



# Will the Real Specification Please Stand Up? A Comment on Andrew Bird and Stephen Karolyi

Alex Young<sup>1</sup>

[LINK TO ABSTRACT](#)

The present paper concerns two versions of an article by Andrew Bird and Stephen Karolyi (henceforth BK) titled “Governance and Taxes: Evidence from Regression Discontinuity.” One version is a manuscript working paper, dated September 2015 (BK 2015); the other is the piece as published in *The Accounting Review* (BK 2017). Between the two versions, the numbers in all 11 tables stayed exactly the same (to three decimal places), so it is clear that the specifications generating the results were the same. But the description of the specifications changed. The 2015 working paper version says that, for their main results, the authors used one method of ranking firms by market capitalization, while the 2017 published version says that they used another method.

Here I respectfully press the issue of whether the 2017 description in *The Accounting Review* is accurate. If I were able to see, without any problem, that the 2017 description simply corrects some confusion in the earlier version, I would be glad of the resolution. Unfortunately, I see problems: Replication along the lines of the 2015 description closely replicates the paper’s results, whereas replication along the lines of the 2017 published description does not.

To provide some background, I recount some personal experience. In October 2015, I attended the UNC/Duke Fall Camp conference for accounting at Chapel Hill, North Carolina, where Professor Bird presented the 2015 version

---

1. North Dakota State University, Fargo, ND 58108. I thank Wei Wei for helpful comments and Sergio Correia, Scott Dyreng, and Bill Mayew for helpful comments in an earlier version. All remaining errors are my own.

of the paper. I approached Professor Bird and raised some concerns about the specification used in the paper. In January 2017, the paper was published in *The Accounting Review*, and I noticed that the description of the specifications had changed from the 2015 version, while the numbers stayed the same in all 11 tables. On 18 April 2017 I emailed Professor Bird, reminding him of our conversation in October 2015, and requested that they share their data and code. On 20 April 2017 Professor Karolyi responded with a brief message that ignored my request. On 5 May 2017 I explained my concerns in an email to Professor Edward Maydew (the editor who was assigned to the BK paper at *The Accounting Review*). On 9 May 2017 Professor Maydew emailed me to say that he had forwarded my message to the journal's Senior Editor. After waiting for six months with no reply, I contacted *Econ Journal Watch* on 9 November 2017. On 15 November 2017 the *EJW* chief editor, Daniel Klein, emailed Professors Bird and Karolyi requesting that they share their data and code. The next day they responded, without addressing the request. On 21 November 2017, Klein replied and repeated the request; later that day they replied, declining the request. Meanwhile, after my also having sent in November 2017 follow-up emails to Professor Mary E. Barth, the current Senior Editor of *The Accounting Review*, Professor Barth emailed me on 1 December 2017 to say that *The Accounting Review* was commencing its process for the investigation of possible misconduct.

I would be pleased to discover that there was no misconduct, that my concern was misplaced. If Professors Bird and Karolyi make public the data and code, everything should become clear. I press the issue here from a sense of duty to scholarship in general and to the accounting and finance research communities in particular.

In this article, I proceed in the following steps: First, I explain the two different ways of ranking firms by market capitalization. Next, I turn to the textual descriptions of the specifications in the two versions for the BK paper. Next, I turn to replicating, trying to follow from scratch the descriptions provided. I attempt to replicate BK's results by following the 2017 version's description, but that specification delivers a statistically insignificant null result—whereas when I follow the 2015 description, I am able to replicate the main results very closely. Additional analysis further supports the conclusion that the 2015 version describes the main analysis accurately, while the published version misstates it. I conclude by demonstrating that the main result, when understood on the 2015 description, is spurious: When I conduct a placebo test by lagging the dependent variable by one year, holding all else equal, I find that the 'effect' was already present *before* the 'cause.' On the 2015 description, therefore, the main result is a manifestation of selection bias and has no causal interpretation. That problem is what I brought to Professor Bird's attention when I approached him at the conference at Chapel Hill

in 2015.

## Background regarding the two methods of ranking

The Russell 1000 and Russell 2000 are U.S. stock market indices under the purview of the company now called FTSE Russell but which I will here refer to as simply ‘Russell.’ The Russell 1000 contains the one thousand largest US stocks as of the end of May in each year, while the Russell 2000 contains the *next* largest two thousand U.S. stocks. ‘Size’ in this context refers to market capitalization.

The appeal of the Russell 1000 and 2000 indices in accounting and finance research is that the rules-based construction of the indices seemingly lends itself to analysis with a regression-discontinuity design. Intuitively, we would not expect significant differences on average between a firm ranked 999 (assigned to the Russell 1000) and a firm ranked 1001 (assigned to the Russell 2000). However, within each index, firms are weighted by their market capitalization relative to the total market capitalization of that index. Since the firm ranked 1001 is the largest in the Russell 2000, it will have a much larger index weight than the firm ranked 999. Small differences in market capitalization at the end of May thus potentially affect capital markets.

Russell uses a proprietary measure of market capitalization to rank all three thousand stocks at the end of May each year, a measure that, alas, is not available to researchers (Chang et al. 2015, 217). As a result, it is not possible to know *exactly* which firm was 999 and which was 1001 at the end of May in each year. Instead, Russell provides index weights at the end of June to researchers. Researchers must then attempt either to reproduce Russell’s end-of-May rankings using publicly available data, or to reverse-engineer the proprietary end-of-May rankings using the Russell-provided June index weights.

For the first method, the most commonly used database is CRSP (the Center for Research in Security Prices). Price and shares outstanding, each adjusted by cumulative factors, can be multiplied together to determine market capitalization, and then all three thousand stocks in the Russell 1000 and 2000 can be ranked on market capitalization. These CRSP rankings (supposedly the one used in the main specification of BK 2017) will not perfectly reproduce the actual Russell 1000 and Russell 2000 index constituents (Chang et al. 2015, 230).

Alternatively, Russell provides researchers with the Russell 1000 and 2000 index weights measured at the end of June. A stock’s weight is the proportion of the stock’s adjusted market capitalization to its actual index’s total adjusted market

capitalization. Note the word ‘adjusted’: the weights are calculated from a *different* market capitalization than the proprietary one Russell uses to rank firms at the end of May. Specifically, the adjustments remove shares that Russell considers to be unavailable for purchase by the public.

The Russell-provided index weights can be used to rank the Russell 1000 and 2000 stocks. Since each index is separately value weighted, the weights are ranked separately by index and then combined such that the largest stock in the Russell 2000 is ranked 1001 overall out of 3000. A problem with the Russell-provided June rankings is that because of stock market activity in the month of June, these rankings do not necessarily reverse-engineer the proprietary rankings Russell uses at the end of May each year (Chang et al. 2015, 222). This is the ranking method that is said to be used, and which I believe was used, in BK 2015.

## Textual descriptions: BK 2015 versus BK 2017

In BK 2015, the authors began their data description as follows (in block quotations the boldfacing is added by me):

Two approaches have been used in the literature to identify these index membership lists. **First**, because the Russell 1000 and 2000 indexes are explicitly determined by market capitalization rank as of the last trading day in May each year, the index memberships can be constructed **using CRSP data**. **Second, Russell provides** data for academic use. Because Crane et al. (2014) argue that predicting index inclusion for a fixed sample size can induce a bias from index misclassifications, **we use the latter approach in our main results**. We use index membership lists generated from CRSP (for the subset of reconstitutions covered by the Russell-provided data) in robustness checks of our main results. (BK 2015, 8)

In BK 2017, the authors stated that they made the opposite choice for their main specification:

Two approaches have been used in the literature to identify these index membership lists. **First**, because the Russell 1000 and 2000 indexes are explicitly determined by market capitalization rank as of the last trading day in May each year, the index memberships can be constructed **using CRSP data; we use these rankings in our main specifications**. **Second, Russell provides** index membership data for academic use, and we use these in **robustness checks** of our main findings. (BK 2017, 32)

A few pages later, a similar change was made. In BK 2015 appears the following:

It is important that we acknowledge the methodological differences among the other papers that exploit the Russell index reconstitution setting for identification. **In the results presented here** we follow Crane et al. (2014) and impose a  $k$ -order polynomial control function on the market capitalization **ranks provided and used by Russell**, but, in results reported in Table 7, we confirm that our main findings hold if we **alternatively** impose a  $k$ -order polynomial control function on the market capitalization ranks that we calculate **using May 31 closing prices from CRSP** as in Appel et al. (2015). (BK 2015, 11)<sup>2</sup>

While in BK 2017:

It is important that we acknowledge the methodological differences among the other papers that exploit the Russell index reconstitution setting for identification. **In the results presented here** we follow Crane et al. (2016) and impose a  $k$ -order polynomial control function on the market capitalization ranks **based on May 31 closing prices from CRSP**, but, in results reported in Table 7, we confirm that our main findings hold if we **alternatively** impose a  $k$ -order polynomial control function on **Russell index ranks, which include the effects of Russell’s proprietary float adjustment**. (BK 2017, 34)

To summarize, BK made the following choices for the older 2015 version and then reversed them for the 2017 published version:

	BK 2015	BK 2017
Main result rankings	Russell-provided	CRSP-calculated
Alternative/robustness check rankings	CRSP-calculated	Russell-provided

Thus, if we compare the tables from the two versions, at the very least, we should see slightly different numbers for the estimated coefficients, standard errors, etc. But row for row, column for column, the numbers for the estimated coefficients, standard errors, and R-squareds are all exactly the same between BK 2015 and BK 2017 to three decimal places in all 11 tables.<sup>3</sup>

The exact sameness cannot be explained by claiming that the Russell-provided and CRSP-constructed rankings are exactly the same. As shown in table 7 of both versions, regardless of which ranking is claimed as “baseline” or “alterna-

2. The “Crane et al. (2014)” cited by BK (2015) is a working-paper version of Crane et al. (2016), and the “Appel et al. (2015)” is a working-paper version of Appel et al. (2016).

3. The only exception is that column 4 in table 10 of BK 2015 version was not included in table 10 of BK 2017. However, the remaining columns in table 10 between the two versions all still have exactly the same numbers for the coefficient estimates, standard errors, and R-squareds. Therefore, the same specification was used to produce table 10 in both versions.

tive,” the coefficient estimates are different ( $-0.205$  vs.  $-0.177$ ). Furthermore, the identical numbers in both versions of table 7 cannot be explained as a typo because the remaining regression tables also all report identical numbers in both versions. Hence, one of the two versions of the paper does not describe the main specification accurately.

## Replication

In this section, I attempt to determine which ranking was used to produce the main results in BK through replication. Following Colin Camerer et al. (2016, 1433–1434), I choose “the most important statistically significant finding, as emphasized by the authors” for replication. As BK state,

To start, we investigate tax avoidance using cash effective tax rates because, at least in the long run, what should concern shareholders is the cash flow associated with the firm rather than the reported book income. ... for comparability with other studies and to better understand the objectives of institutional investors, we also investigate the effect of index reconstitutions on book effective tax rates. (BK 2017, 37)

I accordingly attempt to replicate the results with cash and book (GAAP) effective tax rates (ETRs). The text underneath my Tables 1 and 2 describe the sample period, variables, and empirical specifications, which all follow BK.

Using the Russell-provided ranks, per BK 2015, I am able to closely reproduce the main results of BK both in terms of magnitude and statistical significance. Column 1 of Table 1 shows that with cash effective tax rates, the coefficient estimate on instrumented institutional ownership is  $-0.267$  (significant at the 1 percent level). That result is consistent with BK, who state in the published article that “for a 1 percentage point increase in institutional ownership due to inclusion in the Russell 2000 around the threshold, a firm’s cash effective tax rate falls by between 0.2 and 0.3 percentage points” (BK 2017, 38).<sup>4</sup>

Similarly, column 2 of Table 1 shows that with GAAP effective tax rates, the coefficient estimate on instrumented institutional ownership is  $-0.159$  (significant at the 1 percent level). That too is consistent with BK, who state that “an increase in institutional ownership of 1 percentage point results in a decrease of GAAP effective tax rates of 0.16 percentage points” (2017, 38; 2015, 17). I reiterate that

---

4. An all-but-identical statement appears in the working paper: “for a one percentage point increase in institutional ownership caused by inclusion in the Russell 2000, a firm’s cash effective tax rate falls by between 0.2 and 0.3 percentage points” (BK 2015, 16).

by using the Russell-provided ranks, following BK 2015, I am able to very closely reproduce the main results of BK.

TABLE 1.

	(1)	(2)	(3)	(4)
	Cash ETR	GAAP ETR	Lag Cash ETR	Lag GAAP ETR
IO	-0.267*** (0.069)	-0.159*** (0.046)	-0.252*** (0.072)	-0.229*** (0.051)
Constant	0.407*** (0.041)	0.424*** (0.027)	0.401*** (0.042)	0.454*** (0.030)
Year Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	24,404	24,404	24,718	24,718

*Notes:* This table reports the results of the estimating the following with 2SLS. The outcome variables are (lagged) cash and GAAP effective tax rates (ETRs) in columns 1 and 2 (3 and 4), defined as cash taxes paid (total income tax expense) divided by pre-tax book income minus special items. IO is the total institutional ownership of firm  $i$  in September of year  $t$ ,  $D$  is an indicator equal to 1 (0) if firm  $i$  in year  $t$  is a member of the Russell 2000 (1000), and  $R$  is the rank of stock  $i$  in year  $t$  relative to the smallest firm in the Russell 1000 based on the Russell-provided June rank. Standard errors (in parentheses) are clustered by firm. The sample period is 1996–2006.  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

$$IO_{i,t} = \alpha + \tau D_{i,t} + \delta_j R_{i,t} + \gamma_j D_{i,t} R_{i,t} + v_t + \xi_{i,t}$$

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{IO}_{i,t} + \varphi_j R_{i,t} + \lambda_j D_{i,t} R_{i,t} + v_t + \varepsilon_{i,t}$$

By contrast, my analysis indicates that the main results of BK do not hold with the CRSP-constructed ranks, which BK 2017 claims to use in the main specification. In columns 1 and 2 of Table 2, the standard errors for the IO variable are very large relative to the coefficient estimates, and so the results are not statistically significant at conventional levels. For example, the 95% confidence interval for the coefficient on instrumented institutional ownership in column 1 is (-1.07, +0.31). I stress that the only difference in producing Table 2 from Table 1 was the choice of ranking data.

Michael Clemens has argued persuasively that “the author of a replication test bears the burden of showing that its estimates come from materially the same sampling distribution” (2017, 327). I have shown in Table 1 that I am able to closely reproduce the main results of BK by following the empirical specification described in BK 2015 version with Russell-provided ranks. Since the only difference in Tables 1 and 2 is the ranking, I therefore need to show that the CRSP-constructed ranks I use in Table 2 exhibit properties that they are expected to exhibit.

As the CRSP-constructed ranks are a rank variable of market capitalization at the end of May from CRSP, by construction, a graph of market capitalization on the  $y$ -axis against its ranking on the  $x$ -axis should be continuous and monotonically

decreasing (rank 1 is the largest). Figure 1 shows that the CRSP-constructed ranks exhibit this property.

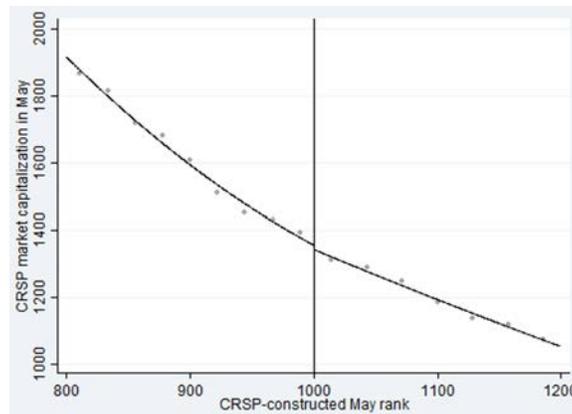
TABLE 2.

	(1)	(2)	(3)	(4)
	Cash ETR	GAAP ETR	Lag Cash ETR	Lag GAAP ETR
IO	-0.382 (0.353)	-0.324 (0.248)	-0.341 (0.300)	-0.501* (0.257)
Constant	0.471** (0.199)	0.517*** (0.140)	0.450*** (0.166)	0.602*** (0.142)
Year Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	24,404	24,404	24,718	24,718

*Notes:* This table reports the results of the estimating the following with 2SLS. The outcome variables are (lagged) cash and GAAP effective tax rates (ETRs) in columns 1 and 2 (3 and 4), defined as cash taxes paid (total income tax expense) divided by pre-tax book income minus special items. IO is the total institutional ownership of firm  $i$  in September of year  $t$ ,  $D$  is an indicator equal to 1 (0) if firm  $i$  in year  $t$  is a member of the Russell 2000 (1000), and  $R$  is the rank of stock  $i$  in year  $t$  relative to the smallest firm in the Russell 1000 based on the researcher-constructed rankings of CRSP market capitalization in May. Standard errors (in parentheses) are clustered by firm. The sample period is 1996–2006. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

$$IO_{i,t} = \alpha + \tau D_{i,t} + \delta_j R_{i,t} + \gamma_j D_{i,t} R_{i,t} + v_t + \xi_{i,t}$$

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{IO}_{i,t} + \varphi_j R_{i,t} + \lambda_j D_{i,t} R_{i,t} + v_t + \varepsilon_{i,t}$$



**Figure 1.** The figure plots market capitalization at the end of May from CRSP against the researcher-constructed ranking of May market capitalization from CRSP, for firms in the 800 to 1200 ranks. The sample period is 1996–2006. Only firms within 200 ranks of the 1000 cutoff by researcher-constructed CRSP rankings are included (to the left/right). The dots (or ‘bins’) represent averages of the  $y$ -axis variable over various ranges of the  $x$ -axis variable throughout the sample period. The black curves are quadratic polynomial estimates for how the  $y$ -axis variable varies at the cutoff with the  $x$ -axis variable.

Hence, I submit that I have met the standard in Clemens (2017, 327) to bear the burden of showing that my estimates come from “materially the same sampling distribution,” and I conclude that the choice of ranking matters: with the CRSP-constructed ranks, as claimed in BK 2017, the main result is a null result. Only by following the BK 2015 description am I able to replicate the main results.

## **Additional evidence as to which specification yielded BK’s main results**

My replication suggests that BK’s main specification is accurately described in BK 2015 and misstated in BK 2017. Nonetheless, it is possible that I have made a mistake, or that I have misunderstood BK.

Additional evidence that BK 2015 accurately describes the main specification can be found in two other, related papers. Shuping Chen and coauthors (2018) and Mozaffar Khan, Suraj Srinivasan, and Liang Tan (2017) also use the Russell 1000/2000 setting in regression discontinuity designs to investigate the same question as BK: How does institutional ownership affect tax avoidance? However, unlike BK 2017, both Chen et al. (2018) and Khan et al. (2017) report that they use the Russell-provided rankings in their main specifications: The former write that they “define *Rank* using the actual index assignment based on June weights provided by Russell Investments” (Chen et al. 2018, 12), while the latter define the relative distance (i.e., the ranking centered at 1000) “of a Russell 1000 firm to the cutoff firm (the 1,000th firm) each year based on June weights” (Khan et al. 2017, 120). While BK 2017 do not use the term “June weights,” they clarify in their companion paper (Bird and Karolyi 2016, 3250) that the Russell-provided data is as of the “end of June.” That is, their Russell-provided rankings are consistent with those used by Chen et al. (2018) and Khan et al. (2017).

Given that:

- two contemporaneous papers use Russell-provided rankings (Chen et al. 2018, Khan et al. 2017),
- BK 2015 claimed to use Russell-provided rankings in the main specification,
- the numbers in all 11 tables of BK 2015 are identical to those of BK 2017, which instead claims to use CRSP-constructed rankings in the main specification, and
- Chen et al. (2018), Khan et al. (2017), and BK 2017 all conclude that an increase in institutional ownership increases tax avoidance,

the simplest explanation for the irregularity of identical numbers yet different specifications between BK 2015 and BK 2017 is the conclusion from my replication: BK 2015 accurately describes the research design, while BK 2017 misstates it.

## The main result is spurious

Although my replication suggests that there is a misstatement in BK 2017, the results with Russell-provided rankings nevertheless *seem* very robust: Chen et al. (2018) and Khan et al. (2017) also use Russell-provided rankings, find similar results, and reach similar conclusions as BK. Isn't there a finding here that advances the literature on governance and tax avoidance?

Now I demonstrate that the main result in BK's work (on my conclusion that it is accurately described with Russell-provided rankings) is spurious. Columns 3 and 4 of Table 1 report the results from a placebo test with *lagged* effective tax rates as the pseudo-'outcome.' Although BK do not perform this test in their paper, they cite David Lee and Thomas Lemieux (2010), who state the intuition behind this test:

Intuitively, if the [regression discontinuity] design is valid, we know that the treatment variable cannot influence variables determined prior to the realization of the assignment variable and treatment assignment; if we observe it does, something is wrong in the design. (Lee and Lemieux 2010, 330)

That is, under this placebo test, there should be no evidence that *this year's* instrumented institutional ownership affects *last year's* tax planning. But the coefficients on instrumented institutional ownership remain negative and statistically significant at the one-percent level in columns 3 and 4 of Table 1. For example, column 3 suggests that a one percentage-point increase in instrumented institutional ownership 'causes' a decrease in *last year's* cash effective tax rates by 0.252 percentage points.<sup>5</sup>

It would be a stretch to think that institutional ownership this year can affect tax planning decisions made in the previous year. Since the only change made for the placebo test is lagging the ETRs, to re-quote Lee and Lemieux, "something is wrong in the design."<sup>6</sup> But what is wrong?

---

5. By contrast, when CRSP-constructed rankings are used in my Table 2, this placebo test passes, as seen in columns 3 and 4. However, as stated previously, with CRSP-constructed rankings, the main result is a *null result*: the results in columns 1 and 2 are very imprecisely estimated with large standard errors.

6. In email correspondence (available upon request), Professor Karolyi argued that a failure of this placebo test does not necessarily imply that the main result of BK is spurious partially because of measurement

Recall that the appeal of the Russell 1000/2000 setting in accounting and finance research is that the Russell 1000 and 2000 indices are constructed using a cutoff or threshold rule (Chang et al. 2015, 217):

- if a firm's market capitalization at the end of May is ranked 1 (largest) through 1000, it is assigned to the Russell 1000;
- if a firm's market capitalization at the end of May is ranked 1001 through 3000, it is assigned to the Russell 2000.

Thus, around the cutoff separating the Russell 1000 from the Russell 2000, membership in the one rather than the other is a close call (Chang et al. 2015, 218), and that would suggest that firms at the bottom of the Russell 1000 are a natural control group for the firms at the top of the Russell 2000.

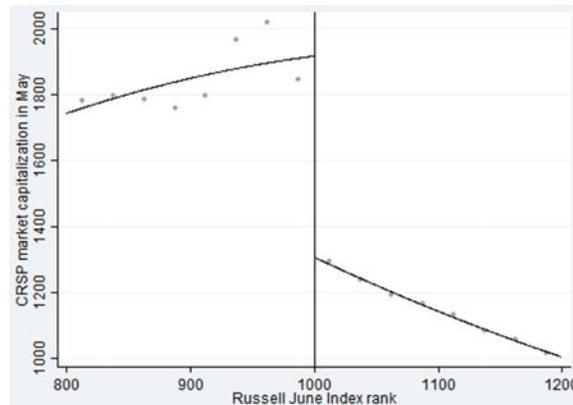
As described previously, Figure 1 shows that the CRSP end-of-May market capitalization on the  $y$ -axis is smoothly and continuously decreasing with its rankings, as we would expect. That is, firms on the left of the cutoff *could plausibly have become* Russell 2000 firms (on the right of the cutoff) and hence *are an appropriate* control group.

Figure 2 replaces the  $x$ -axis with the Russell-provided June rankings, all else being unchanged. It is clear that CRSP May market capitalization is *discontinuous* at the cutoff with the Russell-provided June rankings. That is, with Russell-provided rankings, the firms on the left of the cutoff were *far too large* to have ever become Russell 2000 firms (on the right of the cutoff); hence, the firms on the left are *not an* appropriate control group for the firms on the right.

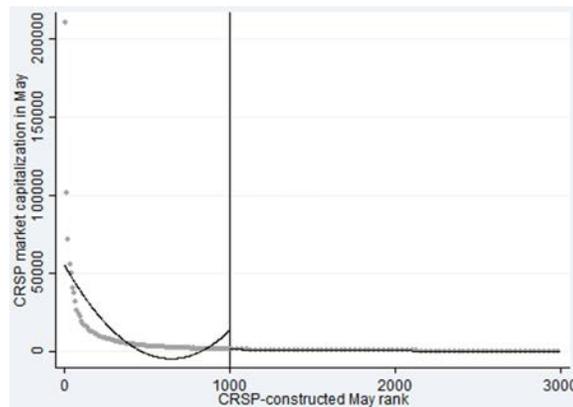
Note that both Figures 1 and 2 are 'zoomed in' at the Russell 1000/2000 cutoff and only include firms within 200 ranks on either side of the vertical line. If CRSP market capitalization in May were instead graphed against either ranking across *all three thousand* firms in the Russell 1000 and 2000, as shown in figures 3 and 4, we would not be able to see the discontinuity in market capitalization at the cutoff. The largest firms in the Russell 1000 (on the leftmost side) distort the  $y$ -axis scale in Figures 3 and 4 such that the two figures are almost indistinguishable, and market capitalization erroneously appears to be continuously decreasing. Only by 'zooming in' do we see that with the Russell-provided June rankings, the firms just to the left of the 1000/2000 cutoff were far too large to have ever been on the right of the cutoff.

---

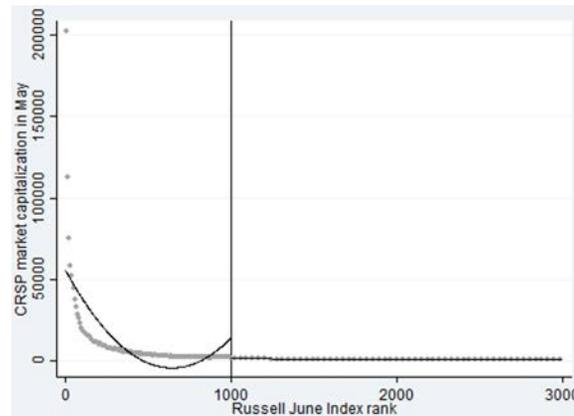
error in fiscal quarter end dates. However, the placebo test lags the effective tax rates (ETRs), which are based on annual Compustat data (BK 2017, 34). The placebo test *does not* lag the quarterly institutional ownership measures. Therefore, measurement error in fiscal quarter end dates cannot explain why the placebo test fails with lagged annual ETRs.



**Figure 2.** The figure plots market capitalization at the end of May from CRSP against the Russell-provided (end-of-June) rankings, for firms in the 800 to 1200 ranks. The sample period is 1996–2006. Only firms within 200 ranks of the 1000 cutoff by Russell-provided rankings are included (to the left/right). The dots (or ‘bins’) represent averages of the  $y$ -axis variable over various ranges of the  $x$ -axis variable throughout the sample period. The black curves are quadratic polynomial estimates for how the  $y$ -axis variable varies at the cutoff with the  $x$ -axis variable.



**Figure 3.** The figure plots market capitalization at the end of May from CRSP against the researcher-constructed ranking of May market capitalization from CRSP. The sample period is 1996–2006. The dots (or ‘bins’) represent averages of the  $y$ -axis variable over various ranges of the  $x$ -axis variable throughout the sample period. The black curves are quadratic polynomial estimates for how the  $y$ -axis variable varies at the cutoff with the  $x$ -axis variable.



**Figure 4.** The figure plots market capitalization at the end of May from CRSP against the Russell-provided (end-of-June) rankings. The sample period is 1996–2006. The dots (or ‘bins’) represent averages of the  $y$ -axis variable over various ranges of the  $x$ -axis variable throughout the sample period. The black curves are quadratic polynomial estimates for how the  $y$ -axis variable varies at the cutoff with the  $x$ -axis variable.

Thus, by using the Russell-provided June rankings, the empirical analysis is contaminated by selection bias: the firms on either side of the Russell 1000/2000 cutoff were *ex ante* already different from each other. Hence, not only did BK misstate their main analysis in BK 2017, but when the main specification is described accurately with Russell-provided rankings, the main result is also spurious: the ‘effect’ was already present *before* the ‘cause.’<sup>7</sup>

## Appendix

Data and code used in this research can be downloaded from the journal website ([link](#)) or from a page that I maintain ([link](#)).

## References

- Appel, Ian R., Todd A. Gormley, and Donald B. Keim.** 2016. Passive Investors, Not Passive Owners. *Journal of Financial Economics* 121(1): 111–141.
- Bird, Andrew, and Stephen A. Karolyi (BK).** 2015. Governance and Taxes: Evidence from Regression Discontinuity. Working paper. [Link](#)
- Bird, Andrew, and Stephen A. Karolyi (BK).** 2016. Do Institutional Investors Demand

7. Wei and Young (2017) document and explain the selection bias problem more generally for the literature that uses the Russell-provided June rankings.

Public Disclosure? *Review of Financial Studies* 29(12): 3245–3277.

**Bird, Andrew, and Stephen A. Karolyi** (BK). 2017. Governance and Taxes: Evidence from Regression Discontinuity. *The Accounting Review* 92(1): 29–50.

**Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu.** 2016. Evaluating Replicability of Laboratory Experiments in Economics. *Science* 351(6280): 1433–1436.

**Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich.** 2015. Regression Discontinuity and the Price Effects of Stock Market Indexing. *Review of Financial Studies* 28(1): 212–246.

**Chen, Shuping, Ying Huang, Ningzhong Li, and Terry J. Shevlin.** 2018 (forthcoming). How Does Quasi-Indexer Ownership Affect Corporate Tax-Planning? *Journal of Accounting and Economics*. [Link](#)

**Clemens, Michael A.** 2017. The Meaning of Failed Replications: A Review and Proposal. *Journal of Economic Surveys* 31(1): 326–342.

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

**Khan, Mozaffar, Suraj Srinivasan, and Liang Tan.** 2017. Institutional Ownership and Corporate Tax Avoidance: New Evidence. *The Accounting Review* 92(2): 101–122.

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

**Wei, Wei, and Alex Young.** 2017. Selection Bias or Treatment Effect? A Re-Examination of Russell 1000/2000 Index Reconstitution. Working paper. [Link](#)

## About the Author



**Alex Young** is an assistant professor of accounting at North Dakota State University. His research explores how corporate governance affects earnings attributes. His email address is [alex.y.young@gmail.com](mailto:alex.y.young@gmail.com).

[Andrew Bird and Stephen Karolyi's reply to this article](#)

[Go to archive of Comments section](#)

[Go to January 2018 issue](#)



Discuss this article at JournalTalk:  
<http://journaltalk.net/articles/5957>