Regression Discontinuity Versus Instrumental Variables: A Response to Appel, Gormley, and Keim (2020)

Supplementary Material

Wei Wei Alex Young

January 8, 2021

Abstract

Appel, Gormley, and Keim (2016, 2019, 2020) claim that Russell 1000/2000 Index assignment is a valid IV after controlling for observed size; while Wei and Young (2020) explain that without controlling for the unobserved Russell size, index assignment is endogenous and must be instrumented. To resolve the debate, we conduct a Monte Carlo simulation using the data procedure from Appel et al. (2020). We find that despite the observed size controls, the Appel et al. IV groups systematically differ on Russell size by over $910 million. However, when index assignment is instrumented by predicted index assignment, this difference is eliminated. Therefore, after controlling for observed size, index assignment is not a valid IV.

[1]Price College of Business, University of Oklahoma and Zarb School of Business, Hofstra University. We thank Ivo Welch for suggesting the use of simulations in this setting; and we are grateful to Ian Appel, Todd Gormley, and Donald Keim for sharing code. Replication files are posted at `https://osf.io/2wgmc/files/`
E-mails: `weiwei@ou.edu` and `alex.young@hofstra.edu`.

1 Introduction

If an ideal research design cannot be implemented because its key control variable is unobserved, then is it a valid alternative to simply control for an observed-but-noisy version of the key variable? In this note, we answer the question in the specific context of the Russell 1000/2000 reconstitution setting. Index assignments are determined by a threshold rule with Russell's descending-order ranking of an unobserved market capitalization. The ideal implementation of the setting would be a sharp regression discontinuity design that estimates the effect of index assignment at the threshold after controlling for the unobserved size ranking [Lee and Lemieux, 2010, p. 307], and the issue is whether the practical alternative of controlling for observed size can recover the true effects.

Appel, Gormley, and Keim [2016, 2019, 2020] (hereafter AGK) claim that researchers do not need the unobserved *Russell* size because after controlling for *observed* size, the remaining component of Russell size does not matter for the outcome of interest. Hence, they conclude that Russell 2000 Index assignment is a conditionally valid instrumental variable (IV) for passive institutional ownership (IO).

On the other hand, Wei and Young [2020] explain that after controlling for the Russell size rankings, Russell 2000 Index assignment has no variation (i.e. its conditional distribution is degenerate). Thus, it trivially follows that index assignment is conditionally exogenous (Imbens and Lemieux [2008, p. 618], Lee and Lemieux [2010, p. 289]). Unfortunately, this desirable property does not apply in practice because the Russell size rankings are unobservable and hence cannot be controlled for. More importantly, because of the threshold rule, firms with similar *observed* size but different index assignments must also differ systematically on *Russell* size. Therefore, conditional on observed size, index assignment is not a valid IV and instead must be instrument**ed**.

In the absence of a setting with known causes and effects, the debate has been difficult to resolve. Recently, Appel et al. [2020] have made a substantial contribution to the literature by simulating the Russell reconstitution setting and thus providing an ideal, controlled testing environment. They showed that given a known null effect of index assignment on total IO and a known positive effect on passive IO, the AGK methodology can correctly detect both effects.

Although the Appel et al. [2020] simulation approach is conceptually sound, their choice of dependent variables has a major drawback: when plotted across the Russell size rankings, the slopes of total and passive IO are zero on both sides of the threshold. Consequently, even a simple univariate comparison of means will recover the effects of index assignment on total and passive IO. If no control variables are needed to recover those effects, then these two dependent variables are not suitable for substantiating the AGK claim that index assignment is a valid IV for passive IO after including observed size controls.

Fortunately, the Appel et al. [2020] simulation contains another variable that can be used to meaningfully assess their claim and thus further the debate. By construction, Russell size is negatively sloped across its descending-order rankings and is continuous at the threshold between the two Indexes. Due to the steepness of the slope, a simple approach such as a univariate comparison of means can no longer recover the known null effect of index assignment on Russell size at the threshold. But if the AGK claim is correct that index assignment is a valid IV after controlling for *observed* size, then the conditional difference in *Russell* size between the AGK IV groups should be negligible.

To test this prediction, we conduct a Monte Carlo simulation based on the code in Appel et al. [2020]. We change only the dependent variable from simulated IO to simulated Russell size. Across 1000 repetitions, we find that after applying a bandwidth and controlling for *observed* size, on average Russell 2000 firms are smaller than Russell 1000 firms by over $**910 million** in *Russell* size. Thus, the AGK methodology spuriously detects a substantial conditional difference in Russell size between the two Indexes that by design should be immaterial. Hence, the Appel et al. [2020] simulation shows that the AGK claim is incorrect: after controlling for observed size, index assignment is **not** a valid IV.

We explain why the AGK claim is incorrect by comparing and contrasting the AGK methodology with a fuzzy regression discontinuity (RD) design. The fuzzy RD design similarly controls for (the ranking of) observed size, but differs by treating Russell 2000 Index assignment as an endogenous variable to be instrument**ed**. Using the same Monte Carlo simulation, we find that the fuzzy RD design correctly detects that the difference in Russell size at the threshold is immaterial. We show that it is the AGK methodology's incorrect use of index assignment as an IV that creates its substantial but spurious conditional difference in Russell size, and we explain the implications for the post-banding period after 2006.

Section 2 outlines the Appel et al. [2020] simulation. Section 3 presents the Monte Carlo simulation results. Section 4 discusses potential objections to our interpretation of the simulation results. Section 5 concludes.

2 The AGK Simulation

2.1 Background

Appel et al. [2020] simulated the Russell reconstitution setting to compare various methodologies. Their approach was to simulate no effect of Russell 1000/2000 Index assignment on total institutional ownership (IO), simulate a 1 percentage point effect of index assignment on passive IO, and then determine which methodologies could correctly recover these known effects.

One of the methodologies discussed was the Appel et al. [2016] (AGK) IV estimation

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

where *R2000* is the IV for passive IO after conditioning on *Mktcap*, and the variables are defined as follows:

- *Passive%* is passive IO
- *R2000* is Russell 1000/2000 Index assignment
- *Mktcap* is observed size
- *Float* is the float-adjusted size from Russell
- δ_t are year fixed effects

It is important to note that prior to 2007, *R2000* was exclusively determined by Russell's descending-order ranking of an *unobserved* size:

$$R2000_{it} = \begin{cases} 1 & \text{if Rank}(Russell\ Mktcap_{it}) > 1000 \\ 0 & \text{if Rank}(Russell\ Mktcap_{it}) \leq 1000 \end{cases}$$

A debate in the literature is whether *R2000* can be a valid IV for passive IO without controlling for the unobserved *Russell Mktcap* ranking. Appel et al. [2020] showed that the AGK methodology can recover both the simulated null effect of index membership on total IO and the simulated 1 percentage point effect on passive IO. They thus concluded,

> [Wei and Young [2020]] 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.

This claim is clearly incorrect, as illustrated by the prior estimations using simulated data (e.g., see Table 7). **Estimators that use index assignment as an IV can recover the correct coefficients even when they are unable to control for Russell's unobserved total end-of-May market caps** (emphasis added).

2.2 Problem

However, in the Appel et al. [2020] simulation, when plotted across the Russell size rankings, on average total IO is flat. Passive IO is initially calculated as 4.6% of total IO; and the effect of index assignment on passive IO is simulated by adding 1 percentage point to the passive IO observation *for all firms in the Russell 2000*, as shown in figure 1. Therefore, the main result of the Appel et al. [2020] simulation is that the AGK approach can successfully detect a horizontal level difference between its IV groups that is already built in.[1]

Although simulations simplify reality, the lack of a slope in total and passive IO trivializes the problem: even a univariate comparison of means will correctly recover the null effect of index assignment on total IO and the 1 percentage point positive effect on passive IO; no control variables are needed. This suggests those two dependent variables are not suitable for examining the validity of an empirical methodology that relies on controls. To substantiate the AGK claim – after controlling for observed size, Russell 2000 Index assignment can be a valid IV for passive IO – a more meaningful dependent variable must be used.

2.3 Solution

By design, the Appel et al. [2020] simulation already contains such a dependent variable. Russell size decreases across its descending-order rankings and is continuous at the threshold

[1] The bottom panel of Figure 5 in Appel et al. [2020] also shows the horizontal level difference within a bandwidth of 250.

between the Russell 1000 and 2000. As shown in the left panel of figure 2, the negative slope is sufficiently steep that a univariate comparison of means will not recover the null effect of index assignment on Russell size at the threshold. But as shown in the right panel of figure 2, a valid methodology will confirm that the difference in Russell size at the threshold between the two Indexes is immaterial.

Hence, if the AGK claim is correct that index assignment is a valid IV after controlling for simulated *observed* size, then we should see no significant conditional difference in simulated *Russell* size between the two Indexes. Observing otherwise would indicate that the AGK methodology spuriously detects an "effect" that by design does not exist, and therefore the AGK claim is incorrect.

3 Results

Our objective is to test whether a given methodology correctly recovers the known null effect of index assignment on Russell size at the threshold. We follow the Appel et al. [2020] code Appendix – including the reported bandwidth and polynomial order – and make only two modifications to their code:

1. We replace the dependent variable from simulated passive institutional ownership (IO) to simulated Russell size.

2. The Appel et al. [2020] results are based on a single instance of simulated data. We conduct a Monte Carlo simulation with 1000 repetitions.[2]

The average difference in simulated Russell size between the Russell 1000 and 2000 by

[2]To generate the proportion of publicly-tradable shares, Appel et al. [2020] use the Stata module `sknor`. However, even after setting a seed, the results from `sknor` are not reproducible. To facilitate reproducibility, we save one `sknor` output instance to be used throughout all 1000 repetitions. Using a fixed output of `sknor` does not affect Russell size because Russell size is generated using the Stata function `rnormal`, independently of `sknor`. To facilitate replication, we have provided the code for readers to verify that all results remain qualitatively the same if `sknor` output varies by repetition (`https://osf.io/2wgmc/files/`).

methodology is presented in table 1 and figure 3.

3.1 Univariate Comparison of Means

As a benchmark, we first estimate the univariate mean difference in simulated Russell size between Russell 1000 and Russell 2000 firms within a bandwidth of 250 based on simulated observed size rankings:

$$\textit{Russell Mktcap}_{it} = \beta_0 + \beta_1 \textit{R2000}_{it} + \varepsilon_{it} \tag{2}$$

Across 1000 repetitions, there is an unconditional mean difference in Russell size between the two Indexes that averages to almost **\$1.7 billion** with a 95% confidence interval of $[-\$1.688\text{B}, -\$1.678\text{B}]$. While Russell 2000 firms are always smaller than Russell 1000 firms within a given year, the difference in size at the threshold is expected to be negligible. It is clear that even after applying a bandwidth, a simple comparison of means does not recover the known null effect of index assignment on Russell size at the threshold; and more sophisticated analysis is necessary.

3.2 AGK Methodology

Next, we estimate the conditional difference in simulated Russell size between Russell 1000 and Russell 2000 firms using the AGK reduced-form regression:

$$\textit{Russell Mktcap}_{it} = \eta + \lambda \textit{R2000}_{it} + \sum_{n=1}^{3} \chi_n (\log(\textit{Mktcap}_{it}))^n + \sigma \log(\textit{Float}_{it}) + \delta_t + u_{it} \tag{3}$$

The IV in the reduced-form regression is *R2000*, the bandwidth remains at 250 based on simulated observed size rankings, and all variables are as defined previously.

Across 1000 repetitions, there is a conditional mean difference in Russell size between the two

Indexes that averages to slightly more than $**910 million** with a 95% confidence interval of [−$914.82M, −$905.22M]. That is, for two firms in the same year with the same *observed* size but different index assignments, the firm assigned to the Russell 2000 is substantially smaller in *Russell* size than the firm assigned to the Russell 1000. While the AGK methodology cuts the unconditional difference in Russell size by almost half, it still falsely detects a significant difference in simulated *Russell* size that by design does not exist.

3.3 Fuzzy Regression Discontinuity

We then compare the AGK methodology to the fuzzy regression discontinuity (RD) design:

$$
\begin{aligned}
\mathit{R2000}_{it} &= \beta + \lambda \mathit{Treatment}_{it} + \sum_{n=1}^{3} \varphi_n (\mathit{Rank}_{it} - 1000)^n + \mu_{it} \\
\mathit{Russell\ Mktcap}_{it} &= \alpha + \chi \widehat{\mathit{R2000}}_{it} + \sum_{n=1}^{3} \phi_n (\mathit{Rank}_{it} - 1000)^n + \varepsilon_{it}
\end{aligned} \tag{4}
$$

This specification exactly follows the Appel et al. [2020] Appendix and constrains the regression function $f(\mathit{Rank}_{it} - 1000)$ to be the same on both sides of the threshold.[3] Rank_{it} is the ranking of firm i's simulated *observed* size (*Mktcap*) in reconstitution year t, the bandwidth remains at 250 based on these rankings, and $\mathit{R2000}_{it}$ is an endogenous variable to be instrumented by

$$
\mathit{Treatment}_{it} = \begin{cases} 1 & \text{if } \mathit{Rank}_{it} > 1000 \\ 0 & \text{if } \mathit{Rank}_{it} \leq 1000 \end{cases}
$$

Whereas the AGK methodology controls for observed size and then directly compares Russell 1000 and 2000 firms, the fuzzy RD design compares firms *predicted to be in the* Russell 1000 and 2000 at the threshold (the reduced form); and then divides this comparison by the discontinuity in the probability of being assigned to the Russell 2000 at the threshold (the first stage).

[3] Wei and Young [2020] separately estimate the intercepts for the treatment and control groups by allowing the regression function $f(\mathit{Rank}_{it} - 1000)$ to differ on both sides of the threshold, as recommended by Lee and Lemieux [2010, p. 318] and Pei and Shen [2017, p. 487].

The fuzzy RD design requires a significant first stage; and across 1000 repetitions, the F-statistic for the null hypothesis that there is no first-stage jump averages to over **257** (95% C.I. $[254.44, 261.21]$), which vastly exceeds the critical value of 10 from Staiger and Stock [1997, p. 557].[4] Across the same repetitions, the discontinuity in Russell size at the threshold averages to less than $\$250,000$ (95% C.I. $[-\$9.09\text{M}; \$9.57\text{M}]$). Because Russell size around the threshold is in the **b**illions of dollars, a discontinuity in the hundreds of thousands is both statistically and economically insignificant from $0. Thus, unlike the AGK methodology, the fuzzy RD design correctly recovers the null effect of index assignment on simulated Russell size at the threshold.

3.4 Index Assignment: IV or Instrumented

To explain the difference in results between the AGK methodology and the fuzzy RD design, we first note that both are applications of instrumental variables. They differ in two ways, one minor and one major:

1. The AGK methodology controls for observed size, while the fuzzy RD design controls for observed size rankings.

2. The AGK methodology uses Russell 2000 Index assignment as an IV for passive IO, conditional on observed size; while the fuzzy RD design treats Index assignment as an endogenous variable to be instrument**ed**.

To test which of these drives the difference in results between the two methodologies, we implement a "mixed" design that combines the AGK IV *R2000* with the fuzzy RD observed

[4] Using graphical analysis, Appel et al. [2020] remark that the fuzzy RD design can have a weak first stage and thus inferences may be unreliable. However, Roberts and Whited [2013, p. 541] have cautioned that "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."

size ranking controls:

$$Russell\ Mktcap_{it} = \alpha + \lambda R2000_{it} + \sum_{n=1}^{3} \phi_n (Rank_{it} - 1000)^n + \varepsilon_{it} \tag{5}$$

Compared to the AGK methodology, the "mixed" design holds *R2000* constant and varies the controls; while compared to the fuzzy RD design, it holds the controls constant and varies whether *R2000* is to be instrumented or is an IV. To clarify, the "mixed" design is not the fuzzy RD reduced form because the "mixed" design uses *R2000* directly.

Across 1000 repetitions, after holding constant observed size rankings, the conditional difference in Russell size between Russell 1000 and 2000 firms averages to over $**1.4 billion** with a 95% confidence interval of $[-\$1.421\text{B}, -\$1.407\text{B}]$. That is, whether controlling for *observed* size or its rankings, Russell 1000 and Russell 2000 firms will substantially and systematically differ on *Russell* size. Therefore, after conditioning on observed size, *R2000* is not a valid IV; and the difference in results between the fuzzy RD design and the AGK methodology is explained by the latter's use of *R2000* as an IV.

4 Discussion

We discuss several potential objections to our conclusion that the systematic conditional difference in simulated Russell size between the Appel et al. [2016] IV groups demonstrates that the AGK IV is invalid.

4.1 Does the AGK methodology require that the conditional difference in Russell size be insignificant?

Appel et al. [2020] write,

> The identifying assumption for the AGK (2016) framework is that after condi-

> tioning on stocks' end-of-May CRSP market cap, inclusion in the Russell 2000 is associated with an increase in *Passive%* (relevance condition) but does not directly affect their outcomes of interest except through its impact on ownership by passive investors (exclusion assumption).

Does a $910 million conditional difference in simulated Russell size violate any of the identifying assumptions for the AGK framework? The answer is yes. The AGK methodology also relies on the conditional independence assumption: after conditioning on observed size, inclusion in the Russell 2000 is "as good as randomly assigned" [Angrist and Pischke, 2009, p. 176]. Conditional randomization implies that holding observed size constant, firm characteristics should be balanced between the two Indexes [Angrist and Pischke, 2009, p. 18, 55].

However, in the Monte Carlo simulation, after controlling for *observed* size, there is a $910 million average difference in *Russell* size between Russell 1000 and 2000 firms. This imbalance is both substantial and systematic. Hence, after conditioning on observed size, Russell 1000/2000 Index assignment is not "as good as randomly assigned," which invalidates the AGK methodology.

4.2 Does the systematic difference matter for the outcome?

Appel et al. [2020] write,

> . . . the AGK IV estimation assumes that the unobservable component of Russell's total market cap does not directly matter for the IV estimation's outcome of interest after robustly controlling for the observable end-of-May market cap.

That is, Appel et al. [2020] express Russell size as

$$Russell\ Mktcap_{it} = Mktcap_{it} + \xi_{it}$$

such that after robustly controlling for the observable *Mktcap*, any difference in *Russell Mktcap* is solely due to the unobservable component ξ. Does a systematic conditional difference in ξ between the AGK IV groups still matter for the outcome?

The answer is yes. If we rewrite Russell size as observable size plus the unobservable component, the threshold rule becomes

$$R2000_{it} = \begin{cases} 1 & \text{if } \mathrm{Rank}(Mktcap_{it} + \xi_{it}) > 1000 \\ 0 & \text{if } \mathrm{Rank}(Mktcap_{it} + \xi_{it}) \leq 1000 \end{cases}$$

It follows that for two firms in year t with the <u>same</u> *Mktcap* but different *R2000* assignments, the firm with $R2000 = 1$ will **always** have a <u>smaller</u> value of ξ than the firm with $R2000 = 0$. Therefore, if index assignment matters for the outcome (even indirectly through an exclusion restriction), then *the only reason* a firm is **always** assigned to one index or the other must also matter for the outcome.

4.3 Do these criticisms still apply after banding?

After 2006, Russell imposed a "banding" rule which made index assignments more persistent. To address the rule change, Appel et al. [2019, 2020] modified the AGK methodology

$$\begin{aligned} Passive\%_{it} = {} & \eta + \lambda R2000_{it} + \sum_{n=1}^{N} \chi_n (\log(Mktcap_{it}))^n + \sigma \log(Float_{it}) \\ & + \phi_1 band_{it} + \phi_2 R2000_{it-1} + \phi_3 \left(band_{it} \times R2000_{it-1}\right) + \delta_t + u_{it} \end{aligned} \tag{6}$$

where $R2000_{it}$ is still the IV for passive IO, conditional on observed size; $band_{it}$ is an indicator variable that firm i is within the reconstitution year t banding range and thus will retain its year $t-1$ index assignment; and all other variables are as defined before.

Do the criticisms in the preceding sections still apply to the modified AGK methodology for the post-banding period? The answer is yes. While the banding rule as described in Appel et al. [2019, 2020] may appear convoluted, it simply results in multiple thresholds for index assignment. To see this, consider the FTSE Russell [2020, p. 22-23] illustration:

Rank	**Previous index**	**Market cap ($M)**	**Cumulative percentile**	**New index**
995	R1	$2,115	84.38%	R1
996	R2	$2,105	85.54%	R1
997	R1	$2,100	86.69%	R1
998	*R2*	*$2,011*	*87.79%*	*R2*
999	*R2*	*$2,010*	*88.89%*	*R2*
1000	*R2*	*$2,000*	*89.99%*	*R2*
1001	*R1*	*$1,995*	*91.08%*	*R1*
1002	R2	$1,950	92.15%	R2
1003	R1	$1,923	93.20%	R2

The cumulative market capitalization percentile of the 1000$^{\text{th}}$ firm is 89.99%, and a firm whose percentile is within the banding range $89.99 \pm 2.5 = (87.49\%, 92.49\%)$ will remain in its current index. In this example, the firms ranked 998$^{\text{th}}$ through 1001$^{\text{st}}$ (italicized) are all within the banding range and thus retain their previous index assignments. Outside of the range, a firm that was in the Russell 1000 must be ranked **below 1002** to be *added to* the Russell 2000; while a firm that was in the Russell 2000 must be ranked **above 998** to be *deleted from* the Russell 2000.

The Appel et al. [2020] simulation is limited to the pre-banding period through 2006. Hence, post-banding Monte Carlo simulations cannot be conducted using their data. Nonetheless, based on the FTSE Russell [2020] example, we can rewrite the post-banding assignment rule as follows. For a firm that was in the Russell 1000 in the previous year,

$$R2000_{it} = \begin{cases} 1 & \text{if } \mathrm{Rank}(Mktcap_{it} + \xi_{it}) > 1002 \\ 0 & \text{if } \mathrm{Rank}(Mktcap_{it} + \xi_{it}) \leq 1002 \end{cases}$$

while for a firm that was in the Russell 2000 in the previous year,

$$
R2000_{it} = \begin{cases} 1 & \text{if } \mathrm{Rank}(Mktcap_{it} + \xi_{it}) > 998 \\ 0 & \text{if } \mathrm{Rank}(Mktcap_{it} + \xi_{it}) \leq 998 \end{cases}
$$

After banding, the conclusion from the previous section is fundamentally unchanged. The modified AGK methodology still relies on the conditional independence assumption, albeit also modified: after conditioning on observed size, *banded status, and prior year index assignment*, inclusion in the Russell 2000 is "as good as randomly assigned" [Angrist and Pischke, 2009, p. 176].

However, for two firms in year t with the <u>same</u> $R2000_{it-1}$ and the <u>same</u> *Mktcap* but different $R2000_{it}$ assignments, the firm with $R2000 = 1$ will **always** have a <u>smaller</u> value of ξ than the firm with $R2000 = 0$. Therefore, the problems described in the previous sections still apply in the post-banding period: Russell 1000/2000 Index assignment is not "as good as random" after controlling for observed size, banded status, and prior year index assignment.

5 Conclusion

The fundamental error in the AGK methodology is the mistaken assumption that conditional on *observed* size, Russell 2000 Index assignment is "as good as random." It is only at the Russell 1000/2000 threshold where the difference in *Russell* size is expected to be the smallest between the two Indexes such that index assignment is plausibly "as good as random." But the only way to isolate the threshold is to use rankings of Russell size, which are unobservable to researchers. As we have demonstrated through both simulation and theory, firms with the same observable size but different index assignments will substantially differ on Russell size. Therefore, the AGK methodology does not compare the appropriate treatment and control groups.

We conclude by noting that Imbens and Lemieux [2008, p. 621] have cautioned against methodologies which use reasoning similar to Appel et al. [2016, 2019, 2020]:

> Unconfoundedness is fundamentally based on units being comparable if their covariates are similar. This is not an attractive assumption in the current setting where the probability of receiving the treatment is discontinuous in the covariate. Thus, **units with similar values of the forcing variable (but on different sides of the threshold) must be different in some important way related to the receipt of treatment**. Unless there is a substantive argument that this difference is immaterial for the comparison of the outcomes of interest, an analysis based on unconfoundedness is not attractive.

Applying their warning to this setting, AGK assume that Russell 1000 and Russell 2000 firms are comparable if their observed size is similar. However, Russell 1000/2000 Index assignment (treatment receipt) is a discontinuous and deterministic function of the Russell size rankings; and our Monte Carlo simulation shows that firms with similar values of *observed* size (but on different sides of the threshold) differ on *Russell* size by over $910 million. Because a ranking of Russell size exclusively determines index assignment, and AGK argue that index assignment affects their outcomes via an exclusion restriction, the $910 million conditional difference in Russell size is material for the comparison of their outcomes of interest. Therefore, an analysis based on the AGK IV estimation is not attractive.

Despite the flaws of some existing approaches, the Russell 1000/2000 reconstitution setting (both before and after banding) remains promising for future research. However, the analysis must be done more carefully based on econometric principles.

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, 2020. URL `https://cfr.pub/forthcoming/appel2020identification.pdf`. Critical Finance Review, forthcoming.

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

I. R. Appel, T. A. Gormley, and D. B. Keim. Standing on the Shoulders of Giants: The Effect of Passive Investors on Activism. *The Review of Financial Studies*, 32(7):2720–2774, 2019.

FTSE Russell. Russell U.S. Equity Indexes: Construction and Methodology, 2020. URL `https://research.ftserussell.com/products/downloads/Russell-US-indexes.pdf`. Version 4.9.

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

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

Z. Pei and Y. Shen. *The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable*. Emerald Publishing Limited, 2017.

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.

D. Staiger and J. H. Stock. Instrumental Variables Regression with Weak Instruments . *Econometrica*, 65(3):557–586, 1997.

W. Wei and A. Young. Selection Bias or Treatment Effect? A Re-Examination of Russell 1000/2000 Index Reconstitution, 2020. URL `https://cfr.ivo-welch.info/forthcoming/wei2020selection.pdf`. Critical Finance Review, forthcoming.

	Mean	95% C.I. Lower	95% C.I. Upper
Univariate Mean	-$1683.05	-$1687.87	-$1678.22
AGK IV	-$910.02	-$914.82	-$905.22
Fuzzy RD	$0.24	-$9.09	$9.57
Fuzzy RD First Stage F-statistic	257.82	254.44	261.21
Mixed Design	-$1413.84	-$1420.97	-$1406.71
Repetitions	1000		

Table 1: **Description:** The table presents the average difference in Russell size (in millions) between the Russell 1000 and 2000 by methodology within a bandwidth of 250, the average fuzzy RD first-stage F-statistic, and the 95% confidence interval bounds across 1000 repetitions in the Monte Carlo simulation. *Univariate Mean* compares the unconditional average of Russell size between the Indexes. *AGK IV* holds observed size constant and then compares the average of Russell size between the Indexes. *Fuzzy RD* estimates the discontinuity in Russell size at the predicted threshold between the Indexes, and then scales this by the jump in the probability of being assigned to the Russell 2000 at the predicted threshold between the Indexes. *Fuzzy RD First Stage F-statistic* tests the null hypothesis that there is no first-stage jump in the probability of being assigned to the Russell 2000 at the predicted threshold between the Indexes. *Mixed Design* holds observed size rankings constant and then compares the average of Russell size between the Indexes.

Interpretation: The AGK methodology uses Russell 2000 Index assignment as an IV and produces a substantial-but-spurious conditional difference of over $910 million in Russell size between the Indexes. On the other hand, the fuzzy RD design treats index assignment as an endogenous variable to be instrumented and correctly recovers the null effect of index assignment on Russell size at the threshold.

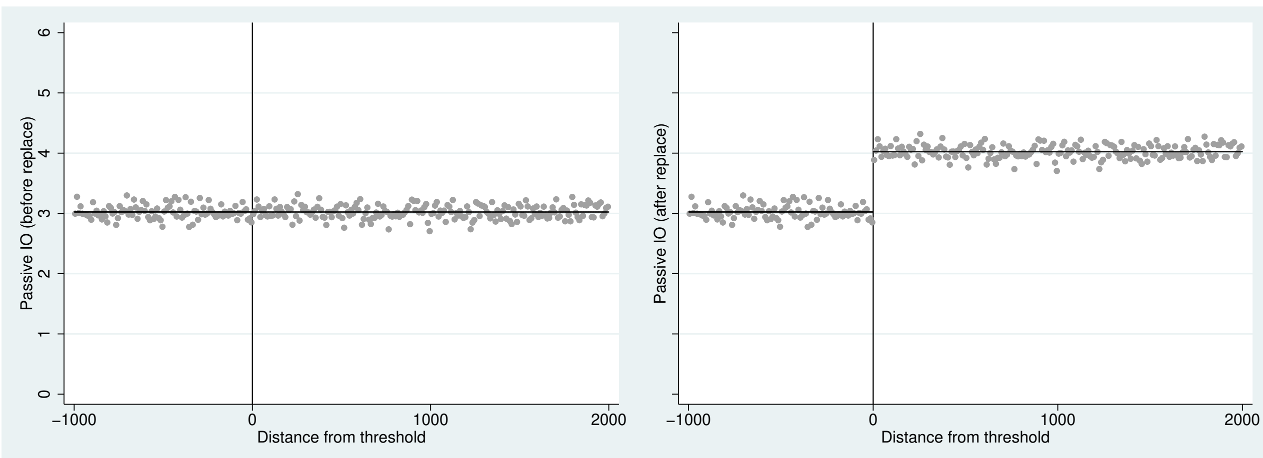

Figure 1: **Description:** The left (right) panel plots simulated passive institutional ownership in percentage points against rankings of simulated Russell size centered at the 1000^{th} rank, before (after) one percentage point is added to all observations in the Russell 2000. Each bin represents the average of the y-axis variable over 10 ranks throughout the simulated sample period of 1998–2006.

Interpretation: In the Appel et al. [2020] simulation, even a univariate comparison of means will recover the effects of index assignment on both total and passive IO. If no control variables are needed to detect those effects, then these two dependent variables are not suitable for substantiating the AGK claim that Russell 2000 Index assignment is a valid IV after *controlling* for observed size.

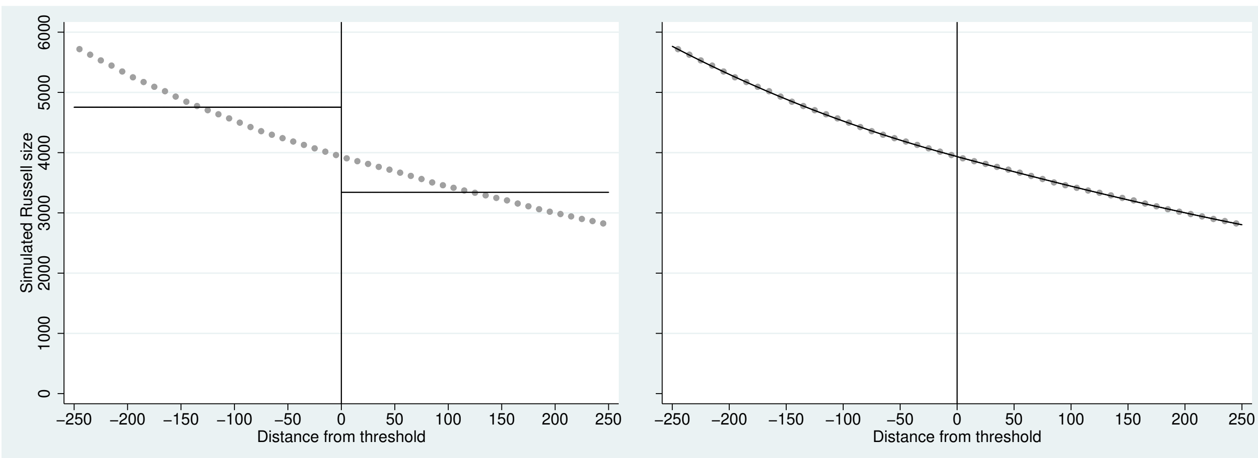

Figure 2: **Description:** The figure plots simulated Russell size (in millions) against its rankings centered at the 1000$^{\text{th}}$ rank. The bandwidth is 250. The left panel estimates the discontinuity at the threshold by taking the simple average on either side of the threshold, while the right panel estimates the discontinuity as the difference between the intercepts of two separate regressions, one for each side of the threshold. Each bin represents the average of the y-axis variable over 10 ranks throughout the simulated sample period of 1998–2006.

Interpretation: By construction, Russell size is negatively sloped across its descending-order ranking and is continuous at the threshold between the Russell 1000 and 2000. However, due to the steepness of the slope, a simple approach such as a univariate comparison of means will falsely detect a significant effect at the threshold; while a valid methodology will correctly verify that the difference in Russell size at the threshold is immaterial.

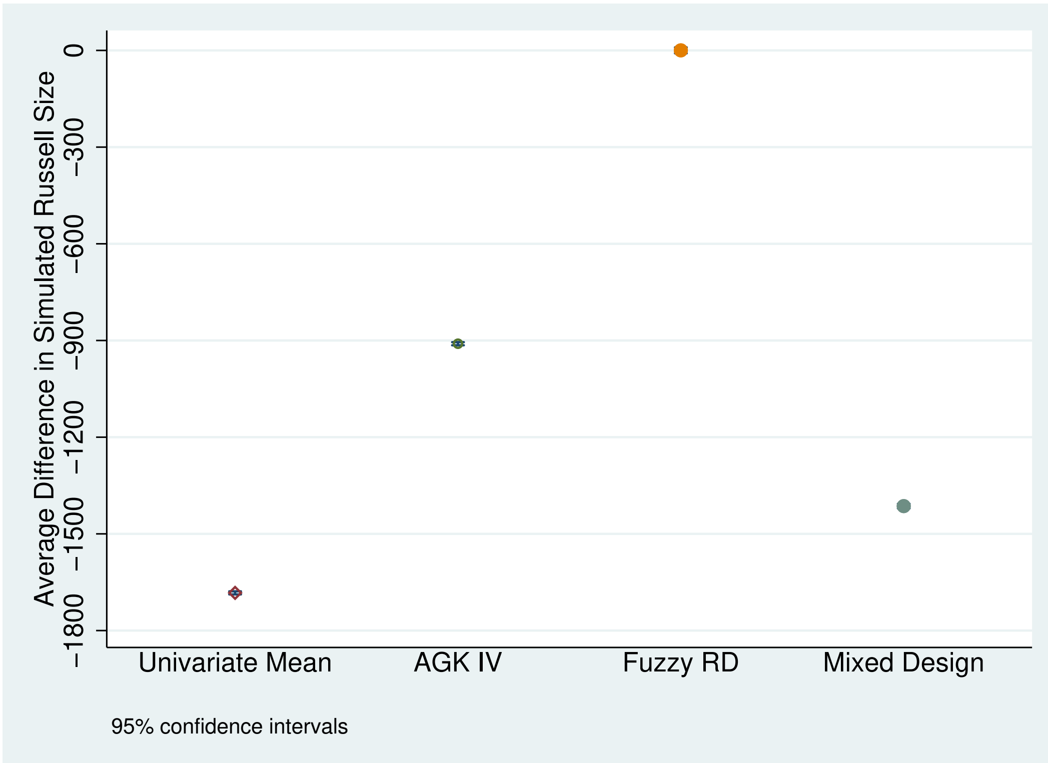

Figure 3: **Description:** The figure is a visual representation of table 1 and plots the average difference in Russell size (in millions) between the Russell 1000 and 2000 by methodology within a bandwidth of 250 and the corresponding 95% confidence interval bounds across 1000 repetitions in the Monte Carlo simulation. *Univariate Mean* compares the unconditional average of Russell size between the Indexes. *AGK IV* holds observed size constant and then compares the average of Russell size between the Indexes. *Fuzzy RD* estimates the discontinuity in Russell size at the predicted threshold between the Indexes, and then scales this by the jump in the probability of being assigned to the Russell 2000 at the predicted threshold between the Indexes. *Mixed Design* holds observed size rankings constant and then compares the average of Russell size between the Indexes.

Interpretation: The AGK methodology uses Russell 2000 Index assignment as an IV and produces a substantial-but-spurious conditional difference of over $910 million in Russell size between the Indexes. On the other hand, the fuzzy RD design treats index assignment as an endogenous variable to be instrumented and correctly recovers the null effect of index assignment on Russell size at the threshold.