Online Appendix to "Selection Bias or Treatment Effect? A Re-Examination of Russell 1000/2000 Index Reconstitution"

Wei Wei   Alex Young

February 6, 2024

Contents

1 Data Availability

We have posted the code and replication datasets for the main paper and this internet appendix at `https://osf.io/gku6j`.

2 Index Construction Methodology

The annual reconstitution of the Russell Indices takes place in May and June of each year. On the last trading day in May, Russell ranks all eligible securities by their market capitalization. The largest 4000 become the Russell 3000E Index; all Russell U.S. equity indices are subsets of the Russell 3000E [FTSE Russell, 2017a, p. 12].

After index membership is determined, Russell adjusts shares outstanding to include only the "free float": shares that are available to the public for purchase. Examples of shares which are excluded from the free float (i.e. *not* available to the public) include

- Shares held by directors, senior executives, and managers of the company.
- Shares where the holder is subject to a lock-in clause.
- Shares subject to on-going contractual agreements that would ordinarily be treated as restricted [FTSE Russell, 2017b, p. 2].

Index weights are then calculated on the last Friday in June based on this float-adjusted market capitalization [FTSE Russell, 2017a, p. 25].

Until 2007, the largest 1000 by market capitalization were assigned to the Russell 1000, and the next 2000 were assigned to the Russell 2000. Starting in 2007, Russell Investments instituted a "banding policy" designed to reduce turnover between the indices:

> After the initial market capitalization breakpoints are determined by the ranges listed above, new members are assigned on the basis of the breakpoints, and existing members

are reviewed to determine if they fall within a cumulative 5% market cap range around these new market capitalization breakpoints. If an existing member's market cap falls within this cumulative 5% of the market capitalization breakpoint, it will remain in its current index rather than be moved to a different market capitalization-based Russell index [FTSE Russell, 2017a, p. 21].

For example, a firm that was in the Russell 1000 in the previous year could stay in the Russell 1000 if its ranking was from 1 to 1222; and a firm that was in the Russell 2000 in the previous year could stay in the Russell 2000 if its ranking was from 836 to 3000.

3 Interpreting Our Measure of The Float Adjustment

Suppose we have four firms with CRSP market capitalizations of 40, 30, 20, and 10, respectively. Our float adjustment measure is defined as

$$\text{Float adjustment}_{i,j,t} = \frac{\text{Russell June weight}_{i,j,t} - \text{CRSP weight}_{i,j,t}}{\text{CRSP weight}_{i,j,t}}$$

3.1 Equal proportional adjustments imply a float adjustment measure of zero

If half of every firm's shares are publicly unavailable, then the the float-adjusted market capitalizations will be 20, 15, 10, and 5, respectively. In this case, both the CRSP and float adjusted weights are 40%, 30%, 20%, and 10%, respectively. Hence, the float adjustment measure is zero for every firm even though all firms have 50% publicly unavailable shares.

As long as every firm has the same proportional adjustment, the weights before and after the adjustments remain the same, so the float adjustment measure, which is the percentage difference between the two sets of weights, is equal to zero. Zero does not mean no adjustments; it means no percentage difference in weights because all firms are adjusted the same way.

3.2 Larger adjustments imply a more negative value of the float adjustment measure

Now suppose the firm with 40 CRSP market capitalization has 75% of its shares as publicly unavailable, so the float adjusted market cap is only 10. All the other firms still have half of their shares as publicly unavailable. Hence, the float adjusted market capitalizations are 10, 15, 10, and 5. In this case, the float adjusted weights become 25%, 37.5%, 25%, and 12.5%.

As a result, the values of the float adjustment measure are -.375, .25, .25, and .25, respectively. The firm with 40 CRSP market capitalization was adjusted the most, so its shares were reduced proportionally more than all other firms. Since all the weights have to sum up to 1, it must be that its weight decreases while the weights of all other firms increase; indeed, we see that its weight gets reduced by 37.5% while other firms all get a 25% increase in weights after the adjustments.

3.3 Smaller adjustments imply a more positive value of the float adjustment measure

Now suppose the firm with 10 CRSP market capitalization has no publicly unavailable shares, so the float adjusted market capitalization is also 10. All the other firms still have half their shares as publicly unavailable. Hence, the float adjusted market capitalizations are 20, 15, 10, and 10. In this case, the float adjusted weights become 36.4%, 27.3%, 18.2%, and 18.2%. As a result, the values of the float adjustment measure are -.091, -.091, -.091, and 0.818 respectively. As we see, the firm with the 10 CRSP market capitalization was adjusted the least. Thus, its shares were reduced proportionally less than all other firms. Since all the weights have to sum up to 1, it must be that its weight increases while the weights of all other firms decrease. Indeed, we see that its weight gets increased by 81.8% while other firms all had a 9.1% decrease in weights after the adjustments.

Overall, each firm's change in market capitalization affects not only its own weights but also the weights of other firms. In the end, all the weights must sum to 1. Therefore, if one goes up in weight, others have to go down and vice versa.

4 Illustration of Selection Bias

The prior literature uses an indicator for Russell 2000 Index membership as an instrumental variable for institutional ownership, conditional on the Russell June rankings. A simplified model with no covariates can illustrate how to interpret the prior literature's results. Suppose that we restrict our attention to observations close to the Russell 1000/2000 threshold based on the Russell June rankings and consider the following:

$$
\begin{aligned}
\mathrm{IO}_{i,t} &= \alpha_0 + \alpha_1 \mathrm{R2000}_{i,t} + \xi_{i,t} \\
Y_{i,t} &= \beta_0 + \beta_1 \mathrm{IO}_{i,t} + \varepsilon_{i,t}
\end{aligned} \tag{1}
$$

where

- $Y_{i,t}$ is an outcome variable (e.g. payout, disclosure, etc.),
- $\mathrm{IO}_{i,t}$ is the September-dated institutional ownership of firm i in year t,
- $\mathrm{R2000}_{i,t}$ is an indicator variable for Russell 2000 Index membership (equal to 0 if the firm is in the Russell 1000) used as an instrumental variable for $\mathrm{IO}_{i,t}$.

Hahn et al. [2001, p. 203-204] prove that β_1 can be consistently estimated by taking the sample analogue of the following ratio:[1]

$$
\frac{\lim_{c \to 0^+} \mathbb{E}[Y_{i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[Y_{i,t} \,|\, \text{June Rank}_{i,t} = c]}{\lim_{c \to 0^+} \mathbb{E}[\mathrm{IO}_{i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[\mathrm{IO}_{i,t} \,|\, \text{June Rank}_{i,t} = c]} \tag{2}
$$

where $\text{June Rank}_{i,t}$ is the Russell June ranking centered at the 1000$^{\text{th}}$ rank. Note that while $\text{June Rank}_{i,t}$ did not appear as a variable in eq. (1), imposing a bandwidth based on the June

[1] See also Angrist and Pischke [2009, p. 262].

Rankings is equivalent to conditioning on it.

The key assumption for the consistency of the estimator is that the average potential outcome in the absence of treatment ($Y_{0i,t}$), is equal for firms arbitrarily close to the threshold [Hahn et al., 2001, p. 202]:[2]

$$\lim_{c \to 0^+} \mathbb{E}[\varepsilon_{i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[\varepsilon_{i,t} \,|\, \text{June Rank}_{i,t} = c] = 0$$

While it is not possible to test this or any assumption involving the error term [Roberts and Whited, 2013, p. 512], our placebo tests suggest that the assumption is untenable; firms around the threshold based on Russell June rankings are *ex-ante* different, not similar. Hence, the sample analogue of the ratio instead converges to

$$\beta_1 + \frac{\lim_{c \to 0^+} \mathbb{E}[\varepsilon_{i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[\varepsilon_{i,t} \,|\, \text{June Rank}_{i,t} = c]}{\lim_{c \to 0^+} \mathbb{E}[\text{IO}_{i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[\text{IO}_{i,t} \,|\, \text{June Rank}_{i,t} = c]}$$

The numerator of the term on the right can be seen as a form of selection bias: the average potential outcome in the absence of treatment is different between Russell 1000 and Russell 2000 firms at the threshold based on the Russell June rankings [Angrist and Pischke, 2009, p. 22].

While the sample analogue of eq. (2) does not consistently estimate β_1, it can still be interpreted as the difference in the average outcome at the threshold for Russell 2000 and Russell 1000 firms, scaled by the difference in the average institutional ownership of Russell 2000 and Russell 1000 firms at the threshold. Thus, prior literature has documented differences in outcomes between Russell 1000 and 2000 firms close to the Russell 1000/2000 cutoff, but these differences are only descriptive: they cannot be causally attributed to institutional investors' responding to Russell

[2] Let $\varepsilon_{i,t} \equiv Y_{0i,t} - \mathbb{E}[Y_{0i,t}]$ [Angrist and Pischke, 2009, p. 22]. By the law of iterated expectations,

$$\begin{aligned} &\lim_{c \to 0^+} \mathbb{E}[\varepsilon_{i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[\varepsilon_{i,t} \,|\, \text{June Rank}_{i,t} = c] \\ = &\lim_{c \to 0^+} \mathbb{E}[Y_{0i,t} \,|\, \text{June Rank}_{i,t} = c] - \lim_{c \to 0^-} \mathbb{E}[Y_{0i,t} \,|\, \text{June Rank}_{i,t} = c] \end{aligned}$$

1000/2000 Index reconstitution.

5 Response to Crane et al. (2016)

5.1 Proof that fuzzy RD estimator is consistent

5.1.1 Crane et al. model

In their internet appendix, Crane et al. [2016] present a model and claim to show that the fuzzy RD design in the Russell 1000/2000 setting does not consistently estimate the treatment effect. We first summarize their model and conclusion, and then we explain that their proof is incorrect because they did not complete the final step.

Let X_i denote market capitalization, and Y_i institutional ownership. Suppose

$$X_1 \leq c$$

$$X_0 > c$$

$$Y_1 - Y_0 = \tau > 0$$

where c is an arbitrary threshold, and τ is the treatment effect. Without loss of generality, let $i = 1$ be the treatment observation, and $i = 0$ the control observation.

X_i is unobservable. Instead, $\tilde{X}_i$ is observable, where

$$\tilde{X}_1 = X_1 + \varepsilon_1$$

$$\tilde{X}_0 = X_0 + \varepsilon_0$$

$$\varepsilon_i \sim \text{i.i.d}(0, \sigma_i) \text{ for } i = 1, 2$$

Let $P(\tilde{X}_1 - \tilde{X}_0 > 0) \equiv \alpha$. That is, although $X_1 < X_0$, we will observe the opposite, $\tilde{X}_1 - \tilde{X}_0 > 0$ with probability α. Define $Z_i = \mathbf{1} \cdot [\tilde{X}_1 - \tilde{X}_0 \leq 0]$. Then $P(Z_i = 1) = 1 - \alpha$ and $P(Z_i = 0) = \alpha$. We can assume $\alpha \neq 0$, otherwise the unobservability of X_i becomes trivial.

Assume $\tilde{X}_i$ can be treated as an exogenous variable such that we can consistently estimate the following model with OLS:

$$Y_i = \alpha + \rho Z_i + \varepsilon_i \tag{3}$$

It follows that

$$\begin{aligned}
\alpha &= \mathbb{E}[Y_i \,|\, Z_i = 0] \\
\alpha + \rho &= \mathbb{E}[Y_i \,|\, Z_i = 1] \\
&\Rightarrow \\
\rho &= \mathbb{E}[Y_i \,|\, Z_i = 1] - \mathbb{E}[Y_i \,|\, Z_i = 0] \\
&= (Y_1(1-\alpha) + Y_0\alpha) - (Y_1\alpha + Y_0(1-\alpha)) \\
&= Y_1(1-2\alpha) + Y_0(2\alpha - 1) \\
&= (Y_1 - Y_0)(1-2\alpha) \\
&= \tau(1-2\alpha)
\end{aligned}$$

Since $\alpha \neq 0$, ρ does not equal the treatment effect, τ. Crane et al. [2016, p. 9, internet appendix] thus conclude that a fuzzy RD design in the Russell 1000/2000 reconstitution setting fails to consistently estimate the treatment effect.

5.1.2 Our response

The mistake in Crane et al. [2016]'s conclusion is that in a fuzzy RD design, Z_i is not used as the treatment indicator but rather as the *instrumental variable* for the treatment indicator. That is, actual treatment status must be known and must be used for the **first stage**. Define actual treatment as

$$D_i = \begin{cases} 1 & \text{if } X_i \leq c \\ 0 & \text{if } X_i > c \end{cases}$$

where X_i is the unobservable market capitalization. Since X_i is unobservable, it is part of the error term. As $D_i = f(X_i)$, it follows that D_i is almost surely correlated with the error term [Roberts and Whited, 2013, p. 533]. That is, D_i is an **endogenous variable**.

But we can use Z_i - an indicator variable equal to 1 if we observe $\tilde{X}_1 < \tilde{X}_0$ - to instrument for D_i in the following first stage:

$$\begin{aligned} D_i &= \omega + \pi Z_i + \varepsilon_i \\ &\Rightarrow \\ \pi &= \mathbb{E}[D_i \mid Z_i = 1] - \mathbb{E}[D_i \mid Z_i = 0] \\ &= (1 - \alpha) - \alpha \\ &= 1 - 2\alpha \end{aligned}$$

Recall that Crane et al. [2016] ended their proof by noting that ρ does not equal τ (the treatment effect) because $\alpha \neq 0$:

$$\rho = \tau(1 - 2\alpha)$$

As long as $\alpha \neq \frac{1}{2}$, dividing ρ by π **recovers** the treatment effect, τ:

$$\begin{aligned} \frac{\rho}{\pi} &= \frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[D_i \mid Z_i = 1] - \mathbb{E}[D_i \mid Z_i = 0]} \\ &= \frac{\tau(1 - 2\alpha)}{(1 - 2\alpha)} \\ &= \tau \end{aligned}$$

Hence, the mistake in Crane et al. [2016]'s conclusion is that the numerator of this ratio *alone* **does not** implement the fuzzy RD design; rather, the numerator is the reduced form. The reduced form must be *divided by* the first stage to correctly implement the fuzzy RD design. This ratio is the population analog of a Wald estimator, which is the simplest instrumental variables estimator [Angrist and Pischke, 2009, p. 127]; and fuzzy RD is a special case of instrumental variables [Angrist and Pischke, 2009, p. 259].

$\alpha = \frac{1}{2}$ implies that Z_i is a weak instrument; that is, Z_i has no predictive power for D_i. We have shown in the paper that Z_i, predicted Russell 2000 Index membership, is a very strong instrument for D_i, actual Russell 2000 Index membership. Thus, the fuzzy RD estimator in the Russell 1000/2000 reconstitution setting can consistently estimate the treatment effect.

5.2 Strong fuzzy RD first stage does not require the full support

Crane et al. [2016, p. 10-11, internet appendix] state,

> ...given that the identification of the causal effect is at the threshold, F-tests of instrument strength are misspecified. Predicted market capitalization ranks may look like a very strong instrument over the full support because it predicts more than 98% of the actual index inclusions. But at the threshold, it is a weak instrument, which is where it really matters for the RD, because around the threshold observed market capitalization ranks have low predictive power.

However, Crane et al. [2016] provided no citation or proof that F-tests of instrument strength are misspecified in the fuzzy RD design. To the contrary, econometricians have emphasized the connection between fuzzy RD and instrumental variables [Hahn et al., 2001, p. 206], and have derived tests for weak instruments in fuzzy RD designs using F-statistics [Feir et al., 2016, p. 186].

Furthermore, we have shown in the paper (table 5) that researchers do not need to use "the full support" to find a significant first-stage discontinuity in the Russell 1000/2000 fuzzy RD design. A bandwidth of 100 with a linear specification, the same as in Crane et al. [2016, p. 1387], results in a 71.5 percentage point discontinuity at the cutoff and a Kleibergen and Paap [2006] F-statistic of 299.96.

5.3 Is a pre-existing discontinuity in market capitalization a problem? Yes.

Crane et al. [2016, p. 3, internet appendix] claim that a pre-existing discontinuity in market capitalization at the Russell 1000/2000 threshold is not necessarily a problem:

> Appel et al. (2015) argue that the regression specified in equation (2) of our paper may violate the assumption of local continuity/exclusion because it creates a discontinuity in market capitalization, which is related to the forcing variable, *Rank*. However, the local continuity assumption requires that the *potential outcomes* of firms around the threshold are the same, conditional on treatment. The discontinuity in observed market capitalization arises mechanically and obtains even if our identifying assumptions hold perfectly.

That is, Crane et al. make two claims:

1. There is a distinction between covariate balance (e.g. in market capitalization) and the local continuity assumption.
2. A discontinuity in market capitalization at the threshold will naturally occur as a result of the Crane et al. [2016] approach.

Regarding the first claim, Lee and Lemieux [2010, p. 291-292] explain that covariate balance is a *direct consequence* of the local continuity assumption; they are distinct but *related*.

To see this, we start with the definition of potential outcome. Assume that given a binary treatment, for each unit of observation i, there is

- an outcome Y_{1i} for what would occur if the unit were treated, and

- an outcome Y_{0i} for what would occur if the unit were not treated [Lee and Lemieux, 2010, p. 287].

Since we only ever observe Y_{1i} or Y_{0i} for unit i, but never both, these are "potential" outcomes.

Local continuity in the RD design context assumes that the functions

- $\mathbb{E}[Y_{1i} \,|\, X]$
- $\mathbb{E}[Y_{0i} \,|\, X]$

are continuous for all values of X, the assignment variable. (More technically, Hahn et al. [2001, p. 202] show that the minimum assumption is that $\mathbb{E}[Y_{0i} \,|\, X]$ is continuous at $X = c$, the RD threshold.) As a result of the local continuity assumption, we can use the average outcome of those right below the RD threshold (who are not treated) as a valid counterfactual for those right above the cutoff (who are treated) [Lee and Lemieux, 2010, p. 288-289]:

$$\underbrace{\lim_{X \to c^-} \mathbb{E}[Y_{0i} \,|\, X = c]}_{\text{Average outcome of the untreated right below the RD threshold}} = \underbrace{\lim_{X \to c^+} \mathbb{E}[Y_{0i} \,|\, X = c]}_{\text{What would have happened to the treated if they were not treated}}$$

That is, we are assuming that but for the treatment, units around the threshold would have been similar to each other. Lee and Lemieux [2010, p. 295] prove that covariate balance is a **consequence** of the local continuity assumption:[3]

> all observed and unobserved predetermined characteristics will have identical distributions on either side of $X = c$, in the limit, as we examine smaller and smaller neighborhoods of the threshold.

Phrased as a logical "if, then" statement, we have

> If the local continuity assumption holds, then we will observe covariate balance at the threshold in the data.

[3] See also Lee [2008, p. 679].

A statement is logically equivalent to its contrapositive:

> If we *do not* observe covariate balance at the threshold in the data, then the local continuity assumption *does not* hold.

Thus, we see that the presence of a pre-existing discontinuity in market capitalization at the threshold with Russell June rankings demonstrates a **violation** of the local continuity assumption.

Regarding the second claim, whether this pre-existing discontinuity arises mechanically as a result of Crane et al. [2016]'s methodology is irrelevant; the end result is that their design compares *ex-ante* different firms with each other around the threshold, which fails to identify a treatment effect.

6 Response to Appel et al. (2018, 2019)

During the review process, we revised our paper in response to successive revisions of Appel et al.:

- Appel et al. [2018] posted a revision of Appel et al. [2016a] to SSRN in October 2018.
- In November 2018, we received an invitation to revise and resubmit to *Critical Finance Review*.
- Based on reviewer comments (which referred to e.g., the content in Appel et al. [2018]), we revised our paper and resubmitted in February 2019.
- In May 2019, Appel et al. [2019] posted a revision of Appel et al. [2018] to SSRN, in which several points in Appel et al. [2018] were removed; and a new example was given.
- We received a conditional acceptance in July 2019, revised our paper (including e.g., the new content in Appel et al. [2019]), and resubmitted the final version in September 2019.
- In April 2020, Appel et al. [2020] posted a revision of Appel et al. [2019] to SSRN, in which the new example given in Appel et al. [2019] was removed.

Thus, in this section we respond to points made in earlier versions of Appel et al. "for pedagogical, historical, and scholarly reasons" [Dye, 2001, p. 182].

6.1 Why covariate imbalance invalidates using the Russell June rankings

To Summary

6.1.1 Treatment persistence does not imply that covariate imbalance is expected

In the first submission to *Critical Finance Review*, we initially wrote,

> One may ask whether the covariate balance test is inappropriate in the Russell 1000/2000 reconstitution setting because although reconstitution occurs annually, Russell 1000/2000 Index membership may be fairly stable over time. Under this line of thinking, covariate *imbalance* may be natural and expected and thus may not necessarily invalidate the research design choice of using the Russell June rankings.
>
> However, the RD design does not compare Russell 1000 and Russell 2000 firms overall. Instead, the focus is on the firms *around* the cutoff, and it is at the 1000$^{\text{th}}$ ranking cutoff that we would expect index membership to resemble a coin toss. The argument that covariate imbalance may be natural and expected implies the opposite, that a firm ranked 999 (1001) is *not likely* to be ranked 1001 (999), which would indeed invalidate the research design choice of using the Russell June rankings.

Appel et al. [2018, p. 19] responded,

> In our view, Wei and Young's use of covariate balance tests is flawed conceptually because it looks at outcomes argued to be affected by index assignment and the pre-reconstitution outcomes are not truly predetermined when index assignment is persistent (i.e., this year's index assignment is correlated with last year's index assignment). If there was a setting where the Russell indexes did not exist in year $t-1$, then a covariate

> imbalance before the creation of the indexes in year t would be compelling evidence against that methodology.

In section 3.3 of our paper, we addressed this response. In Appel et al. [2019], the preceding "In our view, Wei and Young's use of covariate balance tests is flawed because ..." Appel et al. [2018, p. 19] paragraph was removed entirely.

6.1.2 Use data away from the threshold to estimate effect at threshold

In the first submission to *Critical Finance Review*, we initially wrote,

> RD designs fundamentally rely upon extrapolation, such that data *away* from the threshold must be used to provide an estimate of the effect *at* the threshold:
>
> - **Angrist and Pischke [2009, p. 253]** - "...the validity of RD turns on our willingness to extrapolate across covariate values, at least in a neighborhood of the discontinuity."
> - **Lee and Lemieux [2010, p. 286]** - "...in order to produce a reasonable guess for the treated and untreated states at $X = c$ with finite data, one has no choice but to use data *away* from the discontinuity" (emphasis original).
> - **Roberts and Whited [2013, p. 542]** - "As such, widening the area of analysis around the threshold to mitigate power concerns is often necessary."

Appel et al. [2018, p. 19] responded,

> The authors are aware of this critique and claim on page 15 that the only thing that matters is that index assignment in year t is not predictive of index assignment in year $t - 1$ for observations sufficiently close to the threshold (e.g., stocks with end-of-May year t rankings of 999 and 1,001). But, it's not clear this is correct. As the authors accurately point out on page 25, "RD designs fundamentally rely upon extrapolation, such that data away from the threshold must be used to provide an estimate of an effect

> at the threshold". And, the further one moves from the threshold, the more and more predictive this year's ranking will be of last year's index assignment. For example, a stock with a ranking of 1,100 in year *t* is more likely to have been in the Russell 2000 last year than a stock with a ranking of 900.

However, their "the further one moves from the threshold, the more ..." response demonstrates a misunderstanding of

> RD designs fundamentally rely upon extrapolation, such that data *away* from the threshold must be used to provide an estimate of the effect *at* the threshold.

To provide an estimate of the effect *at* the threshold, we must use data *away* from the threshold, but we are still interested in the effect ***at* the threshold**. In Appel et al. [2019], the preceding "The authors are aware of this critique and claim ... But, it's not clear this is correct ..." Appel et al. [2018, p. 19] paragraph was removed entirely.

6.2 Why actual index membership is not conditionally exogenous

To Summary

In section 4.1 of the paper, we stated,

> Hence, in practice, actual index membership is an endogenous variable, is not conditionally exogenous, and thus cannot be a valid instrumental variable.

In the first submission to *Critical Finance Review*, this statement (and accompanying argument preceding the "hence") was originally abbreviated in footnote 15:

> Note that based on our discussion, actual index membership is an endogenous variable and therefore cannot be an instrumental variable for institutional ownership. If we could control for the true Russell May rankings, then actual membership could be conditionally exogenous, but since the true rankings are unobservable, we cannot control for them. Hence, in practice, actual index membership cannot be conditionally exogenous and thus cannot be an instrumental variable.

Appel et al. [2018, p. 20] responded,

> To see why we believe this statement is incorrect, consider the example of a RD estimation where the indicator for treatment is always an endogenous outcome of the forcing variable. This is exactly why RD relies on a robust set of controls for the forcing variable. The idea is that the endogenous treatment is *conditionally exogenous* after including the controls that are known to drive treatment. The same logic applies to IVs and fuzzy RD (which is a special type of IV estimation). In the fuzzy RD proposed by Wei and Young (2018) one can view the predicted index assignment as the endogenous outcome of end-of-May market rankings, and yet, fuzzy RD still uses this predicted index assignment as an instrument for actual index assignment in the first stage of the estimation. This is not problematic, however, because the fuzzy RD estimation controls for the variable driving predicted treatment, end-of-May market capitalization rankings. The same is true for AGK's use of actual index assignment as the IV: One can use actual index assignment as an instrument if one is able to include controls that sufficiently capture the underlying factors that determine treatment that might also pose problems for the exclusion assumption. The exclusion assumption does not depend on whether the unconditional IV is endogenous; it instead requires conditional exogeneity.

Indeed, in our original footnote 15, we ourselves discussed in the 2nd and 3rd sentences (the lack of) conditional exogeneity. In Appel et al. [2019], the preceding "To see why we believe this statement is incorrect . . . " Appel et al. [2018, p. 20] paragraph was removed; but Appel et al. [2019, p. 19] reiterated their disagreement,

> We disagree with footnote #15 of Wei and Young (2018) which claims that because actual index assignment is an endogenous outcome of Russell's end-of-May market cap rankings, it cannot be used as an instrument when Russell's end-of-May market cap rankings are unobserved. In other words, they seem to argue that it is impossible to isolate exogenous variation in firms' ownership structures using index assignment as an IV unless one can control for Russell's unobserved end-of-May market cap rankings.

and added a new example for illustration.

We demonstrate why their argument (that "it is *possible* to isolate exogenous variation in firms' ownership structures using index assignment as an IV *without controlling* for Russell's unobserved end-of-May market cap rankings") is incorrect with both theory and simulation.

6.2.1 Theory

Denote

- X as the assignment variable, the true end-of-May market capitalization rankings used by Russell;
- c, the threshold for Russell 2000 Index assignment; and
- $D = \mathbf{1} \cdot [X > c]$, a 0/1 indicator variable for Russell 1000/2000 Index membership.

Appel et al. [2018, p. 7] agree that the actual end-of-May market capitalization ranking used by Russell (X) should perfectly predict index assignment (D), therefore D is a deterministic function of X. As explained in Lee and Lemieux [2010, p. 289], when $X > c$, D is always equal to 1; and when $X \leq c$, D is always equal to 0. That is, conditional on X, D is a constant; and a constant has zero correlation with any other factor. Hence, if we could control for X, we *do not have to assume* conditional exogeneity: the conditional exogeneity of D is *trivially satisfied*.[4]

However, we cannot control for X because X is unobserved. Instead, we observe $Z = X + \xi$ (e.g. researcher-constructed May rankings). Since Z does not perfectly predict D, when $Z > c$, D **is not** always equal to 1; and when $Z \leq c$, D **is not** always equal to 0. That is, conditional on Z, there is still some variation in D; it is not guaranteed that D is uncorrelated with the regression error term. Therefore, without controlling for X, the conditional exogeneity of D has to be *assumed*.

[4]We still have to assume local continuity, but that is a separate issue.

Conditional exogeneity (also known as ignorability or unconfoundedness) implies that all relevant factors are controlled for, and no omitted variables are correlated with the treatment D [Lee and Lemieux, 2010, p. 289]. However, in this setting, firms with similar Z but different D **must have** values of X that are on opposite sides of the threshold c. The difference in X is what causes the difference in D, therefore X is a correlated omitted variable. Hence, as we stated, actual Russell 1000/2000 Index membership cannot be conditionally exogenous in practice.

6.2.2 Simulation

We can also demonstrate our conclusion through simulation (code available at `https://osf.io/gku6j`). For comparability, we follow the example introduced in Appel et al. [2019, p. 19] where a researcher is interested in estimating β_2, the effect of x_2 on y:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

with the following conditions:

- x_1 and x_2 are correlated: $\text{Cov}(x_1, x_2) \neq 0$.
- x_1 is exogenous: $\text{Cov}(x_1, u) = 0$.
- x_2 is endogenous: $\text{Cov}(x_2, u) \neq 0$.

To address the problem that x_2 is endogenous, suppose we have an instrumental variable z for x_2:

$$x_2 = \alpha_0 + \alpha_1 x_1 + \gamma z + v$$

with the following properties:

- Non-zero first stage: $\gamma \neq 0$.
- z is an indicator variable based on a threshold rule with x_3:

$$z = \begin{cases} 1 & \text{if } x_3 \leq 1000^{\text{th}} \text{ descending-order rank value of } x_3 \\ 0 & \text{if } x_3 > 1000^{\text{th}} \text{ descending-order rank value of } x_3 \end{cases}$$

- x_3 is equal to x_1 plus noise: $x_3 = f(x_1) + \varepsilon$

Adapted to this example,

- our paper's argument is that conditional on the true-but-unobserved end-of-May Russell ranking of market capitalization x_3, index assignment z is a constant, and therefore conditionally, $\text{Cov}(z, u) = 0$ is trivially satisfied.

- Appel et al. [2019, p. 20] argue that conditional on observed-but-noisy end-of-May market capitalization x_1, index assignment z can be a valid instrument for passive institutional ownership x_2 if the modified exclusion restriction $\text{Cov}(\varepsilon, u) = 0$ is **assumed**.

With 1,000 repetitions, our simulation shows that if

- $\text{Cov}(\varepsilon, u) \neq 0$,

- the true $\beta_2 = 0$,

- and x_3 were observable,

then

- the Appel et al. [2019] approach **fails to consistently estimate** $\beta_2 = 0$, the effect of x_2 on y. The $\widehat{\beta_2}$'s are economically and statistically different from zero in 100% of the simulations.

- using x_3 as the assignment variable for z in a "continuous endogenous regressor" regression discontinuity design [Lee and Lemieux, 2010, p. 301] **consistently estimates** $\beta_2 = 0$. The $\widehat{\beta_2}$'s are economically and statistically non-zero in approximately 5% of the simulations, which is in line with the $\alpha = 0.05$ size of the test.

We emphasize that the regression discontinuity design

- consistently estimated β_2
  - without controlling for x_1, although x_1 affects both x_2 and y:

$$y = \beta_0 + \boldsymbol{\beta}_1\mathbf{x_1} + \beta_2 x_2 + u \qquad (\beta_1 \neq 0)$$

$$x_2 = \alpha_0 + \boldsymbol{\alpha}_1\mathbf{x_1} + \gamma z + v \qquad (\alpha_1 \neq 0)$$

  - and instead controlling for $x_3 = f(x_1) + \varepsilon$, which **does not** directly affect either x_2 or y.
- and **is robust to** $\text{Cov}(\varepsilon, u) \neq 0$.

One may argue that our simulation is unfair because $\text{Cov}(\varepsilon, u) \neq 0$ implies a violation of the modified exclusion restriction required for the Appel et al. [2019] approach. The question becomes, what is the meaning of $\text{Cov}(\varepsilon, u) \neq 0$? It means

> ...there is something systematic about different end-of-May market caps used by Russell to construct its rankings that is associated with a factor that directly matters for the outcome of interest even after robustly controlling for size [Appel et al., 2019, p. 20].

Despite Appel et al. [2019] implying otherwise, there *is* something systematic at play here. Firms i and j with similar x_1 (size) but different z (index assignment) **must have** values of x_3 (Russell size) **on opposite sides** of the z threshold. Thus, firms i and j are systematically different on the sole dimension that determines index assignment, z. It is for this reason that Imbens and Lemieux [2008, p. 621] in general cautioned against approaches which condition on x_1 and assume that the remaining variation in $z = f(x_3) \equiv g(x_1) + \varepsilon$ is as good as random.

We believe that the Appel et al. [2019, p. 19]

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

example has a key advantage of being both applicable to the Russell reconstitution setting and relatively generalizable. Nevertheless, in Appel et al. [2020], the preceding $y = f(x_1, x_2)$ example [Appel et al., 2019, p. 19] was removed; while their disagreement with our statement that "…in practice, actual index membership is an endogenous variable, is not conditionally exogenous, and thus cannot be a valid instrumental variable" was further reiterated.

6.3 Graphs do not substitute for tables

To Summary

In the paper, we show

- in column 3 of table 5 and in the bottom left panel of figure 5 that there is a large and significant first-stage discontinuity *at* the threshold with the same bandwidth of 250 and third-order polynomial as in the main analysis of Appel et al. [2016b, p. 121]
- and in column 4 of table 5 and in the bottom right panel of figure 5 that there is a large and significant first-stage discontinuity *at* the threshold even with a much smaller bandwidth of 50 and a first-order (linear) polynomial.

Appel et al. [2018, p. 11] responded,

> …Wei and Young (2018) report compelling evidence of a strong first stage that is not sensitive to the polynomial order used to control for the forcing variable, which in their specification is the end-of- May market cap ranking as calculated using a combination of data from CRSP and Compustat. Examples of this are provided in the two left panels of Figure 5 in Wei and Young (2018). These findings suggest the Russell 1000/2000 setting is amenable to using a fuzzy RD in wider bandwidths that make use of observations further from the discontinuity. While the use of wider bandwidths increases the risk of bias that might occur from any failure to control adequately for the importance of the forcing variable, *Rank*, these concerns can be overcome with sufficiently robust

controls for *Rank*.

> However, at smaller bandwidths, the robustness of the fuzzy RD still remains questionable even using this alternative way to calculate end-of-May market cap rankings. The potential weakness of the first stage when using smaller bandwidths is seen in the right two panels of Figure 5 in Wei and Young (2018). In both cases (especially in the bottom right panel), one can see that the probability of being in the Russell 2000 is converging to 50-50 as you near the threshold. While their inclusion of linear trends on either side give the impression of a large discontinuity, the use of a second- or third-order polynomial (as done in the left two panels of their Figure 5) would likely attenuate this difference.

However, the large discontinuity in the bottom right panel of figure 5 is confirmed in the corresponding first-stage regression: Column 4 of table 5 uses the same bandwidth of 50 and same local linear specification; and column 4 reports a 54.5 percentage point discontinuity in the probability of Russell 2000 Index assignment *at* the cutoff, with an adjusted R^2 of 71.2% and Kleibergen and Paap [2006] F-statistic of 64.53.

As Roberts and Whited [2013, p. 541] remarked,

> Graphical analysis can be helpful but should not be relied upon. There is too much room for researchers to construct graphs in a manner that either conveys the presence of treatment effects when there are none, or masks the presence of treatment effects when they exist. Therefore, graphical analysis should be viewed as a tool to guide the formal estimation, rather than as a necessary or sufficient condition for the existence of a treatment effect.

In Appel et al. [2019], the preceding Appel et al. [2018, p. 11] claim that the bottom right panel of our figure 5 gives an "impression of a large discontinuity" was removed.

6.4 Summary

Thus, to summarize,

- Covariate imbalance – in March-dated total and quasi-indexer institutional ownership and end-of-May market capitalization etc. – invalidates using the Russell June rankings; more generally, imbalance – in relevant covariates or lagged outcomes measured pre-treatment – suggests that "something is wrong in the design" (see e.g., Lee and Lemieux [2010, p. 330]).

- In practice, actual Russell 1000/2000 Index membership is not conditionally exogenous; more generally, whenever a treatment variable is a deterministic function of an unobserved assignment variable relative to a cutoff, such a treatment variable is **not** conditionally exogenous after controlling for a noisy-but-observed measure of the assignment variable (see e.g., Imbens and Lemieux [2008, p. 621]).

- First-stage discontinuity graphs are complements to first-stage discontinuity regressions, **not** substitutes (see e.g., Roberts and Whited [2013, p. 541]).

7 Robustness

We assess the robustness of the results to changes in polynomial order and bandwidth.

7.1 Discontinuity in market capitalization with Russell June rankings

Table 1 holds the bandwidth constant at 200 and varies the polynomial order from first-, second-, and third-degree. The third-degree case in column 1 is also presented in the main paper. The discontinuity in CRSP and Compustat May market capitalization at the threshold ranges from $726 million to $1.028 billion, and the standard errors range from $92 million to $115 million.

Table 2 holds the bandwidth constant at 100 and varies the polynomial order from first-, second-, and third-degree. The discontinuity at the threshold ranges from $353 million to $1.023 billion, and the standard errors range from $93 million to $114 million. Thus, across all six specifications, there are extremely large, significant, pre-existing discontinuities in market capitalization at the threshold such that with the Russell June rankings we are comparing firms of very different sizes with each other. Hence, it is not credible to argue that the firms at the bottom of the Russell 1000 by the June rankings *could have been* assigned to the top of the Russell 2000.

7.2 Discontinuity in lagged institutional ownership with Russell June rankings

7.2.1 March

- In tables 3 and 4, we hold the bandwidth constant at 200 and use second- and first-order polynomials, respectively. We presented the third-order case in the main paper.
- In tables 5, 6, and 7, we hold the bandwidth constant at 100 and use third-, second-, and first-order polynomials, respectively.

The large pre-existing discontinuities in March-dated total, quasi-indexer, and transient institutional ownership persist throughout a variety of specifications.

7.2.2 Last December

One could argue that there is measurement error in fiscal quarter end dates, such that the pre-existing discontinuities in March-dated institutional ownership are troubling but not fatal for the research design. To explore this possibility, we extend the lags to December *of the previous year*.

- In tables 8, 9, and 10, we hold the bandwidth constant at 200 and vary the polynomial orders.
- In tables 11, 12, and 13, we hold the bandwidth constant at 100 and vary the polynomial orders.

There continue to be large pre-existing discontinuities in *last* December's total, quasi-indexer, and transient institutional ownership in each set of specifications.

7.2.3 Last September

We also lag institutional ownership by 4 quarters to September *of the previous year*.

- In tables 14, 15, and 16, we hold the bandwidth constant at 200 and vary the polynomial orders.
- In tables 17, 18, and 19, we hold the bandwidth constant at 100 and vary the polynomial orders.

Once again, there are large pre-existing discontinuities in total, quasi-indexer, and transient institutional ownership in each set of specifications. Thus, the pre-existing discontinuities in March-dated institutional ownership cannot be explained by measurement error in the fiscal quarter end dates; that these discontinuities can be found a full year *prior to* reconstitution indicates that reconstitution cannot cause the discontinuity in current year September-dated institutional ownership. Hence, using the Russell June rankings results in *ex-ante* very different firms being compared with each other at the threshold.

7.3 No discontinuity in the *change* in institutional ownership with Russell June rankings

- In tables 20 and 21, we hold the bandwidth constant at 200 and vary the polynomial order. We presented the third-order case in the main paper.
- In tables 22, 23, and 24, we hold the bandwidth constant at 100 and vary the polynomial order.

Across all 5 tables, we find no evidence of any significant discontinuities in the *changes* in institutional ownership from March to September. In addition, the estimated coefficients on $R2000_{i,t}$ are also economically small and close to zero. Together with the pre-existing discontinuities in the *level* of institutional ownership, it is clear that firms around the cutoff by June rankings are different before reconstitution, and remain different *in much the same way* after reconstitution.

7.4 Is there a discontinuity in institutional ownership with researcher-constructed end-of-May rankings? No.

Chen et al. [2017] claim to show a discontinuity in total institutional ownership at the Russell 1000/2000 cutoff with CRSP May rankings: for their figure 1, Chen et al. [2017] write

> This figure presents graphical analyses [sic] of the institutional holdings for firms around the Russell 1000/2000 threshold based on end-of-May market capitalization rank calculated from CRSP.

In the left panel of our figure 1, we graph total institutional ownership against the researcher-constructed end-of-May rankings with the Chen et al. [2017] sample period of 2003–2006. Consistent with our paper's findings using a longer sample period, there is no discontinuity in total institutional ownership at the Russell 1000/2000 cutoff with these rankings. If instead we graph total institutional ownership against the Russell June rankings, as shown in the right panel of our figure 1, we observe a discontinuity similar to that shown in Chen et al. [2017].

Given that

1. we find no discontinuity in total institutional ownership at the Russell 1000/2000 cutoff with the researcher-constructed end-of-May rankings,

2. the lack of a discontinuity in total institutional ownership with the researcher-constructed rankings is also confirmed by Chang et al. [2015, p. 234] and Schmidt and Fahlenbrach [2017, p. 304] using different specifications,

3. Bird and Karolyi [2019] claimed to find a discontinuity in total institutional ownership with CRSP May rankings but were unable to provide the original data to reproduce the alleged discontinuity, resulting in their paper's retraction,

we are skeptical that Chen et al. [2017] used the CRSP May rankings in their figure 1. Chen et al.

[2017] or any other authors who claim to find a discontinuity in total or quasi-indexer institutional ownership with the CRSP May rankings are invited to post their code and data.

7.5 First stage robustness

- In table 25, we hold the bandwidth constant at 250 and vary the polynomial order. The Kleibergen and Paap [2006] F-statistics range from 29.309 to 1831.502, and the adjusted R^2s are all around 91.5%.

- In table 26, we hold the bandwidth constant at 200 and vary the polynomial order. The Kleibergen and Paap [2006] F-statistics range from 17.527 to 1200.024, and the adjusted R^2s are all around 90%.

- In table 27, we hold the bandwidth constant at 100 and vary the polynomial order. The adjusted R^2s are all around 83%. The Kleibergen and Paap [2006] F-statistics are 2.962, 33.538, and 299.960, respectively.

Only in column 1 of table 27 is there some evidence of a weak first stage based on the Kleibergen and Paap F-statistic, and this is likely due to the nature of the multivariate F-statistic calculation. For example, if we test the null hypothesis that the excluded instruments are jointly zero with the $\text{R2000}_{i,t}$ first stage, the F-statistic is 265.61. However, if we test the null with the $\text{R2000}_{i,t} \times \text{Rank}_{i,t}^3$ first stage, the F-statistic is only 1.48. The multivariate F-statistic is "dragged down" as a result of the $\text{R2000}_{i,t} \times \text{Rank}_{i,t}^3$ first stage. Thus, overall, the first stage in the fuzzy RD is very strong and robust to various bandwidths and polynomial orders.

7.6 Discontinuity in float adjustment

In table 28, we hold the bandwidth constant at 100 and vary the polynomial order. As in the main paper, there is approximately a 100 percentage point discontinuity in the float adjustment at the Russell 1000/2000 threshold.

7.7 Controlling for the float adjustment does not solve the problem

- In tables 29 and 30, we hold the bandwidth constant at 200 and vary the polynomial orders. (We presented the third-order case in the main paper.)

- In tables 31, 32, and 33, we hold the bandwidth constant at 100 and vary the polynomial orders.

Across all the tables, we continue to find significant pre-existing discontinuities in March-dated institutional ownership even after controlling for our measure of the float adjustment.

References

J. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics*. Princeton University Press, Princeton, New Jersey, 2009.

I. Appel, T. A. Gormley, and D. B. Keim. Identification Using Russell 1000/2000 Index Assignments: A Discussion of Methodologies, 2016a. URL `https://repository.upenn.edu/handle/20.500.14332/34365`. Dated: March 29, 2016.

I. Appel, T. A. Gormley, and D. B. Keim. Identification Using Russell 1000/2000 Index Assignments: A Discussion of Methodologies, 2018. URL `https://rodneywhitecenter.wharton.upenn.edu/wp-content/uploads/2018/11/04-18.Keim_.pdf`. Dated: October 17, 2018.

I. Appel, T. A. Gormley, and D. B. Keim. Identification Using Russell 1000/2000 Index Assignments: A Discussion of Methodologies, 2019. URL `https://osf.io/rx6mc`. Dated: May 20, 2019.

I. Appel, T. A. Gormley, and D. B. Keim. Identification Using Russell 1000/2000 Index Assignments: A Discussion of Methodologies, 2020. URL `https://osf.io/f3bze`. Dated: April 27, 2020.

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

A. Bird and S. A. Karolyi. **Retraction:** Governance and Taxes: Evidence from Regression Discontinuity. *The Accounting Review*, 2019. URL `https://doi.org/10.2308/1558-7967-92.1.000`.

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

Y. Chang, H. Hong, and I. Liskovich. Regression Discontinuity and the Price Effects of Stock Market Indexing. *Review of Financial Studies*, 28(1):212–246, 2015.

T. Chen, H. Dong, and C. Lin. Institutional Shareholders and Corporate Social Responsibility, 2017. URL https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CICF2017&paper_id=1104. Accepted at *Journal of Financial Economics* on February 26, 2018; available online on June 15, 2019.

A. D. Crane, S. Michenaud, and J. Weston. The Effect of Institutional Ownership on Payout Policy: Evidence from Index Thresholds. *Review of Financial Studies*, 29(6):1377–1408, 2016.

R. A. Dye. An evaluation of "essays on disclosure" and the disclosure literature in accounting. *Journal of Accounting and Economics*, 32(1):181–235, 2001.

D. Feir, T. Lemieux, and V. Marmer. Weak Identification in Fuzzy Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, 34(2):185–196, 2016.

FTSE Russell. Russell U.S. Equity Indexes: Construction and Methodology, 2017a. URL http://www.ftse.com/products/downloads/Russell-US-indexes.pdf. Version 2.9.

FTSE Russell. Free Float Restrictions: Update to become effective from June 2017, 2017b. URL http://www.ftse.com/products/downloads/Free_Float_Restrictions.pdf. Version 1.8.

J. Hahn, P. Todd, and W. Van der Klaauw. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209, 2001.

G. W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

F. Kleibergen and R. Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126, 2006.

D. S. Lee. Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, 142(2):675–697, 2008.

D. S. Lee and T. Lemieux. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

M. R. Roberts and T. M. Whited. *Endogeneity in Empirical Corporate Finance*, volume 2A of *Handbook of the Economics of Finance*. Elsevier, Amsterdam, The Netherlands, 2013.

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

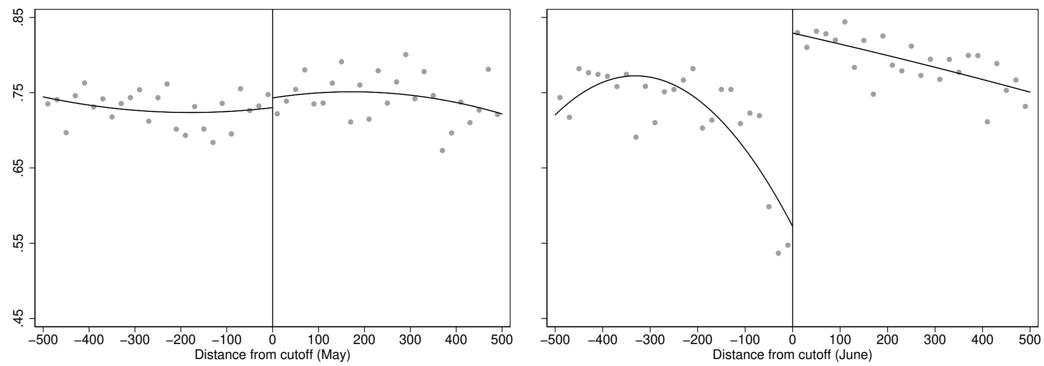

Figure 1: This figure is an attempt to replicate figure 1 of Chen et al. [2017]. The left (right) plots September-dated total institutional ownership against rankings based on CRSP and Compustat market capitalization in May (Russell's index weights in June) centered at the 1000$^{\text{th}}$ ranking. In the left panel, firms on the left (right) side of the vertical line are predicted to be in the Russell 1000 (Russell 2000) in June of year t; in the right panel, firms on the left (right) side of the vertical line are in the Russell 1000 (Russell 2000) in June of year t. Each bin represents the average of the y-axis variable over 20 ranks through the sample period. The curves fit the data using local polynomial (cubic) regression. The sample period is 2003–2006.

Table 1: Discontinuity in market capitalization, bandwidth 200

	(1) Market Cap.	(2) Market Cap.	(3) Market Cap.
R2000	-0.726***	-1.028***	-1.016***
	(0.107)	(0.115)	(0.092)
June Rank	-0.024***	0.009**	0.007***
	(0.009)	(0.004)	(0.001)
R2000 × June Rank	0.010**	-0.005***	-0.004***
	(0.004)	(0.002)	(0.001)
June Rank^2	-0.000***	0.000	
	(0.000)	(0.000)	
R2000 × June Rank^2	0.000***	-0.000	
	(0.000)	(0.000)	
June Rank^3	-0.000***		
	(0.000)		
R2000 × June Rank^3	0.000***		
	(0.000)		
Constant	2.099***	2.382***	2.357***
	(0.104)	(0.113)	(0.091)
Observations	4395	4395	4395
Adjusted R^2	0.247	0.241	0.241

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variable is CRSP and Compustat May market capitalization. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 2: Discontinuity in market capitalization, bandwidth 100

	(1) Market Cap.	(2) Market Cap.	(3) Market Cap.
R2000	-0.353***	-0.611***	-1.023***
	(0.101)	(0.093)	(0.114)
June Rank	-0.103***	-0.039***	0.007***
	(0.028)	(0.010)	(0.003)
R2000 × June Rank	0.051***	0.018***	-0.005***
	(0.014)	(0.005)	(0.001)
June Rank^2	-0.002***	-0.000***	
	(0.001)	(0.000)	
R2000 × June Rank^2	0.001***	0.000***	
	(0.000)	(0.000)	
June Rank^3	-0.000***		
	(0.000)		
R2000 × June Rank^3	0.000***		
	(0.000)		
Constant	1.717***	1.980***	2.379***
	(0.094)	(0.088)	(0.112)
Observations	2194	2194	2194
Adjusted R^2	0.247	0.242	0.228

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variable is CRSP and Compustat May market capitalization. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 3: Discontinuity in March institutional ownership, second-order, bandwidth 200

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.377***	0.264***	-0.011	0.119***
	(0.039)	(0.023)	(0.032)	(0.013)
June Rank	-0.008***	-0.007***	0.000	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.004***	0.003***	-0.000	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)
June Rank^2	-0.000***	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000***	0.000***	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.315***	0.139***	0.098***	0.080***
	(0.035)	(0.019)	(0.032)	(0.010)
Observations	4079	4079	4029	4078
Adjusted R^2	0.103	0.133	-0.000	0.043

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 4: Discontinuity in March institutional ownership, first-order, bandwidth 200

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.250***	0.179***	-0.008	0.076***
	(0.027)	(0.016)	(0.020)	(0.010)
June Rank	-0.003***	-0.002***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	0.001***	0.001***	-0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.422***	0.215***	0.095***	0.113***
	(0.023)	(0.013)	(0.019)	(0.008)
Observations	4079	4079	4029	4078
Adjusted R^2	0.087	0.114	0.000	0.035

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^{n} + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^{n} + \varepsilon_{i,t}$$

Table 5: Discontinuity in March institutional ownership, third-order, bandwidth 100

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.376***	0.234***	0.006	0.129***
	(0.057)	(0.032)	(0.042)	(0.022)
June Rank	-0.005	-0.002	-0.005	0.003
	(0.008)	(0.004)	(0.004)	(0.003)
R2000 × June Rank	0.002	0.001	0.003	-0.002
	(0.005)	(0.003)	(0.002)	(0.002)
June Rank^2	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	0.000	0.000	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	-0.000	-0.000	0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.319***	0.158***	0.075*	0.090***
	(0.046)	(0.022)	(0.040)	(0.015)
Observations	2035	2035	2002	2034
Adjusted R^2	0.146	0.185	0.000	0.073

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \mathrm{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \mathrm{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \mathrm{R2000}_{i,t} \times \mathrm{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 6: Discontinuity in March institutional ownership, second-order, bandwidth 100

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.391***	0.251***	-0.010	0.145***
	(0.049)	(0.028)	(0.042)	(0.017)
June Rank	-0.009***	-0.007***	-0.000	-0.002
	(0.003)	(0.002)	(0.003)	(0.001)
R2000 × June Rank	0.004**	0.004***	0.000	0.000
	(0.002)	(0.001)	(0.001)	(0.001)
June Rank^2	-0.000	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000	-0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.302***	0.139***	0.095**	0.073***
	(0.042)	(0.022)	(0.041)	(0.012)
Observations	2035	2035	2002	2034
Adjusted R^2	0.147	0.185	-0.001	0.072

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 7: Discontinuity in March institutional ownership, first-order, bandwidth 100

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.354***	0.253***	-0.014	0.111***
	(0.038)	(0.022)	(0.031)	(0.013)
June Rank	-0.006***	-0.005***	0.000	-0.001***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.003***	0.002***	-0.000	0.001***
	(0.001)	(0.000)	(0.000)	(0.000)
Constant	0.332***	0.149***	0.100***	0.086***
	(0.033)	(0.018)	(0.031)	(0.010)
Observations	2035	2035	2002	2034
Adjusted R^2	0.147	0.186	-0.000	0.069

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 8: Discontinuity in last December's institutional ownership, third-order, bandwidth 200

	(1) IO L12	(2) QIX L12	(3) DED L12	(4) TRA L12
R2000	0.407***	0.282***	-0.016	0.132***
	(0.051)	(0.029)	(0.046)	(0.017)
June Rank	-0.015***	-0.011***	0.000	-0.004***
	(0.003)	(0.002)	(0.003)	(0.001)
R2000 × June Rank	0.008***	0.006***	0.000	0.002**
	(0.002)	(0.001)	(0.001)	(0.001)
June Rank^2	-0.000***	-0.000**	0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000**	0.000**	-0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	-0.000**	-0.000**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	0.000**	0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.259***	0.108***	0.095**	0.063***
	(0.045)	(0.024)	(0.045)	(0.013)
Observations	4027	4027	3976	4025
Adjusted R^2	0.083	0.084	-0.001	0.039

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) December-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \mathrm{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \mathrm{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \mathrm{R2000}_{i,t} \times \mathrm{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 9: Discontinuity in last December's institutional ownership, second-order, bandwidth 200

	(1) IO L12	(2) QIX L12	(3) DED L12	(4) TRA L12
R2000	0.370***	0.259***	-0.008	0.113***
	(0.041)	(0.023)	(0.035)	(0.014)
June Rank	-0.009***	-0.007***	0.000	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.004***	0.003***	-0.000	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)
June Rank^2	-0.000***	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000***	0.000***	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.307***	0.140***	0.093***	0.078***
	(0.036)	(0.020)	(0.034)	(0.011)
Observations	4027	4027	3976	4025
Adjusted R^2	0.082	0.083	-0.001	0.039

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) December-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 10: Discontinuity in last December's institutional ownership, first-order, bandwidth 200

	(1) IO L12	(2) QIX L12	(3) DED L12	(4) TRA L12
R2000	0.237***	0.169***	-0.007	0.073***
	(0.029)	(0.019)	(0.020)	(0.011)
June Rank	-0.003***	-0.002***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	0.001***	0.001***	-0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.423***	0.222***	0.091***	0.112***
	(0.025)	(0.016)	(0.019)	(0.009)
Observations	4027	4027	3976	4025
Adjusted R^2	0.066	0.069	-0.000	0.032

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) December-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^{n} + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^{n} + \varepsilon_{i,t}$$

Table 11: Discontinuity in last December's institutional ownership, third-order, bandwidth 100

	(1) IO L12	(2) QIX L12	(3) DED L12	(4) TRA L12
R2000	0.355***	0.226***	-0.001	0.118***
	(0.061)	(0.032)	(0.049)	(0.023)
June Rank	-0.008	-0.003	-0.005	0.001
	(0.008)	(0.004)	(0.005)	(0.003)
R2000 × June Rank	0.005	0.002	0.003	-0.001
	(0.005)	(0.003)	(0.002)	(0.002)
June Rank^2	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.298***	0.151***	0.074	0.082***
	(0.051)	(0.024)	(0.048)	(0.017)
Observations	1993	1993	1959	1992
Adjusted R^2	0.110	0.101	-0.001	0.066

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) December-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 12: Discontinuity in last December's institutional ownership, second-order, bandwidth 100

	(1) IO L12	(2) QIX L12	(3) DED L12	(4) TRA L12
R2000	0.398***	0.265***	-0.012	0.135***
	(0.054)	(0.031)	(0.047)	(0.018)
June Rank	-0.013***	-0.009***	-0.000	-0.003**
	(0.004)	(0.003)	(0.003)	(0.001)
R2000 × June Rank	0.006***	0.005***	0.000	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
June Rank^2	-0.000	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000	0.000	-0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.269***	0.117***	0.092**	0.067***
	(0.048)	(0.027)	(0.047)	(0.013)
Observations	1993	1993	1959	1992
Adjusted R^2	0.110	0.101	-0.002	0.066

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) December-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 13: Discontinuity in last December's institutional ownership, first-order, bandwidth 100

	(1) IO L12	(2) QIX L12	(3) DED L12	(4) TRA L12
R2000	0.350***	0.251***	-0.011	0.104***
	(0.039)	(0.022)	(0.033)	(0.013)
June Rank	-0.006***	-0.005***	0.000	-0.001***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.003***	0.002***	-0.000	0.001***
	(0.001)	(0.000)	(0.000)	(0.000)
Constant	0.323***	0.148***	0.094***	0.085***
	(0.034)	(0.019)	(0.032)	(0.010)
Observations	1993	1993	1959	1992
Adjusted R^2	0.108	0.100	-0.001	0.064

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) December-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 14: Discontinuity in last September's institutional ownership, third-order, bandwidth 200

	(1) IO L9	(2) QIX L9	(3) DED L9	(4) TRA L9
R2000	0.389***	0.273***	-0.025	0.131***
	(0.054)	(0.030)	(0.051)	(0.018)
June Rank	-0.012***	-0.008***	0.001	-0.003***
	(0.003)	(0.002)	(0.003)	(0.001)
R2000 × June Rank	0.006***	0.004***	-0.000	0.001**
	(0.002)	(0.001)	(0.001)	(0.001)
June Rank^2	-0.000*	-0.000**	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000	0.000*	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.273***	0.121***	0.106**	0.059***
	(0.049)	(0.026)	(0.051)	(0.014)
Observations	3942	3941	3872	3939
Adjusted R^2	0.091	0.117	-0.000	0.034

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) September-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock *i* in year *t* centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 15: Discontinuity in last September's institutional ownership, second-order, bandwidth 200

	(1) IO L9	(2) QIX L9	(3) DED L9	(4) TRA L9
R2000	0.372***	0.260***	-0.018	0.122***
	(0.043)	(0.024)	(0.039)	(0.014)
June Rank	-0.009***	-0.006***	0.000	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
R2000 × June Rank	0.004***	0.003***	-0.000	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)
June Rank^2	-0.000***	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000***	0.000***	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.297***	0.138***	0.101***	0.067***
	(0.039)	(0.021)	(0.038)	(0.012)
Observations	3942	3941	3872	3939
Adjusted R^2	0.091	0.116	0.000	0.034

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) September-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 16: Discontinuity in last September's institutional ownership, first-order, bandwidth 200

	(1) IO L9	(2) QIX L9	(3) DED L9	(4) TRA L9
R2000	0.247***	0.176***	-0.012	0.078***
	(0.028)	(0.017)	(0.022)	(0.011)
June Rank	-0.003***	-0.002***	0.000	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	0.001***	0.001***	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.410***	0.214***	0.095***	0.105***
	(0.025)	(0.014)	(0.022)	(0.009)
Observations	3942	3941	3872	3939
Adjusted R^2	0.076	0.100	0.000	0.027

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) September-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 17: Discontinuity in last September's institutional ownership, third-order, bandwidth 100

	(1) IO L9	(2) QIX L9	(3) DED L9	(4) TRA L9
R2000	0.345***	0.223***	-0.004	0.107***
	(0.066)	(0.035)	(0.057)	(0.024)
June Rank	-0.011	-0.005	-0.005	0.001
	(0.008)	(0.004)	(0.006)	(0.003)
R2000 × June Rank	0.007	0.004	0.003	-0.000
	(0.005)	(0.003)	(0.003)	(0.002)
June Rank^2	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.293***	0.150***	0.078	0.083***
	(0.057)	(0.027)	(0.056)	(0.017)
Observations	1938	1937	1891	1935
Adjusted R^2	0.138	0.173	0.001	0.067

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) September-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 18: Discontinuity in last September's institutional ownership, second-order, bandwidth 100

	(1) IO L9	(2) QIX L9	(3) DED L9	(4) TRA L9
R2000	0.367***	0.248***	-0.018	0.126***
	(0.056)	(0.031)	(0.053)	(0.018)
June Rank	-0.010***	-0.007***	-0.000	-0.002
	(0.004)	(0.002)	(0.003)	(0.001)
R2000 × June Rank	0.005**	0.004***	0.000	0.001
	(0.002)	(0.001)	(0.002)	(0.001)
June Rank^2	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.289***	0.135***	0.098*	0.069***
	(0.050)	(0.026)	(0.053)	(0.014)
Observations	1938	1937	1891	1935
Adjusted R^2	0.139	0.173	0.000	0.067

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) September-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \mathrm{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \mathrm{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \mathrm{R2000}_{i,t} \times \mathrm{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 19: Discontinuity in last September's institutional ownership, first-order, bandwidth 100

	(1) IO L9	(2) QIX L9	(3) DED L9	(4) TRA L9
R2000	0.353***	0.254***	-0.021	0.114***
	(0.041)	(0.023)	(0.037)	(0.014)
June Rank	-0.006***	-0.005***	0.000	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.003***	0.002***	-0.000	0.001***
	(0.001)	(0.000)	(0.000)	(0.000)
Constant	0.316***	0.147***	0.104***	0.073***
	(0.037)	(0.020)	(0.037)	(0.011)
Observations	1938	1937	1891	1935
Adjusted R^2	0.139	0.173	0.001	0.068

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are (last year's) September-dated total, quasi-indexer, dedicated, and transient institutional ownership. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 20: No discontinuity in the change in ownership, second-order, bandwidth 200

	(1) $\Delta IO_{3\to9}$	(2) $\Delta QIX_{3\to9}$	(3) $\Delta DED_{3\to9}$	(4) $\Delta TRA_{3\to9}$
R2000	-0.013	-0.000	-0.002	-0.013
	(0.016)	(0.007)	(0.004)	(0.014)
June Rank	0.001**	0.000	-0.000	0.001**
	(0.001)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	-0.001*	-0.000	0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^2	0.000*	0.000	-0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000*	-0.000	0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.048***	0.019***	0.003	0.027**
	(0.014)	(0.005)	(0.002)	(0.012)
Observations	4000	3989	3915	3987
Adjusted R^2	0.004	0.004	-0.000	0.004

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are the changes in total, quasi-indexer, dedicated, and transient institutional ownership (from March to September). R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\Delta Y_{i,3\to9,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 21: No discontinuity in the change in ownership, first-order, bandwidth 200

	(1) $\Delta IO_{3\to9}$	(2) $\Delta QIX_{3\to9}$	(3) $\Delta DED_{3\to9}$	(4) $\Delta TRA_{3\to9}$
R2000	0.005	0.006	-0.002	0.004
	(0.010)	(0.005)	(0.003)	(0.007)
June Rank	0.000***	0.000	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	-0.000***	-0.000	-0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.032***	0.015***	0.005***	0.013**
	(0.008)	(0.004)	(0.002)	(0.007)
Observations	4000	3989	3915	3987
Adjusted R^2	0.003	0.004	0.000	0.002

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are the changes in total, quasi-indexer, dedicated, and transient institutional ownership (from March to September). R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\Delta Y_{i,3\to9,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 22: No discontinuity in the change in ownership, third-order, bandwidth 100

	(1) $\Delta IO_{3\to 9}$	(2) $\Delta QIX_{3\to 9}$	(3) $\Delta DED_{3\to 9}$	(4) $\Delta TRA_{3\to 9}$
R2000	-0.042	0.002	-0.003	-0.038
	(0.031)	(0.015)	(0.009)	(0.026)
June Rank	0.007*	0.002	0.000	0.004
	(0.004)	(0.002)	(0.001)	(0.003)
R2000 × June Rank	-0.004	-0.001	-0.000	-0.002
	(0.002)	(0.001)	(0.001)	(0.002)
June Rank^2	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.078***	0.023**	0.006	0.049**
	(0.027)	(0.011)	(0.005)	(0.023)
Observations	1993	1990	1941	1988
Adjusted R^2	0.003	0.001	-0.001	0.006

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are the changes in total, quasi-indexer, dedicated, and transient institutional ownership (from March to September). R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\Delta Y_{i,3\to 9,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 23: No discontinuity in the change in ownership, second-order, bandwidth 100

	(1) $\Delta IO_{3\to9}$	(2) $\Delta QIX_{3\to9}$	(3) $\Delta DED_{3\to9}$	(4) $\Delta TRA_{3\to9}$
R2000	-0.025	0.005	-0.002	-0.031
	(0.025)	(0.011)	(0.006)	(0.022)
June Rank	0.003	0.000	0.000	0.002
	(0.002)	(0.001)	(0.000)	(0.001)
R2000 × June Rank	-0.001	-0.000	-0.000	-0.001
	(0.001)	(0.000)	(0.000)	(0.001)
June Rank^2	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.060***	0.016**	0.006*	0.040**
	(0.022)	(0.008)	(0.004)	(0.020)
Observations	1993	1990	1941	1988
Adjusted R^2	0.003	0.002	0.000	0.006

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are the changes in total, quasi-indexer, dedicated, and transient institutional ownership (from March to September). R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\Delta Y_{i,3\to9,t} = \beta_0 + \beta_1 \mathrm{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \mathrm{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \mathrm{R2000}_{i,t} \times \mathrm{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 24: No discontinuity in the change in ownership, first-order, bandwidth 100

	(1) $\Delta IO_{3\to 9}$	(2) $\Delta QIX_{3\to 9}$	(3) $\Delta DED_{3\to 9}$	(4) $\Delta TRA_{3\to 9}$
R2000	-0.014	-0.000	0.000	-0.015
	(0.015)	(0.007)	(0.004)	(0.013)
June Rank	0.001**	0.000	-0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	-0.000**	-0.000	0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.048***	0.018***	0.002	0.029**
	(0.013)	(0.005)	(0.002)	(0.012)
Observations	1993	1990	1941	1988
Adjusted R^2	0.004	0.002	0.000	0.006

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are the changes in total, quasi-indexer, dedicated, and transient institutional ownership (from March to September). R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock *i* in year *t* centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\Delta Y_{i,3\to 9,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 25: First stage regression, bandwidth 250

	(1) R2000	(2) R2000	(3) R2000
τ	0.604***	0.737***	0.861***
	(0.041)	(0.027)	(0.014)
Rank	0.005***	0.002***	0.000***
	(0.001)	(0.000)	(0.000)
$\tau \times$ Rank	0.001	-0.000	0.000
	(0.001)	(0.000)	(0.000)
Rank^2	0.000***	0.000***	
	(0.000)	(0.000)	
$\tau \times$ Rank^2	-0.000***	-0.000***	
	(0.000)	(0.000)	
Rank^3	0.000***		
	(0.000)		
$\tau \times$ Rank^3	0.000		
	(0.000)		
Constant	0.192***	0.133***	0.070***
	(0.028)	(0.018)	(0.010)
Observations	5053	5053	5053
Adjusted R^2	0.917	0.915	0.912

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variable is an indicator variable for actual membership in the Russell 2000 Index. τ is an indicator variable for predicted membership in the Russell 2000 Index. Rank is the researcher-constructed ranking of end-of-May market capitalization for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 250 [Appel et al., 2016b]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\text{R2000}_{i,t} = \alpha_0 + \alpha_1 \tau_{i,t} + \sum_{n=1}^{k} \alpha_{2n} \text{Rank}_{i,t}^{n} + \sum_{n=1}^{k} \alpha_{3n} \tau_{i,t} \times \text{Rank}_{i,t}^{n} + \nu_{i,t}$$

Table 26: First stage regression, bandwidth 200

	(1) R2000	(2) R2000	(3) R2000
τ	0.549***	0.688***	0.833***
	(0.048)	(0.032)	(0.017)
Rank	0.006***	0.003***	0.001***
	(0.001)	(0.000)	(0.000)
$\tau \times$ Rank	0.002	0.000	-0.000
	(0.002)	(0.001)	(0.000)
Rank^2	0.000***	0.000***	
	(0.000)	(0.000)	
$\tau \times$ Rank^2	-0.000***	-0.000***	
	(0.000)	(0.000)	
Rank^3	0.000***		
	(0.000)		
$\tau \times$ Rank^3	0.000		
	(0.000)		
Constant	0.211***	0.156***	0.085***
	(0.032)	(0.022)	(0.012)
Observations	4041	4041	4041
Adjusted R^2	0.903	0.901	0.896

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variable is an indicator variable for actual membership in the Russell 2000 Index. τ is an indicator variable for predicted membership in the Russell 2000 Index. Rank is the researcher-constructed ranking of end-of-May market capitalization for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\text{R2000}_{i,t} = \alpha_0 + \alpha_1 \tau_{i,t} + \sum_{n=1}^{k} \alpha_{2n} \text{Rank}_{i,t}^{n} + \sum_{n=1}^{k} \alpha_{3n} \tau_{i,t} \times \text{Rank}_{i,t}^{n} + \nu_{i,t}$$

Table 27: First stage regression, bandwidth 100

	(1) R2000	(2) R2000	(3) R2000
τ	0.310***	0.514***	0.715***
	(0.074)	(0.052)	(0.029)
Rank	0.018***	0.007***	0.002***
	(0.004)	(0.001)	(0.000)
$\tau \times$ Rank	0.004	0.002	0.000
	(0.005)	(0.002)	(0.000)
Rank^2	0.000***	0.000***	
	(0.000)	(0.000)	
$\tau \times$ Rank^2	-0.001***	-0.000***	
	(0.000)	(0.000)	
Rank^3	0.000***		
	(0.000)		
$\tau \times$ Rank^3	0.000		
	(0.000)		
Constant	0.317***	0.226***	0.142***
	(0.051)	(0.036)	(0.020)
Observations	2026	2026	2026
Adjusted R^2	0.836	0.830	0.822

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variable is an indicator variable for actual membership in the Russell 2000 Index. τ is an indicator variable for predicted membership in the Russell 2000 Index. Rank is the researcher-constructed ranking of end-of-May market capitalization for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$\text{R2000}_{i,t} = \alpha_0 + \alpha_1 \tau_{i,t} + \sum_{n=1}^{k} \alpha_{2n} \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \alpha_{3n} \tau_{i,t} \times \text{Rank}_{i,t}^n + \nu_{i,t}$$

Table 28: Float adjustment discontinuity, Russell June rankings, bandwidth 100

	(1) Float Adj.	(2) Float Adj.	(3) Float Adj.
R2000	1.023***	1.029***	0.987***
	(0.096)	(0.061)	(0.046)
June Rank	-0.012	-0.016***	-0.011***
	(0.009)	(0.003)	(0.001)
R2000 × June Rank	0.005	0.008***	0.006***
	(0.008)	(0.003)	(0.001)
June Rank^2	0.000	-0.000	
	(0.000)	(0.000)	
R2000 × June Rank^2	0.000	0.000	
	(0.000)	(0.000)	
June Rank^3	0.000		
	(0.000)		
R2000 × June Rank^3	-0.000		
	(0.000)		
Constant	-0.729***	-0.744***	-0.703***
	(0.035)	(0.031)	(0.027)
Observations	2193	2193	2193
Adjusted R^2	0.326	0.326	0.327

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variable is the float adjustment, estimated as the percentage difference between the Russell June index weight and the CRSP-calculated June index weight. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \varepsilon_{i,t}$$

Table 29: Controlling for the float adjustment, second-order, bandwidth 200

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.329***	0.237***	-0.011	0.110***
	(0.044)	(0.025)	(0.031)	(0.014)
June Rank	-0.008***	-0.006***	0.000	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.004***	0.003***	-0.000	0.001**
	(0.001)	(0.000)	(0.001)	(0.000)
June Rank^2	-0.000***	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	0.000***	0.000***	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Float Adj.	0.044**	0.025**	-0.000	0.008**
	(0.020)	(0.011)	(0.005)	(0.004)
Constant	0.348***	0.158***	0.098***	0.087***
	(0.037)	(0.020)	(0.031)	(0.011)
Observations	4078	4078	4028	4077
Adjusted R^2	0.112	0.139	-0.000	0.044

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. The float adjustment is estimated as the percentage difference between the Russell June index weight and the CRSP-calculated June index weight. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \beta_2 \text{Float Adj.}_{i,t} + \varepsilon_{i,t}$$

Table 30: Controlling for the float adjustment, first-order, bandwidth 200

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.206***	0.154***	-0.008	0.067***
	(0.032)	(0.018)	(0.018)	(0.011)
June Rank	-0.002***	-0.002***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank	0.001***	0.001***	-0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Float Adj.	0.049**	0.029**	-0.000	0.010**
	(0.021)	(0.012)	(0.005)	(0.004)
Constant	0.451***	0.232***	0.094***	0.119***
	(0.025)	(0.014)	(0.018)	(0.008)
Observations	4078	4078	4028	4077
Adjusted R^2	0.097	0.122	0.000	0.037

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. The float adjustment is estimated as the percentage difference between the Russell June index weight and the CRSP-calculated June index weight. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock *i* in year *t* centered at the cutoff for Russell 2000 membership. The bandwidth is 200 [Boone and White, 2015]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \beta_2 \text{Float Adj.}_{i,t} + \varepsilon_{i,t}$$

Table 31: Controlling for the float adjustment, third-order, bandwidth 100

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.287***	0.190***	0.006	0.107***
	(0.065)	(0.037)	(0.041)	(0.024)
June Rank	-0.004	-0.002	-0.005	0.003
	(0.007)	(0.004)	(0.004)	(0.003)
R2000 × June Rank	0.001	0.001	0.003	-0.002
	(0.005)	(0.003)	(0.002)	(0.002)
June Rank^2	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^2	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
June Rank^3	0.000	0.000	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 × June Rank^3	-0.000	-0.000	0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Float Adj.	0.086***	0.042**	-0.001	0.021**
	(0.031)	(0.019)	(0.007)	(0.009)
Constant	0.383***	0.189***	0.075*	0.106***
	(0.050)	(0.026)	(0.040)	(0.017)
Observations	2034	2034	2001	2033
Adjusted R^2	0.172	0.199	-0.000	0.080

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. The float adjustment is estimated as the percentage difference between the Russell June index weight and the CRSP-calculated June index weight. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \beta_2 \text{Float Adj.}_{i,t} + \varepsilon_{i,t}$$

Table 32: Controlling for the float adjustment, second-order, bandwidth 100

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.302***	0.208***	-0.010	0.123***
	(0.057)	(0.033)	(0.040)	(0.019)
June Rank	-0.008**	-0.006***	-0.000	-0.001
	(0.003)	(0.002)	(0.003)	(0.001)
R2000 $\times$ June Rank	0.003*	0.003***	0.000	-0.000
	(0.002)	(0.001)	(0.001)	(0.001)
June Rank^2	-0.000	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
R2000 $\times$ June Rank^2	0.000	-0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Float Adj.	0.086***	0.042**	-0.001	0.021**
	(0.031)	(0.020)	(0.007)	(0.009)
Constant	0.367***	0.170***	0.095**	0.089***
	(0.047)	(0.026)	(0.039)	(0.014)
Observations	2034	2034	2001	2033
Adjusted R^2	0.172	0.200	-0.001	0.080

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. The float adjustment is estimated as the percentage difference between the Russell June index weight and the CRSP-calculated June index weight. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \beta_2 \text{Float Adj.}_{i,t} + \varepsilon_{i,t}$$

Table 33: Controlling for the float adjustment, first-order, bandwidth 100

	(1) IO 3	(2) QIX 3	(3) DED 3	(4) TRA 3
R2000	0.268***	0.211***	-0.013	0.090***
	(0.048)	(0.028)	(0.027)	(0.015)
June Rank	-0.005***	-0.004***	0.000	-0.001***
	(0.001)	(0.001)	(0.001)	(0.000)
R2000 × June Rank	0.002***	0.002***	-0.000	0.000**
	(0.001)	(0.000)	(0.000)	(0.000)
Float Adj.	0.086***	0.042**	-0.001	0.021**
	(0.031)	(0.020)	(0.007)	(0.008)
Constant	0.393***	0.179***	0.100***	0.101***
	(0.038)	(0.022)	(0.028)	(0.012)
Observations	2034	2034	2001	2033
Adjusted R^2	0.172	0.200	-0.001	0.077

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

This table reports the results of estimating the following equation with OLS. The outcome variables are March-dated total, quasi-indexer, dedicated, and transient institutional ownership. The float adjustment is estimated as the percentage difference between the Russell June index weight and the CRSP-calculated June index weight. R2000 is an indicator variable for membership in the Russell 2000 Index. June Rank is the within-index ranking of the Russell June index weights for stock i in year t centered at the cutoff for Russell 2000 membership. The bandwidth is 100 [Crane et al., 2016]. Standard errors are clustered by firm. The sample period is 1996–2006.

$$Y_{i,q,t} = \beta_0 + \beta_1 \text{R2000}_{i,t} + \sum_{n=1}^{k} \gamma_n \text{Rank}_{i,t}^n + \sum_{n=1}^{k} \delta_n \text{R2000}_{i,t} \times \text{Rank}_{i,t}^n + \beta_2 \text{Float Adj.}_{i,t} + \varepsilon_{i,t}$$