robust against misclassifications than the fuzzy RD design because the IV approach does not rely on the unobservable end-of-May ranks. Instead, the IV approach includes different polynomials of the May market caps to account for differences in the firms' market capitalizations.

I first estimate the IV approach exactly as described in Appel *et al.* (2016). Panel A of Table 6 presents the results of the first-stage regressions. The panel shows the difference in ownership of quasi-index investors and passive funds between the firms that lie close to either side of the threshold. Columns 1 to 4 indicate that firms at the top of the Russell 2000 have a statistically significant 1.1–2.7 percentage points higher ownership of quasi-index investors than firms at the bottom of the Russell 1000. These estimates are slightly lower than those presented by Appel *et al.* (2016). Columns 5 to 8 show that much of the difference in ownership of quasi-index investors stems from passive mutual funds, which are a subset of quasi-indexers.

I recommend two modifications to the IV approach by Appel *et al.* (2016). The first is to use unadjusted May ranks (instead of float-adjusted June ranks) to identify the firms that lie within the bandwidth around the threshold. Using the float-adjusted June ranks will introduce a selection bias because Russell makes many free-float adjustments to the June ranks. Russell 1000 members with low free-float market capitalizations will gather into the bottom of the Russell 1000, whereas Russell 2000 members with large free-float market capitalizations will be assigned to the top of the Russell 2000. In fact, even with a large June bandwidth of ±250 firms, there is still a 23 percentage points difference in relative free float between the firms on either side of the threshold.

Appel *et al.* (2016) use the float-adjusted June ranks in their main specification (and the unadjusted May ranks in robustness tests) to select the sample around the threshold because, according to them, a disadvantage of using the May ranks is that one is no longer comparing the very bottom firms of the Russell 1000 against the very top firms of the Russell 2000, which is where one would expect to find the biggest difference in passive ownership. Contrary to their preferred specification, I recommend to select the bandwidth based on unadjusted May ranks to avoid a non-trivial selection bias in the IV approach.

The second modification to the IV approach by Appel *et al.* (2016) is to use better estimates of Russell's end-of-May market capitalizations. Appel *et al.* (2016) use end-of-May market capitalizations from CRSP to calculate the control variable $\text{Mktcap}_{i,t}$. Instead of using data only from CRSP, I recommend to calculate the May market caps based on data from CRSP and Compustat (see Appendix A). The previous subsection has shown that this generates better estimates of Russell's true May market caps.

This modification is important because the IV approach by Appel *et al.* (2016) uses actual index assignment as an instrumental variable in the first-stage regression. Wei and Young (2018) point out that actual index assignment is a deterministic function of the true end-of-May rankings (i. e. the May rankings that Russell uses during the annual index reconstitution). This implies that actual index assignment is only uncorrelated with the error term *conditional* on Russell's unadjusted end-of-May rankings, which are kept proprietary. Consequently, the approach by Appel *et al.* (2016) requires precise estimates of Russell's true end-of-May market capitalizations to avoid a possible bias.

Panel B of Table 6 presents the results of a modified IV approach. First, I rely on the end-of-May ranks (instead of the end-of-June ranks) to choose the firms inside the bandwidth. Second, I calculate the control variable $\text{Mktcap}_{i,t}$ by multiplying stock prices from CRSP by outstanding shares from Compustat. Columns 1 to 4 present the first-stage regression for ownership of quasi-index investors. The results indicate that, with these two modifications, the difference in ownership of quasi-index investors shrinks from values between 1.1 and 2.7 percentage points to values between 0.7 and 1.9 percentage points.[10] The difference in ownership of quasi-indexers remains statistically significant in two of the four regressions. As indicated by columns 5 to 8, the difference in passive mutual funds is also lower, with a value of 0.5 percentage points, but it remains statistically significant at the 1% level.

[10]Table IA3 of the Internet Appendix shows that the first modification (using unadjusted May ranks to select the bandwidth) leads to the lower coefficients.

4.4 Instrumental Variable Approach With Index Switchers

Schmidt and Fahlenbrach (2017) and Ben-David *et al.* (2018) exploit the discontinuity in institutional ownership by running an IV approach that uses dummies for the firms that switch between the Russell indexes. This subsection presents a modified version of the IV approach by Schmidt and Fahlenbrach (2017). It is specified by

$$\begin{aligned}\Delta IO_{i,t} &= \alpha_0 + \tau \text{R1000toR2000}_{i,t} + \gamma \text{R2000toR1000}_{i,t} \\ &\quad + \kappa_0(\text{Rank}_{i,t}^{May} - \text{Rank}_{i,t-1}^{May}) + \beta_0 \Delta X_{i,t} + \nu_t + u_{i,t}\end{aligned} \tag{4a}$$

$$\begin{aligned}\Delta Y_{i,t} &= \alpha_1 + \delta \widehat{\Delta IO_{i,t}} + \kappa_1(\text{Rank}_{i,t}^{May} - \text{Rank}_{i,t-1}^{May}) \\ &\quad + \beta_1 \Delta X_{i,t} + \nu_t + \epsilon_{i,t},\end{aligned} \tag{4b}$$

where R1000toR2000 is a dummy indicating whether firm i switches from the Russell 1000 to the Russell 2000 during the annual index reconstitution of year t, R2000toR1000 indicates switches from the Russell 2000 to the Russell 1000, $\text{Rank}_{i,t}^{May}$ are the end-of-May ranks of firm i at year t, $\Delta X_{i,t}$ is a vector of control variables measured in changes, ν_t are year dummies, and $\epsilon_{i,t}$ and $u_{i,t}$ are the error terms. Variable $\Delta IO_{i,t}$ measures the change in institutional ownership of firm i from the fourth quarter of year $t-1$ to the third quarter of year t.[11]

The basic idea of this approach is to investigate how much institutional ownership changes when firms actually move between the Russell indexes during the annual index reconstitution. Institutional ownership should increase when firms move from the Russell 1000 to the Russell 2000 and it should decrease when firms move from the Russell 2000 to the Russell 1000. As firms do only switch between the indexes when their market capitalizations change, it is important to control for these changes. The IV approach by Schmidt and Fahlenbrach (2017) accounts for this by controlling for the difference between the May ranks of year t and $t-1$, formally specified by the term $(\text{Rank}_{i,t}^{May} - \text{Rank}_{i,t-1}^{May})$.

Table 7 reports how much institutional ownership changes when firms switch between the Russell indexes during the annual index reconstitution.

[11]Equation 4 is not exactly the same specification as in Schmidt and Fahlenbrach (2017). The IV approach by Schmidt and Fahlenbrach (2017) includes the term $(\text{Rank}_{i,t}^{May} - \text{Rank}_{i,t-1}^{May})$ only in its first-stage regression, but *not* in its second-stage regression. Appel *et al.* (2018) point out that this is likely inappropriate because changes in market capitalizations are no valid instrument of institutional ownership. In line with this argument, I recommend to control for changes in May ranks in both stages.

Column 5, which uses no bandwidth, shows that ownership of quasi-index investors increases by 1 percentage point when firms move from the Russell 1000 to the Russell 2000. The column also shows that ownership of quasi-index investors decreases by 2.6 percentage points when firms move from the Russell 2000 to the Russell 1000. Column 10 shows that about one third of the change in ownership of quasi-index investors stems from changes in ownership of passive mutual funds.

Switches between the Russell indexes are only exogenous conditional on changes in market capitalizations. This assumption may not be satisfied with only one linear control variable for the change in May ranks. For robustness, I therefore also estimate regressions that include only the firms that stay close to the threshold throughout the year.[12] Table 7 shows lower estimates when dropping observations that lie far away from the threshold (columns 1 to 4). In fact, after introducing the bandwidth, firms that switch from the Russell 1000 to the Russell 2000 show a 0.7–1.5 percentage points increase in ownership of quasi-indexers, whereas firms that switch from the Russell 2000 to the Russell 1000 are associated with a 0.6–1.5 percentage points decrease in ownership of quasi-index investors.

5 Discussion of the Empirical Approaches and Simulation Results

Out of the four empirical approaches described in this paper, which approaches are appropriate to exploit the discontinuity at the Russell index threshold? This section evaluates the approaches and gives practical recommendations.

An appropriate approach should use (precise) estimates of Russell's unadjusted May rankings to control for the distance between the observations and the index threshold. Russell's float-adjusted June rankings cannot be used because they introduce a severe selection bias (see Subsection 2.3; Wei and Young, 2018; Appel *et al.*, 2018).

Three approaches were designed specifically for using estimated May rankings. The first approach is the fuzzy RD design, which uses predicted index membership as an instrumental variable for actual index assignment. The second approach is the IV approach by Appel *et al.* (2016),

[12] Appel *et al.* (2018) propose a different robustness check, namely to use different polynomial forms for the change in the end-of-May rankings.

which uses actual index assignment as an instrument for institutional ownership. And the third approach is the IV approach based on index switchers, which uses actual index switching as an instrument for the change in institutional ownership.

To investigate the robustness of the three approaches against false-positive results, I proceed with a simulation. Specifically, I generate a hypothetical dataset without a discontinuity in institutional ownership around the threshold. I model Russell's May market capitalizations, its free-float measure, institutional ownership, and changes between two index reconstitutions. I then add different levels of noise to Russell's true May market caps and investigate how the noise impacts the approaches. Appendix B explains the simulation in detail.

Table 8 shows that the three approaches perform well (i.e. they do not falsely reject the null hypothesis of no discontinuity in institutional ownership) when the estimated May market caps resemble Russell's May market caps. For example, the first row shows that the three approaches falsely reject (at a significance level of 10%) the null hypothesis of no discontinuity in 10 to 12 percent of the 100,000 runs. Adding more noise to the estimated May market caps, however, reveals that the IV approaches increasingly find significant discontinuities in institutional ownership around the threshold. For example, the third row shows that, with noisy May market caps, the IV approaches falsely reject the null hypothesis in 26 percent (IV approach by Appel *et al.* (2016)) and 34 percent (IV approach with index switchers) of the 100,000 runs. The fuzzy RD approach, by contrast, remains robust to noisy estimates of Russell's May rankings. Its first-stage regression shows weaker results, but it does not reject the null hypothesis of no discontinuity more often.

Based on these simulation results, I recommend to use the fuzzy RD approach to implement the Russell quasi-experiment. The fuzzy RD approach is the most robust approach against false-positive results because it uses predicted index membership (instead of actual index assignment or actual index switching) as an instrumental variable in its first-stage regression. Predicted index membership is a potentially exogenous instrument given that one can easily control for CRSP/Compustat end-of-May market capitalizations (Wei and Young, 2018).

Several papers claim that the fuzzy RD approach cannot be used because the observed CRSP/Compustat May ranks are a weak predictor of actual index assignment near the cutoff (e. g. Appel *et al.*, 2016; Crane *et al.*, 2016). My tests do not confirm this. Instead, my analyses show that predicted index assignment is a strong instrumental variable because it explains most of the variability in actual index assignment (see Subsection 4.2). I therefore conclude, in line with Wei and Young (2018), that the fuzzy RD approach is a well-suited methodology to implement the Russell quasi-experiment.

6 Example: Corporate Social Responsibility

This section investigates the effect of passive investors on CSR. Subsection 6.1 explains how I measure CSR. Subsection 6.2 provides the empirical results of the Russell quasi-experiment, and Subsection 6.3 discusses the results.

6.1 Description of the CSR Data Sample

The CSR ratings come from the KLD Database, which now belongs to MSCI ESG Research. KLD provides annual ratings for all firms in the Russell 1000 and Russell 2000 since 2003. KLD analysts assess each identified firm according to eight CSR dimensions: community, diversity, employment, environment, product, human rights, corporate governance, and controversial businesses. Within each dimension, a firm is rated based on several different binary criteria.[13]

In constructing CSR firm scores, I use all available KLD ratings that focus on environmental and social criteria (the first six dimensions). Following previous literature, I do not include corporate governance and controversial business ratings. I obtain a net KLD score for every firm by counting its strengths and deducting its concerns. In addition, I create a strengths-only and a concerns-only firm score. This procedure gives me three annual firm scores that measure the social responsibility of a firm in a given year.

I match the KLD firm scores to the sample consisting of Russell index constituents, institutional investors data, mutual funds data, accounting

[13] For example, in 2006, the company Caterpillar had one strength and two concerns in the environmental dimension. It was assigned with one strength because it had taken measurements to reduce its impact on climate change and air pollution. However, the firm also had two concerns because of regulatory problems and of manufacturing products that have a negative effect on climate change.

data, and stock returns. The full sample consists of 10,205 firm-year observations from 2003 to 2006, with 3,271 unique firms. The sample does not include observations after 2006 because of Russell's banding policy.

6.2 Empirical Results on CSR

This subsection investigates the effect of passive investors on CSR. I run the fuzzy RD based on CRSP/Compustat May ranks and a modified version of the IV approach by Appel *et al.* (2016). Both approaches do not use Russell's float-adjusted June rankings.

The fuzzy RD design is formally specified by equation 2. I estimate the approach exactly as described in Subsection 4.2. Variable $Y_{i,t}$ of the second-stage regression is the CSR score of firm i in year $t+1$. I measure the CSR score in year $t+1$ because an ownership change takes time to materialize in a change in CSR. Table 9 present the results. Columns 1 to 4 measure the difference in ownership of passive mutual funds between the firms that lie close to either side of the threshold. Firms at the top of the Russell 2000 index have a statistically significant 0.8–0.9 percentage points higher ownership of passive mutual funds than firms at the bottom of the Russell 1000. Columns 5 to 16 estimate the effect of passive mutual funds on CSR. The coefficients are insignificant, suggesting that passive funds do not significantly affect CSR.

I also estimate a modified version of the IV approach by Appel *et al.* (2016), formally specified by equation 3. As described in Subsection 4.3, I modify the approach in two ways. First, I use the unadjusted end-of-May ranks instead of the float-adjusted end-of-June ranks to select the firms around the threshold. Second, I compute the firms' May market caps by multiplying stock prices from CRSP by shares outstanding from Compustat (see Appendix A). Table 10 presents the results of this IV approach. Panel A shows that firms at the top of the Russell 2000 have a statistically significant 0.6–0.9 percentage point higher ownership of passive mutual funds than firms at the bottom of the Russell 1000. Panel B shows that the difference in passive ownership does not result in a CSR change.

Taken together, both approaches show insignificant coefficients. I therefore cannot reject the null hypothesis that passive mutual funds have no significant effect on CSR.

6.3 Discussion of the Results on CSR

Three previous papers also use the Russell 1000/2000 quasi-experiment to investigate the relationship between institutional investors and CSR, but find contradicting results.

Two of these studies rely on IV approaches to identify the effect of institutional investors on CSR. Rubio and Vazquez (2018) use the IV approach by Appel *et al.* (2016), while Chen *et al.* (2018) rely on an IV approach based on sharp RD. Both papers find that firms at the top of the Russell 2000 have 7–13 percentage points higher institutional ownership than firms at the bottom of the Russell 1000. Surprisingly, although both papers find similar differences in institutional ownership, they come to opposite conclusions: Rubio and Vazquez (2018) find a statistically significant *negative* effect of institutional investors on CSR, while Chen *et al.* (2018) find a significant *positive* effect.

Table IA4 of the Internet Appendix shows the replication of Rubio and Vazquez (2018). In contrast to their results, I find no significant effect on the net CSR score in the second-stage regression. The authors explained to me that replicating their results is difficult because they use proprietary data from Russell to construct the market capitalization control variables. They further explained that they use Russell's float-adjusted June ranks to choose the firms in the bandwidth. Table IA5 of the Internet Appendix replicates the findings by Chen *et al.* (2018). I find similar results as them when I use Russell's float-adjusted June market caps to construct the ranking controls. The results, however, become insignificant with ranking controls based on unadjusted May market caps. Given that both papers use Russell's float-adjusted June ranks, I conclude that both papers employ biased identification strategies.

The third study is by Hou and Zhang (2017). It investigates the effect of passive mutual funds on CSR. The authors use an IV approach that is similar to the one by Appel *et al.* (2016). Hou and Zhang (2017) find that firms at the top of the Russell 2000 have a 1 percentage point higher ownership of passive funds than firms at the bottom of the Russell 1000. The authors then show that this passive ownership difference leads to significantly less CSR strengths between 2003 and 2006, suggesting that passive funds have a negative effect on CSR.

Table IA6 of the Internet Appendix shows the replication of Hou and Zhang (2017). I find similar results as them when I use the original version of the IV approach by Appel *et al.* (2016). However, as described in Subsection 4.3, the original version of this IV approach can be improved in two ways. First, using unadjusted May ranks to select the firms that lie close to either side of the threshold avoids a possible selection bias. Second, using data from CRSP and Compustat to calculate the firm's end-of-May market caps provides a better estimate of Russell's May market caps. When I modify the approach accordingly in Panel B, I find insignificant coefficients in the second-stage regressions.

7 Conclusions

My paper discusses a popular quasi-experiment that uses the Russell index reconstitutions to identify the effects of institutional investors on firm outcomes. The basic idea of the quasi-experiment is to compare firms ranked at the top of the Russell 2000 index to firms ranked at the bottom of the Russell 1000 index. Previous studies claim that firms at the top of the Russell 2000 have significantly higher institutional ownership, but firms cannot control which index they are assigned to.

I evaluate four empirical approaches of the Russell quasi-experiment. My main finding is that an appropriate approach should use (precise) estimates of Russell's unadjusted end-of-May ranks to control for the distance between the observations and the index threshold (and also to select the bandwidth around the threshold). Russell's float-adjusted end-of-June ranks cannot be used because they introduce a severe selection bias. Simulations indicate that the fuzzy RD approach, which uses predicted index membership as an instrumental variable for actual index assignment, is most robust against false-positive results.

My empirical approaches find that between 1998 and 2006, firms at the top of the Russell 2000 had at most a 2 percentage points higher ownership of passive investors than firms at the bottom of the Russell 1000. There is no significant difference in total institutional ownership around the index threshold. Overall, these results contribute to the literature by providing evidence that the Russell quasi-experiment can be used to identify the effects of *passive investors* (primarily index funds) on firm outcomes, but not the effects of *overall institutional investors*.

This paper also studies the effect of passive investors on CSR, which is measured with social ratings from KLD. I use two different approaches of the Russell quasi-experiment to identify the effect of passive investors on CSR. Both approaches show insignificant coefficients, suggesting that passive investors have no effect on CSR. These results contradict previous findings that passive investors influence CSR.

Appendix A Estimating Russell's May Ranks

My paper uses the following procedure to estimate Russell's proprietary May ranks. The procedure has four steps:

1. I identify the index members of the Russell 1000 and 2000 in July. For these firms, I estimate the end-of-May market capitalization. Stock prices are obtained from CRSP and outstanding shares come from Compustat. I use the outstanding shares from the most recent quarterly report that is available to the public before the end of May. Compustat's variable "RDY" gives me the date on which a quarterly report is reported.[14]

2. I recalculate the market capitalizations for firms with dual share classes in CRSP. For these firms, I multiply stock prices from CRSP by outstanding shares from CRSP[15] for each stock class and aggregate the stock classes. I do not aggregate individual stock classes when they are individual members of the Russell indexes.

3. I rank the firms according to their end-of-May market caps. The largest firm gets the rank 1 and the smallest firm gets the rank 3,000.

4. I center the rank variable around the cutoff. Specifically, I deduct the number of firms that are a member of the Russell 1000 in July from the ranks estimated in the third step. As a result, firms in the Russell 1000 have negative ranks and firms in the Russell 2000 have positive ranks.

[14] Variable "RDY" is sometimes missing. Following Chang *et al.* (2015), I use the following rules to fill the variable. (1) For annual reports, I set RDY to 90 days after the fiscal year-end. There are some exceptions for this rule: Between 2003 and 2006, I set RDY to 75 days after the fiscal year-end if the firm has a market cap of larger than $75 million. Since 2007, I set RDY to 60 days for firms with a market cap of at least $700 million. (2) For quarterly reports, I set RDY to 45 days after the end of the quarter. An exception for this rule is: Since 2003, I set RDY to 40 days after the quarter-ends when a firm has a market cap of larger than $75 million.

[15] Compustat's outstanding shares cannot be used because Compustat provides the data on a firm level, whereas CRSP provides the data on a share-class level. It is not clear, which stock price from CRSP should be multiplied by the firm-level data from Compustat.

Appendix B Simulation

This paper uses the following procedure to generate a hypothetical Russell dataset. I choose the parameters in such a way that the produced variables resemble the observed Russell data as close as possible. The Internet Appendix shows the R code.

1. I draw 3,000 "May" market caps from a lognormal distribution with a meanlog of 7.0 and a sdlog of 1.4. I then rank these market caps and sort them into two Russell indexes.

2. I calculate 3,000 float-adjusted "June" market caps by multiplying the unadjusted market caps by $(1 - \text{adjustment factor})$. The adjustment factor follows a truncated exponential distribution with a rate of 3.5 and an upper bound of 1.

3. I generate institutional ownership as $IO_t = -0.16\log(\text{mcaps}_t) + 0.20\log(\text{float}_t) + \epsilon_t$.[16] The error ϵ_t is normal distributed with a mean of 0.35 and a standard deviation of 0.23. I set values lower than 0 to 0 and values higher than 1 to 1.

4. The IV approach with index switchers requires two additional steps:

   (a) I calculate the May market caps of period $t+1$ as $\text{mcaps}_{t+1} = \text{mcaps}_t \cdot c$, where c is a lognormal distributed adjustment factor with a meanlog of 0 and a standard deviation of 0.25.

   (b) I calculate institutional ownership of period $t+1$ as $IO_{t+1} = -0.16\log(\text{mcaps}_t \cdot c) + 0.20\log(\text{float}_t \cdot c) + 0.9\epsilon_t + 0.1\epsilon_{t+1}$. The error ϵ_{t+1} follows the same distribution as ϵ_t.

5. I calculate “noisy” May market caps by multiplying the market caps from the first step by an adjustment factor that follows a uniform, normal, triangular, or a (truncated) laplace distribution.

6. I estimate the following approaches:

   (a) The fuzzy RD approach specified by equation 2. Variable Rank is based on the noisy May market caps.

[16] This formula can be rewritten as $IO_t = 0.04\log(\text{mcaps}_t) + 0.20\log(\text{float}_t/\text{mcaps}_t) + \epsilon_t$. A firm thus has relatively lower institutional ownership when a smaller number of its shares are free-floating.

(b) The IV approach by Appel *et al.* (2016) specified by equation 3. Variable Mktcap is the noisy May market cap and variable Float is the float-adjusted June market cap from the second step. To select the bandwidth, I rank either the noisy May market caps or the float-adjusted June market caps.

(c) The IV approach with switchers specified by equation 4. Variable ΔIO is the difference between IO_{t+1} and IO_t and Rank is based on the noisy May market caps.

7. I repeat all steps 100,000 times and count how many times an approach shows a significant discontinuity (at the 10% level) in institutional ownership around the index threshold.

References

Appel, I. R., T. A. Gormley, and D. B. Keim. 2016. “Passive investors, not passive owners”. *Journal of Financial Economics*. 121(1): 111–141.

Appel, I. R., T. A. Gormley, and D. B. Keim. 2018. *Identification using Russell 1000/2000 index assignments: A discussion of methodologies*. Working paper.

Appel, I. R., T. A. Gormley, and D. B. Keim. 2019. “Standing on the shoulders of giants: The effect of passive investors on activism”. *Review of Financial Studies*. 32(7): 2720–2774.

Baghdadi, G. A., I. M. Bhatti, L. H. Nguyen, and E. J. Podolski. 2018. “Skill or effort? Institutional ownership and managerial efficiency”. *Journal of Banking & Finance*. 91: 19–33.

Barko, T., M. Cremers, and L. Renneboog. 2018. *Shareholder engagement on environmental, social, and governance performance*. Working paper.

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

Bird, A. and S. A. Karolyi. 2016. “Do institutional investors demand public disclosure?” *Review of Financial Studies*. 29(12): 3245–3277.

Bird, A. and S. A. Karolyi. 2017. “Governance and taxes: Evidence from regression discontinuity (retracted)”. *Accounting Review*. 92(1): 29–50.

Boone, A. L. and J. T. White. 2015. “The effect of institutional ownership on firm transparency and information production”. *Journal of Financial Economics*. 117(3): 508–533.

Bushee, B. J. 1998. “The influence of institutional investors on myopic R&D investment behavior”. *Accounting Review*. 73(3): 305–333.

Chang, Y.-C., H. Hong, and I. Liskovich. 2015. “Regression discontinuity and the price effects of stock market indexing”. *Review of Financial Studies*. 28(1): 212–246.

Chen, S., Y. Huang, N. Li, and T. Shevlin. 2019. “How does quasi-indexer ownership affect corporate tax-planning?” *Journal of Accounting and Economics*. 67(2–3): 278–296.

Chen, T., H. Dong, and C. Lin. 2017. *Institutional ownership and audit quality: Evidence from Russell index reconstitutions*. Working paper.

Chen, T., H. Dong, and C. Lin. 2018. “Institutional shareholders and corporate social responsibility: Evidence from two quasi-natural experiments”. *Journal of Financial Economics*. Forthcoming.

Coles, J. L., D. Heath, and M. C. Ringgenberg. 2018. *On index investing*. Working paper.

Crane, A. D., S. Michenaud, and J. P. Weston. 2016. "The effect of institutional ownership on payout policy: Evidence from index thresholds". *Review of Financial Studies*. 29(6): 1377–1408.

Dimson, E., O. Karakaş, and X. Li. 2015. "Active ownership". *Review of Financial Studies*. 28(12): 3225–3268.

Dyck, A., K. V. Lins, L. Roth, and H. F. Wagner. 2019. "Do institutional investors drive corporate social responsibility? International evidence". *Journal of Financial Economics*. 131(3): 693–714.

Fang, J. 2018a. *Quasi-indexer institutions and auditor independence: Evidence from non-audit service fees*. Working paper.

Fang, J. 2018b. *Quasi-indexer ownership and financial statements comparability*. Working paper.

Hou, W. and X. Zhang. 2017. *Do passive investors have passive corporate social responsibility?* Working paper.

Khan, M., S. Srinivasan, and L. Tan. 2017. "Institutional ownership and corporate tax avoidance: New evidence". *Accounting Review*. 92(2): 101–122.

Lee, D. S. and T. Lemieux. 2010. "Regression discontinuity designs in economics". *Journal of Economic Literature*. 48(2): 281–355.

Lin, Y., Y. Mao, and Z. Wang. 2018. "Institutional ownership, peer pressure and voluntary disclosures". *Accounting Review*. 93(4): 283–308.

Lu, R. 2013. *How does institutional ownership affect bank loan pricing: Evidence from a regression discontinuity design*. Working paper.

Monks, R. A. G. 1995. *Corporate governance*. Cambridge, Mass.: Blackwell Business.

Mullins, W. 2014. *The governance impact of index funds: Evidence from regression discontinuity*. Working paper.

Porter, M. E. 1992. *Capital choices: Changing the way America invests in industry*. Washington D.C.: Council on Competitiveness.

Rubio, S. and A. B. Vazquez. 2018. *Do institutional investors influence corporate social (ir)responsibility?* Working paper.

Schmidt, C. and R. Fahlenbrach. 2017. "Do exogenous changes in passive institutional ownership affect corporate governance and firm value?" *Journal of Financial Economics*. 124(2): 285–306.

Shleifer, A. and R. W. Vishny. 1986. "Large shareholders and corporate control". *Journal of Political Economy*. 94(3): 461–488.

Shleifer, A. and R. W. Vishny. 1990. "Equilibrium short horizons of investors and firms". *American Economic Review*. 80(2): 148–153.
Wei, W. and A. Young. 2018. *Selection bias or treatment effect? A re-examination of Russell 1000/2000 index reconstitution*. Working paper.
Wong, K. P. and L. Yi. 2017. *Institutional ownership and short-termist pressures*. Working paper.

Table 1: Literature overview of the Russell quasi-experiment

Panel A: Ownership of institutional investors

Study	Subject	Approach	Ranks	ΔIO	Time Period
Baghdadi *et al.* (JBF 2018)	managerial efficiency	IV based on sharp RD	May	12–13	1996–2006
Bird and Karolyi (RFS 2016)	public disclosure	IV based on sharp RD	June	8–11	1996–2006
Bird and Karolyi (AR 2017)	tax avoidance	IV based on sharp RD	June	8–11	1996–2006
Boone and White (JFE 2015)	firm information	sharp regression discontinuity	June	29–39	1996–2006
Chang *et al.* (RFS 2015)	index effects	fuzzy regression discontinuity	May	insignificant	1996–2012
Chen *et al.* (JFE 2018)	social responsibility	IV based on sharp RD	unknown	7–9	2003–2006
Chen *et al.* (WP 2017)	audit quality	IV based on sharp RD	unknown	7–10	2000–2006
Chen *et al.* (JAE 2019)	tax planning	sharp regression discontinuity	June	31	1996–2006
Crane *et al.* (RFS 2016)	payout policy	IV based on sharp RD	June	10	1991–2006
Fang (WP 2018a)	auditor independence	sharp regression discontinuity	June	23	2000–2006
Fang (WP 2018b)	financial statements	sharp regression discontinuity	June	25	1999–2006
Khan *et al.* (AR 2017)	tax avoidance	sharp regression discontinuity	June	26	1988–2006
Lin *et al.* (AR 2018)	voluntary disclosures	sharp regression discontinuity	June	large positive	1998–2006
Lu (WP 2013)	bank loan pricing	sharp regression discontinuity	unknown	35	1990–2006
Mullins (WP 2014)	corporate governance	fuzzy regression discontinuity	May	−10	2002–2006
Rubio and Vazquez (WP 2018)	social responsibility	IV by Appel *et al.* (2016)	none	12–13	2004–2006
Wei and Young (WP 2018)	institutional ownership	fuzzy regression discontinuity	May	insignificant	1996–2006
Wong and Yi (WP 2017)	short-term pressure	sharp regression discontinuity	June	28–31	1984–2006

Table 1: Literature overview of the Russell quasi-experiment (contd.)

Panel B: Ownership of passive mutual funds

Study	Subject	Approach	Ranks	ΔIO	Time Period
Appel *et al.* (JFE 2016)	corporate governance	IV by Appel *et al.* (2016)	none	1	1998–2006
Appel *et al.* (RFS 2019)	shareholder activism	IV by Appel *et al.* (2016)	none	4	2008–2014
Ben-David *et al.* (JF 2018)	stock volatility	IV with index switchers	none	positive	2000–2007
Coles *et al.* (WP 2018)	index investing	sharp regression discontinuity	May	positive	1993–2016
Hou and Zhang (WP 2017)	social responsibility	IV by Appel *et al.* (2016)	none	1	2003–2011
Schmidt and Fahlenbrach (JFE 2017)	corporate governance	IV with index switchers	none	positive	1993–2010

Description: This table shows all studies that use the Russell quasi-experiment to identify the effects of institutional investors on firm outcomes. Column 1 defines the study. Column 2 shows the research topic of the study. The remaining columns describe how the studies model the quasi-experiment. Column 3 gives the main empirical approach of the Russell quasi-experiment. Column 4 shows whether the study uses May or June ranking controls. Column 5 shows the studies' estimates on the difference in institutional ownership between the firms that lie close to either side of the threshold. A number of 10, for example, indicates that firms at the top of the Russell 2000 have a 10 percentage points higher ownership of institutional investors than firms at the bottom of the Russell 1000. Column 6 indicates the time period of the study.

Interpretation: Some studies find a high difference in *total institutional ownership* around the threshold (Panel A), while others find only a moderate difference in ownership of *passive mutual funds* around the threshold (Panel B).

Table 2: Descriptive statistics

Panel A: Firms of the Russell 1000 index

Variable	Mean	Median	Std	Obs
Institutional Ownership	0.659	0.682	0.230	8,880
Quasi-Index Ownership	0.410	0.415	0.155	8,880
Passive Funds Ownership	0.024	0.017	0.020	8,880
Market Capitalization	12.048	3.935	29.165	8,880
Free Float Capitalization	11.254	3.446	28.526	8,880
Index Weights in %	0.101	0.031	0.252	8,880

Panel B: Firms of the Russell 2000 index

Variable	Mean	Median	Std	Obs
Institutional Ownership	0.539	0.545	0.278	17,736
Quasi-Index Ownership	0.335	0.324	0.185	17,736
Passive Funds Ownership	0.021	0.014	0.021	17,736
Market Capitalization	0.574	0.456	0.372	17,736
Free Float Capitalization	0.473	0.351	0.368	17,736
Index Weights in %	0.050	0.038	0.037	17,736

Description: This table presents descriptive statistics for the variables used in this study. For every variable, the mean, median, standard deviation, and the number of observations are presented. Panel A presents descriptive statistics for the members of the Russell 1000 index, and Panel B presents descriptive statistics for the Russell 2000 index. The sample period is from 1998 to 2006.

Table 3: Passive assets benchmarked to Russell indexes

	1996	1997	1998	1999	2000	2001	2002	2003
Russell 2000	11.6	7.6	11.0	13.6	18.9	21.5	26.9	24.6
Russell 1000	20.9	20.7	19.0	25.9	17.3	34.0	35.6	37.2
	2004	2005	2006	2007	2008	2009	2010	2011
Russell 2000	38.9	39.2	43.0	51.7	38.5	38.4	56.8	60.1
Russell 1000	84.9	93.3	151.9	175.8	144.8	104.4	137.1	125.8

Description: This table reports the dollar amount of passive assets, in billions, benchmarked to the Russell 1000 and Russell 2000 by year. The data come from Chang *et al.* (2015), who report data from an internal Russell survey of its clients.

Table 4: Sharp regression discontinuity (incorrectly) based on June ranks

Dependent	Independent	(1)	(2)	(3)	(4)
Total institutional ownership	R2000	0.326***	0.253***	0.211***	0.184***
	(T)	(7.96)	(8.84)	(9.45)	(9.88)
	Bandwidth	100	200	300	400
	Observations	1787	3567	5336	7110
		(5)	(6)	(7)	(8)
Ownership by quasi-index investors	R2000	0.260***	0.196***	0.159***	0.134***
	(T)	(11.67)	(11.93)	(11.85)	(11.57)
	Bandwidth	100	200	300	400
	Observations	1787	3567	5336	7110

Description: This table estimates a sharp regression discontinuity based on Russell's float-adjusted end-of-June ranks. Formally, the approach is specified by

$$Y_{i,t} = \alpha + \tau \text{R2000}_{i,t} + \delta \text{Rank}^{Jun}_{i,t} + \gamma \text{R2000}_{i,t}\text{Rank}^{Jun}_{i,t} + \nu_t + \epsilon_{i,t},$$

where $\text{R2000}_{i,t}$ is a dummy that is one if firm i is a member of the Russell 2000 in year t and zero if firm i belongs to the Russell 1000, $\text{Rank}^{Jun}_{i,t}$ is the float-adjusted end-of-June rank of firm i provided by Russell, ν_t are year dummies, and $\epsilon_{i,t}$ is the error term. The regressions are estimated only on those observations that lie within a bandwidth close to the threshold. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: The sharp regression discontinuity based on float-adjusted June rankings shows that firms at the top of the Russell 2000 have a 18.4 to 32.6 percentage points higher ownership of institutional investors than firms at the bottom of the Russell 1000.

Table 5: Fuzzy regression discontinuity based on CRSP/Compustat May ranks

Panel A: First-stage regressions

Dependent	Independent	(1)	(2)	(3)	(4)
R2000	PredictR2000	0.721***	0.834***	0.875***	0.901***
	(T)	(24.82)	(47.70)	(69.21)	(89.48)
	Bandwidth	100	200	300	400
	Observations	1794	3567	5341	7117
	F-Statistic	668.2	1625.2	3044.5	4612.1
	Adj. R^2	0.83	0.89	0.92	0.94

Panel B: Second-stage regressions

Dependent	Independent	(1)	(2)	(3)	(4)
Firms' index weights	$\widehat{\text{R2000}}$	0.130***	0.126***	0.123***	0.122***
	(T)	(31.26)	(49.49)	(60.63)	(69.50)
	Bandwidth	100	200	300	400
	Observations	1794	3567	5341	7117
		(5)	(6)	(7)	(8)
Ownership by quasi-index investors	$\widehat{\text{R2000}}$	0.012	0.005	0.008	0.018**
	(T)	(0.56)	(0.38)	(0.82)	(1.98)
	Bandwidth	100	200	300	400
	Observations	1794	3567	5341	7117
		(9)	(10)	(11)	(12)
Ownership by passive funds	$\widehat{\text{R2000}}$	0.006***	0.005***	0.005***	0.005***
	(T)	(2.84)	(3.93)	(4.48)	(5.27)
	Bandwidth	100	200	300	400
	Observations	1794	3567	5342	7118

Description: This table estimates a fuzzy regression discontinuity specified by

$$\text{R2000}_{i,t} = \alpha_0 + \tau_0 \text{PredictR2000}_{i,t} + \delta_0 \text{Rank}_{i,t}^{May} + \gamma_0 \text{PredictR2000}_{i,t} \text{Rank}_{i,t}^{May} + v_t + u_{i,t}$$

$$Y_{i,t} = \alpha_1 + \tau_1 \widehat{\text{R2000}_{i,t}} + \delta_1 \text{Rank}_{i,t}^{May} + \gamma_1 \widehat{\text{R2000}_{i,t}} \text{Rank}_{i,t}^{May} + v_t + \epsilon_{i,t},$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 after the annual index reconstitution in June of year t, $\text{Rank}_{i,t}$ is the end-of-May rank of firm i at year t, $\text{PredictR2000}_{i,t}$ is a dummy indicating whether $\text{Rank}_{i,t}$ predicts membership in the Russell 2000, v_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t}$ are the error terms. Variable $\text{Rank}_{i,t}$ is constructed according to the procedure described in Appendix A. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: The fuzzy regression discontinuity based on CRSP/Compustat May rankings shows that firms at the top of the Russell 2000 have a statistically significant 0.5 to 0.6 percentage points higher ownership of passive funds than firms at the bottom of the Russell 1000.

Table 6: Instrumental variable approach by Appel *et al.* (2016)

Panel A: Original IV approach by Appel *et al.* (2016)

Dependent:	Ownership by quasi-indexers				Ownership by passive funds			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R2000	0.011	0.017**	0.023***	0.027***	0.011***	0.007***	0.008***	0.008***
(T)	(0.93)	(2.16)	(3.16)	(3.73)	(9.48)	(10.02)	(11.71)	(12.00)
Bandwidth	100	200	300	400	100	200	300	400
Observations	1784	3563	5332	7105	1784	3563	5332	7106

Panel B: Modified IV approach (CRSP/Compustat mcaps, May bandwidth)

Dependent:	Ownership by quasi-indexers				Ownership by passive funds			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R2000	0.007	0.012	0.016*	0.019**	0.005***	0.005***	0.005***	0.005***
(T)	(0.50)	(1.12)	(1.81)	(2.37)	(3.90)	(4.99)	(5.71)	(5.76)
Bandwidth	100	200	300	400	100	200	300	400
Observations	1794	3567	5341	7117	1794	3567	5342	7118

Description: This table estimates an IV approach by Appel *et al.* (2016). The first-stage regression is specified by

$$IO_{i,t} = \alpha_0 + \tau_0 \text{R2000}_{i,t} + \sum_{n=1}^{3} l_n(\text{Mktcap}_{i,t})^n + \rho_0 \text{Float}_{i,t} + v_t + u_{i,t},$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 index at time t, $\text{Mktcap}_{i,t}$ is the logarithm of the end-of-May market capitalization of firm i in year t, $\text{Float}_{i,t}$ is the logarithm of the float-adjusted end-of-June market capitalization in year t, v_t are year dummies, and $u_{i,t}$ is the error term. Panels A present the original IV approach by Appel *et al.* (2016), which uses CRSP market cap control variables and bandwidths based on float-adjusted June ranks. Panel B present the modified IV approach, which uses CRSP/Compustat market cap control variables and a bandwidth based on unadjusted May ranks. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: The *modified* IV approach shows that firms at the top of the Russell 2000 have a 0.5 percentage points higher ownership of passive funds than firms at the bottom of the Russell 1000.

Table 7: Instrumental variable approach with index switchers

Dependent	Independent	(1)	(2)	(3)	(4)	(5)
Ownership by quasi-index investors	R2000toR1000	−0.006	−0.012**	−0.015***	−0.014***	−0.026***
	(T)	(−0.98)	(−2.19)	(−3.28)	(−3.24)	(−7.03)
	R1000toR2000	0.013**	0.015***	0.010*	0.007	0.010***
	(T)	(1.99)	(2.64)	(1.95)	(1.44)	(2.69)
	Bandwidth	200	300	400	500	no
	Observations	1844	3346	4966	6604	22790
		(6)	(7)	(8)	(9)	(10)
Ownership by passive funds	R2000toR1000	−0.005***	−0.005***	−0.005***	−0.005***	−0.006***
	(T)	(−6.20)	(−8.36)	(−10.11)	(−8.93)	(−14.54)
	R1000toR2000	0.004***	0.005***	0.004***	0.004***	0.005***
	(T)	(3.96)	(5.67)	(5.74)	(5.22)	(8.87)
	Bandwidth	200	300	400	500	no
	Observations	1844	3346	4966	6604	22800

Description: This table estimates an IV approach with index switchers. Formally, the first-stage regression of this approach is specified by

$$\Delta IO_{i,t} = \alpha_0 + \tau \text{R1000toR2000}_{i,t} + \gamma \text{R2000toR1000}_{i,t} + \kappa(\text{Rank}_{i,t}^{May} - \text{Rank}_{i,t-1}^{May}) + v_t + u_{i,t},$$

where R1000toR2000 is a dummy indicating whether firm i switches from the Russell 1000 index to the Russell 2000 index during the index annual reconstitution of year t, R2000toR1000 indicates switches from the Russell 2000 index to the Russell 1000 index, $\text{Rank}_{i,t}^{May}$ are the end-of-May ranks of firm i in year t, v_t are year dummies, and $u_{i,t}$ is the error term. Variable $\Delta IO_{i,t}$ measures the change in institutional ownership of firm i from the fourth quarter of year $t-1$ to the third quarter of year t. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: This approach shows that ownership of quasi-index investors decreases by 0.6–2.6 percentage points when firms switch from the Russell 2000 to the Russell 1000 during the annual index reconstitution. Ownership of quasi-index investors increases by 0.7–1.5 percentage points when firms move the opposite direction.

Table 8: Simulation results

		fuzzy RD		IV by Appel *et al.* (2016)		IV with indexers	
		1-stage	2-stage	1-stage	1-stage	1-stage	
	Coefficient:	PredictR2000	Fit(R2000)	R2000	R2000	R2000toR1000	R1000toR2000
	Bandwidth:	200		200 (May)	200 (June)	No bandwidth	
Normal	Average Coef	0.490	0.000	0.011	0.013	0.000	0.001
mean=1, sd=0.05	Average T	(10.04)	(−0.00)	(0.25)	(0.35)	(0.09)	(0.31)
	|T(Coef)|>=1.64	100%	10%	11%	12%	10%	12%
Normal	Average Coef	0.250	0.006	0.021	0.023	0.002	−0.001
mean=1, sd=0.09	Average T	(3.98)	(−0.00)	(0.61)	(0.64)	(0.64)	(−0.23)
	|T(Coef)|>=1.64	93%	6%	16%	17%	17%	11%
Normal	Average Coef	0.126	0.266	0.030	0.032	0.004	−0.003
mean=1, sd=0.13	Average T	(1.72)	(−0.01)	(1.00)	(0.96)	(1.22)	(−0.82)
	|T(Coef)|>=1.64	52%	2%	26%	26%	34%	21%
Normal	Average Coef	0.065	−0.164	0.038	0.039	0.006	−0.005
mean=1, sd=0.17	Average T	(0.81)	(−0.01)	(1.37)	(1.27)	(1.74)	(−1.32)
	|T(Coef)|>=1.64	26%	1%	39%	36%	54%	38%
Uniform	Average Coef	0.473	0.001	0.010	0.013	0.000	0.001
min=0.92, max=1.08	Average T	(9.78)	(0.00)	(0.21)	(0.33)	(0.05)	(0.36)
	|T(Coef)|>=1.64	100%	10%	11%	12%	11%	12%
Uniform	Average Coef	0.245	−0.043	0.018	0.019	0.001	0.000
min=0.87, max=1.13	Average T	(4.17)	(0.00)	(0.46)	(0.53)	(0.41)	(−0.01)
	|T(Coef)|>=1.64	93%	6%	14%	15%	13%	10%
Uniform	Average Coef	0.093	0.004	0.024	0.026	0.003	−0.002
min=0.82, max=1.18	Average T	(1.39)	(−0.01)	(0.74)	(0.75)	(0.84)	(−0.42)
	|T(Coef)|>=1.64	43%	1%	19%	20%	22%	13%
Uniform	Average Coef	0.019	−0.563	0.031	0.032	0.005	−0.003
min=0.77, max=1.23	Average T	(0.26)	(−0.01)	(1.06)	(0.99)	(1.24)	(−0.83)
	|T(Coef)|>=1.64	19%	0%	29%	26%	35%	21%

Table 8: Simulation results (contd.)

		fuzzy RD		IV by Appel *et al.* (2016)		IV with indexers	
		1-stage	2-stage	1-stage	1-stage	1-stage	
	Coefficient:	PredictR2000	Fit(R2000)	R2000	R2000	R2000toR1000	R1000toR2000
	Bandwidth:	200		200 (May)	200 (June)	No bandwidth	
Triangle	Average Coef	0.484	0.000	0.011	0.013	0.000	0.001
a=0.88, m=1, b=1.12	Average T	(9.93)	(0.00)	(0.24)	(0.34)	(0.08)	(0.33)
	|T(Coef)|>=1.64	100%	10%	11%	12%	11%	12%
Triangle	Average Coef	0.227	−0.126	0.021	0.022	0.002	−0.001
a=0.78, m=1, b=1.22	Average T	(3.59)	(0.00)	(0.61)	(0.63)	(0.63)	(−0.22)
	|T(Coef)|>=1.64	89%	5%	16%	17%	17%	11%
Triangle	Average Coef	0.107	−5.573	0.030	0.032	0.004	−0.003
a=0.68, m=1, b=1.32	Average T	(1.44)	(−0.01)	(1.02)	(0.97)	(1.22)	(−0.81)
	|T(Coef)|>=1.64	44%	1%	27%	26%	34%	21%
Laplace	Average Coef	0.517	0.000	0.013	0.015	0.001	0.001
loc=1, scale=0.04	Average T	(10.44)	(0.00)	(0.30)	(0.39)	(0.18)	(0.23)
lower bound=0.1	|T(Coef)|>=1.64	100%	10%	12%	13%	11%	11%
Laplace	Average Coef	0.284	0.007	0.027	0.028	0.004	−0.002
loc=1, scale=0.08	Average T	(4.29)	(−0.00)	(0.80)	(0.83)	(1.01)	(−0.60)
lower bound=0.1	|T(Coef)|>=1.64	96%	7%	21%	22%	27%	16%
Laplace	Average Coef	0.172	−0.297	0.036	0.038	0.007	−0.005
loc=1, scale=0.12	Average T	(2.27)	(−0.01)	(1.24)	(1.22)	(1.79)	(−1.38)
lower bound=0.1	|T(Coef)|>=1.64	69%	3%	35%	34%	56%	39%

Description: This table presents evidence from simulations. As further described in Appendix B, I first generate hypothetical Russell 1000/2000 indexes without a discontinuity in institutional ownership around the threshold. I then add different levels of noise to Russell's true May market caps (by multiplying the market caps by an adjustment factor shown in the first column) and estimate three empirical approaches: the fuzzy RD approach based on unadjusted May ranks (equation 2), the IV approach by Appel *et al.* (2016) with a bandwidth based either on unadjusted May ranks or on float-adjusted June ranks (equation 3), and the IV approach based on index switchers (equation 4).

Interpretation: The fuzzy regression discontinuity based on estimated May market caps is most robust against false-positive results.

Table 9: CSR and the fuzzy regression discontinuity approach

Dependent	Independent	(1)	(2)	(3)	(4)
Ownership by passive funds	$\widehat{\text{R2000}}$	0.008**	0.009***	0.008***	0.008***
	(T)	(2.30)	(3.67)	(4.02)	(4.80)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934
		(5)	(6)	(7)	(8)
Net CSR score (strengths minus concerns)	$\widehat{\text{R2000}}$	−0.062	−0.172	−0.212	−0.095
	(T)	(−0.16)	(−0.80)	(−1.22)	(−0.62)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934
		(9)	(10)	(11)	(12)
Strenghts-only CSR score	$\widehat{\text{R2000}}$	0.222	0.012	−0.134	−0.149
	(T)	(0.80)	(0.09)	(−1.23)	(−1.49)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934
		(13)	(14)	(15)	(16)
Concerns-only CSR score	$\widehat{\text{R2000}}$	0.284	0.184	0.078	−0.054
	(T)	(1.19)	(1.14)	(0.59)	(−0.47)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934

Description: This table estimates a fuzzy regression discontinuity specified by

$$\text{R2000}_{i,t} = \alpha_0 + \tau_0 \text{PredictR2000}_{i,t} + \delta_0 \text{Rank}_{i,t}^{May} + \gamma_0 \text{PredictR2000}_{i,t} \text{Rank}_{i,t}^{May} + \nu_t + u_{i,t}$$
$$Y_{i,t+1} = \alpha_1 + \tau_1 \widehat{\text{R2000}}_{i,t} + \delta_1 \text{Rank}_{i,t}^{May} + \gamma_1 \widehat{\text{R2000}}_{i,t} \text{Rank}_{i,t}^{May} + \nu_{t+1} + \epsilon_{i,t+1},$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 after the annual index reconstitution in June of year t, $\text{Rank}_{i,t}$ is the end-of-May rank of firm i at year t, $\text{PredictR2000}_{i,t}$ is a dummy indicating whether $\text{Rank}_{i,t}$ predicts membership in the Russell 2000, ν_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t+1}$ are the error terms. Variable $\text{Rank}_{i,t}$ is constructed according to the procedure described in Appendix A. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: Passive mutual funds have no significant effect on CSR.

Table 10: CSR and the instrumental variable approach by Appel *et al.* (2016)

Panel A: First-stage regressions

Dependent	Independent	(1)	(2)	(3)	(4)
Ownership by passive funds (IO)	R2000	0.006***	0.009***	0.009***	0.009***
	(T)	(2.60)	(4.47)	(5.23)	(5.93)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934

Panel B: Second-stage regressions

Dependent	Independent	(1)	(2)	(3)	(4)
Net CSR score (strengths minus concerns)	$\widehat{IO}$	−26.8	−16.7	−15.1	−9.1
	(T)	(−0.56)	(−0.83)	(−0.85)	(−0.57)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934
		(5)	(6)	(7)	(8)
Strenghts-only CSR score	$\widehat{IO}$	11.4	−2.7	−12.4	−12.9
	(T)	(0.34)	(−0.22)	(−1.07)	(−1.20)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934
		(9)	(10)	(11)	(12)
Concerns-only CSR score	$\widehat{IO}$	38.2	14.0	2.8	−3.8
	(T)	(1.24)	(0.94)	(0.22)	(−0.33)
	Bandwidth	100	200	300	400
	Observations	730	1458	2198	2934

Description: This table estimates an IV approach that is specified by

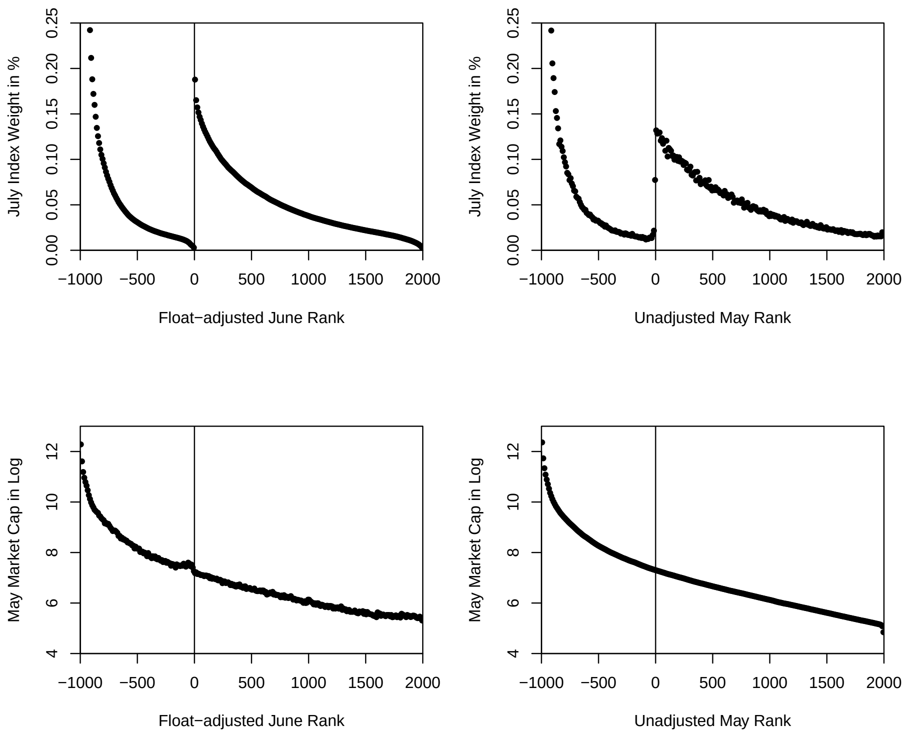
$$IO_{i,t} = \alpha_0 + \tau_0 \text{R2000}_{i,t} + \sum\nolimits_{n=1}^{3} l_n (\text{Mktcap}_{i,t})^n + \rho_0 \text{Float}_{i,t} + \nu_t + u_{i,t}$$

$$Y_{i,t+1} = \alpha_1 + \tau_1 \widehat{IO_{i,t}} + \sum\nolimits_{n=1}^{3} \lambda_n (\text{Mktcap}_{i,t})^n + \rho_1 \text{Float}_{i,t} + \nu_{t+1} + \epsilon_{i,t+1},$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 index at time t, $\text{Mktcap}_{i,t}$ is the logarithm of the end-of-May market capitalization of firm i in year t, $\text{Float}_{i,t}$ is the logarithm of the float-adjusted end-of-June market capitalization of firm i in year t, ν_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t+1}$ are the error terms. Variable $\text{Mktcap}_{i,t}$ is calculated by multiplying stock prices from CRSP by outstanding shares from Compustat (see Appendix A). Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: Passive mutual funds have no significant effect on CSR.

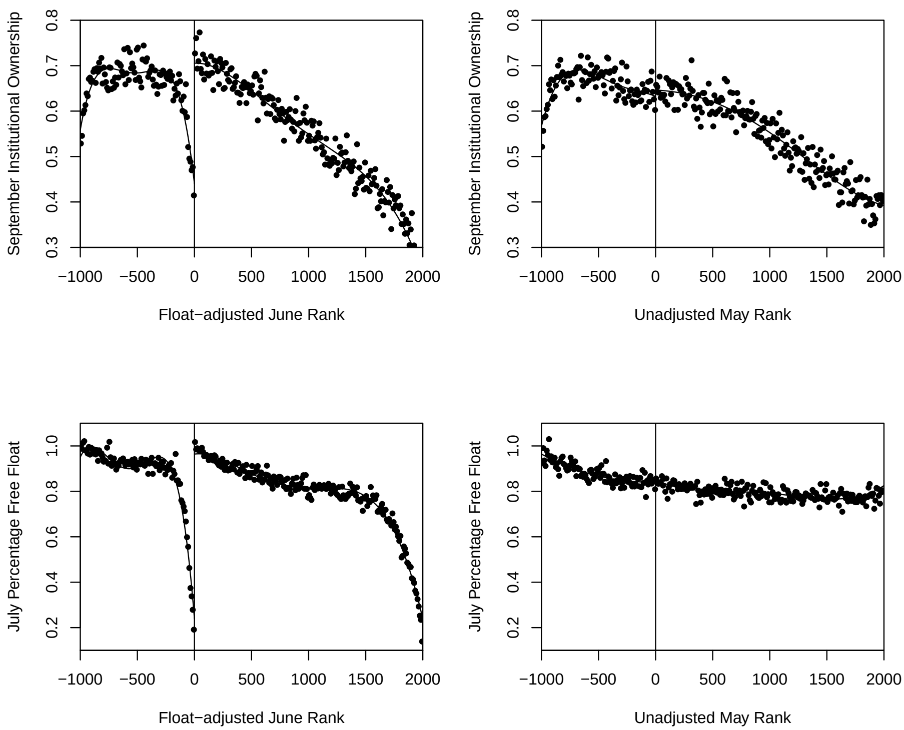
Figure 1: Index weights and market capitalizations around the threshold



Description: This figure presents the index weights and the end-of-May market capitalizations around the Russell index threshold. Ranks with a negative number indicate firms in the Russell 1000, whereas ranks with a positive number indicate firms in the Russell 2000. The top graphs show the firms' index weights (presented in percentages) and the bottom graphs show the logarithm of the firms' CRSP/Compustat end-of-May market capitalizations.

Interpretation: There is a discontinuity in index weights around the Russell index threshold, but no discontinuity in end-of-May market capitalizations.

Figure 2: Institutional ownership and free float around the threshold



Description: This figure presents institutional ownership and free float of the firms that lie close to either side of the Russell index threshold. Firms in the Russell 1000 have negative ranks, whereas firms in the Russell 2000 have positive ranks. The top graphs show the firms' total institutional ownership and the bottom graphs show the ratio of Russell's free-float market capitalization over total CRSP/Compustat market capitalization.

Interpretation: Float-adjusted June ranks show discontinuities in both institutional ownership and free float. Unadjusted May ranks show no discontinuities.