**University of Pennsylvania**
**ScholarlyCommons**

Finance Papers

Wharton Faculty Research

3-29-2016

# Identification Using Russell 1000/2000 Index Assignments: A Discussion of Methodologies

Ian R. Appel

Todd A. Gormley

Donald B. Keim
*University of Pennsylvania*

Follow this and additional works at: http://repository.upenn.edu/fnce_papers

Part of the Finance and Financial Management Commons

# Identification Using Russell 1000/2000 Index Assignments: A Discussion of Methodologies

**Abstract**

This paper discusses tradeoffs of various empirical methods used in recent papers that rely on Russell 1000/2000 index assignments for identification. The paper also addresses why different approaches to this identification appear to reach different conclusions about the effect of index assignment on firm's ownership structure and corporate policies.

# Identification using Russell 1000/2000 index assignments: A discussion of methodologies[*]

Ian R. Appel[†], Todd A. Gormley[‡], and Donald B. Keim[§]

March 29, 2016

## Abstract

This paper discusses tradeoffs of various empirical methods used in recent papers that rely on Russell 1000/2000 index assignments for identification. The paper also addresses why different approaches to this identification appear to reach different conclusions about the effect of index assignment on firm's ownership structure and corporate policies.

(*JEL* D22, G23, G30, G34, G35)
**Keywords: instrumental estimation, regression discontinuity, Russell indexes**

## I. Introduction

There have been numerous recent papers that use assignment to the Russell indexes as a source of plausibly exogenous variation in firms' ownership structures. The underlying idea of this identification strategy is to exploit variation in institutional ownership that occurs around the cutoff point used to construct two widely used market benchmarks—the Russell 1000, which comprises the largest 1,000 U.S. stocks, in terms of market capitalization, and the Russell 2000, which comprises the next largest 2,000 stocks. Because portfolio weights assigned to each stock within these indexes are value-weighted, the assignment of an individual stock to one or the other index can have a significant impact on the extent of ownership by institutional investors such as mutual funds. Stocks at the bottom of the Russell 1000 will have small portfolio weights because they represent the smallest firms in that index, while stocks at the top of the Russell 2000 will have weights that are an order of magnitude larger because they represent the largest firms in that index. Therefore, for each dollar invested in an institutional portfolio using the Russell 1000 as a benchmark, very little of it will be invested in stocks at the bottom of that index; while for each dollar invested in an institutional portfolio using the Russell 2000 as a benchmark, a large proportion of it will be invested in stocks at the top of the index. This benchmarking leads to a sharp difference in ownership by institutional investors for stocks at the top of the Russell 2000 relative to stocks at the bottom of the Russell 1000 even though these stocks are otherwise similar in terms of their overall market capitalization.

And yet, despite this relatively straightforward idea, various papers have reached surprisingly different conclusions on exactly how index assignment affects firms' ownership structures. For example, Appel, Gormley, and Keim (2015) propose an instrumental variable estimation involving the Russell 1000/2000 threshold that differs from previous and contemporaneous papers that use the Russell empirical setting for identification. While the focus of Appel, Gormley, and Keim (2015) is different than other papers in this literature, it nevertheless discusses some findings that differ from those reported elsewhere in the literature. For example, Boone and White (2014), Crane, Michenaud, and Weston (2014), and others argue that there is a much larger 10-25 percentage point difference in total institutional ownership around the cutoff that appears to be driven by a difference in *both* passive and active institutional ownership (as measured using institution-level (13F) data and Bushee (2001) classifications for "quasi-index" (i.e., passive), "transient" (i.e., active), and "dedicated" institutions). Appel, Gormley, and Keim (2015), however, argues that there is a much smaller difference in institutional ownership around the cutoff, and that the difference is confined to ownership by passive investors (as shown using both institution-level 13F data and fund-level S12 data, the latter allowing for a more precise measure of active and passive ownership). Moreover, in further contrast to other papers, the IV methodology of Appel, Gormley, and Keim (2015) also fails to detect significant differences in CEO compensation

around the threshold or differences in firm policies related to investments, acquisitions, capital structure, and cash holdings.

This paper discusses the differences in methodology choices of various papers in this literature and why they seem to reach different conclusions. As discussed below, the primary reason for these differences resides in many papers' comparison of stocks around the threshold *after* endogenously resorting stocks within indexes using Russell's float-adjusted market cap-based rankings. This resorting causes a severe endogeneity problem; in essence, these papers compare the least liquid, highest inside ownership stocks of the Russell 1000 to the most liquid, lowest inside ownership stocks of the Russell 2000, causing them to overstate the impact of index assignment on firms' ownership structure and wrongly attribute differences in firms' capital structure, investments, composition of managerial pay, and other outcomes to index assignment.

## II. Specification choices when using the Russell 1000/2000 threshold for identification

A seemingly attractive approach to estimating the effect of ownership structure in the Russell 1000/2000 setting is to use regression discontinuity estimation (RD). The RD estimation attempts to make use of a discontinuity in ownership between the $1000^{th}$ and $1001^{st}$ largest firms at the end of May each year to identify their effect on corporate outcomes. An advantage of this approach is the ability to focus on a subset of firms very close to cutoff, thus reducing concerns that the estimation is not adequately controlling for the one variable that determines index assignment—the end-of-May market caps calculated by Russell—or other possible differences among firms that might be correlated with a firm's index assignment even after conditioning on market capitalization and other controls.

If the end-of-May market capitalization as calculated by Russell was observable and perfectly predicted index assignment, then researchers interested in determining the effect of the being assigned to the Russell 2000 could estimate the following sharp regression discontinuity estimation:

$$Y_{it} = \alpha + \gamma R2000_{it} + \sum_{n=1}^{N} \phi_n (Rank_{it} - 1000)^n + \varepsilon_{it} \qquad (1)$$

where *Y* is the outcome of interest for firm *i* in year *t*, *Rank* is the ranking of firm *i* in year *t* in terms of end-of-May market capitalization (e.g., the $995^{th}$ largest firm would have a rank of 995), and *R2000* is an indicator that equals one for firms assigned to the Russell 2000. The sample could then be restricted to firms very close to the cutoff threshold of *Rank* = 1000, and the polynomial order of controls, *N*, could also be varied.[1] The above estimation of $\gamma$ would identify the effect of being assigned to Russell 2000 on outcome *Y* by testing for a discontinuity in *Y* between the $1000^{th}$ and $1001^{st}$ largest firms, as determined using end-of-May market capitalization.

---

[1] One could also add an additional set of controls, $R2000_{it} \times \sum_{n=1}^{N} (Rank_{it} - 1000)^n$, to allow the functional form of the relation between *Rank* and outcome *Y* to vary above and below the cutoff. See Angrist and Pischke (2009), Lee and Lemieux (2010), and Roberts and Whited (2013) for more details regarding regression discontinuity estimations.

It is not possible to estimate the above equation, however, since the market capitalization used by Russell to determine firms' index assignment at the end of May is not observable to the econometrician. Specifically, Russell calculates firms' market capitalization using a proprietary calculation that does not perfectly match up to market capitalizations reported elsewhere, such as in CRSP, and because of this, econometricians can only imperfectly predict firms' index assignments.

Some have proposed switching to a fuzzy regression discontinuity to overcome this problem (see e.g., Mullins (2014)). In particular, fuzzy regression estimation could be achieved by estimating Equation (1) and using *Treatment* as an instrument for *R2000*, where *Treatment* is an indicator that equals one for firms with a *Rank* greater than 1000, where *Rank* is determined using end-of-May market capitalizations.

A problem with using the end-of-May market capitalization as an instrument in a fuzzy regression discontinuity, however, is that even the market caps provided by Russell, are a weak predictor of index assignment near the cutoff. As can be seen in the top panel of Figure 1, having a ranking above or below 1000 is a poor predictor of being in the Russell 2000 for firms near threshold between the 1000th and 1001st largest firms. In fact, firms with a ranking of 995-1000 are equally likely to be in the Russell 2000 as firms ranked 1001-1005. While the predictive power of end-of-May market caps is better further from this threshold, this is not helpful in that fuzzy regression discontinuity estimations rely on a discontinuity in probability of treatment *at the threshold*, not at points further away from the threshold (Angrist and Pischke (2009), Lee and Lemieux (2010), Roberts and Whited (2013)). Absent such a discontinuity, the estimation will suffer from a weak instrument problem.

The weakness of using fuzzy regression discontinuity estimation in this setting can be further seen in a graph of average quasi-index ownership by firms' ranking in the vicinity of the threshold. This is shown in the bottom panel of Figure 1, which provides a graphical representation of the reduced form version of the fuzzy regression discontinuity estimation. As shown in Figure 1, there is no meaningful jump in quasi-index ownership close to the 1000/2000 threshold using this approach. The reason is that each missed index assignment is introducing considerable noise in the estimation. For example, every firm ranked between 950 and 1000 that is actually in the Russell 2000 will likely be at the top of their index (and hence receive a large jump in ownership by passive investors), while every firm ranked between 1001 and 1050 that is actually in the Russell 1000 will likely be at the bottom of their index. This correlation in the structure of noise near the threshold can also cause a fuzzy RD estimation to yield estimates that are the opposite of the true effect, thus potentially explaining why Mullins (2014) finds a counterintuitive *decrease* in institutional ownership for firms at the top of the Russell 2000. See Crane, Michenaud, and Weston (2014) for more details regarding this latter issue.

To address this concern, other papers propose to use the actual rankings assigned by Russell in a regression discontinuity framework (e.g., see Boone and White (2014); Crane, Michenaud, and Weston (2014)). However, if actual Russell-assigned rankings, rather than end-of-May market cap rankings, are instead used to calculate the forcing variable, *Rank*, then other variables will no longer be continuous at

the threshold, which violates the underlying identification assumption of the regression discontinuity (Angrist and Pischke (2009), Lee and Lemieux (2010), Roberts and Whited (2013)). In particular, there will be a discontinuity in firms' float-adjusted market cap since Russell resorts firms within each index based on their float-adjusted market cap *after* index assignments are made; firms at the bottom of the Russell 1000 will have a much smaller float-adjusted market cap than firms at the top of the Russell 2000. This is seen in the top half of Figure 2, where we plot the average Ln(float-adjusted market cap) by firms' Russell-assigned ranking. On average, the firm with a Russell-assigned ranking of 1000 (i.e., the bottom firm in the Russell 1000) has a float-adjusted market cap that is more than two log points smaller than the firm with a Russell-assigned ranking of 1001 (i.e., the top firm in the Russell 2000).

This difference in float-adjusted market cap between firms at the bottom of the Russell 1000 and the top of the Russell 2000 causes this modified RD methodology to overstate the actual impact of index assignment on firms' ownership structure and corporate policies. In particular, this RD methodology compares the least liquid stocks of the Russell 1000 index (which active institutions will *endogenously* avoid) against the most liquid stocks of the Russell 2000. Moreover, the differences in float-adjusted market cap around the cutoff will mechanically generate an endogenous shift in ownership structure. Stocks at the very bottom of the Russell 1000 are those that have a much smaller float-adjusted market cap relative to their total market cap, and this occurs when large blocks of a firm's equity is held by insiders or other non-financial companies. These endogenous differences in liquidity and inside ownership likely explain why Boone and White (2014), Crane, Michenaud, and Weston (2014), and other papers that use this identification strategy argue there is a much larger 10-25 percentage difference in institutional ownership (both active and passive) around the cutoff.

The IV estimation of Appel, Gormley, and Keim (2015) is designed to overcome these shortcomings but still allow the Russell 1000/2000 threshold be used as a source of identification. As demonstrated in Appel, Gormley, and Keim (2015), there is only a difference in passive institutional ownership using their IV estimation, and the observed magnitude (which is about an order of magnitude less than reported in other papers) matches what one would predict based on the amount of money passively tracking each index. Their IV specification's ability to isolate the exogenous differences in ownership also explains why their findings for corporate policies and managerial pay differ from previous studies of the Russell 1000/2000 cutoff; in particular, they find little evidence of a difference in firms' leverage, investment, acquisitions, or composition of managerial pay.[2] Further discussion of their methodology and findings can be found in their paper.

---

[2] Using May 31st CRSP market capitalization to determine rankings (within the actual assigned index), as done in a robustness check by Crane, Michenaud, and Weston (2014), will be problematic for a similar reason. Because firms are resorted within an index using total end-of-May CRSP market caps in this alternative approach, there will now exist a discontinuity in Ln(*Mktcap*) near the threshold. The firm with the smallest end-of-May CRSP *Mktcap* within the Russell 1000 will be assigned a rank of 1000, while the firm with the largest end-of-May CRSP *Mktcap* within the Russell 2000 will be assigned a rank of 1001. This discontinuity occurs because the CRSP market caps are only a noisy predictor of the true, but unobserved, forcing variable. This is shown in Figure 2.

**References**

Angrist, Joshua D., and Jörn-Steffen Pischke, 2009, *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press, Princeton.

Boone, Audra L., and Joshua T. White, 2014, The effect of institutional ownership on firm transparency and information production, forthcoming in *Journal of Financial Economics*.

Bushee, Brian J., 2001, Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18, 207–246.

Crane, Alan D., Sébastien Michenaud, and James P. Weston, 2014, The effect of institutional ownership on payout policy: Evidence from index thresholds, working paper.

Mullins, William, 2014, The governance impact of index funds: Evidence from a regression discontinuity, working paper.

Roberts, Michael R., and Toni M. Whited, 2013, Endogeneity in empirical corporate finance, in George Constantinides, Milton Harris, and Rene Stulz, eds. *Handbook of the Economics of Finance*, Vol. 2., Part A, pp. 493–572.

Wooldridge, J. M. 1995. Score diagnostics for linear models estimated by two stage least squares. In *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao*, ed. G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66-87. Oxford: Blackwell.
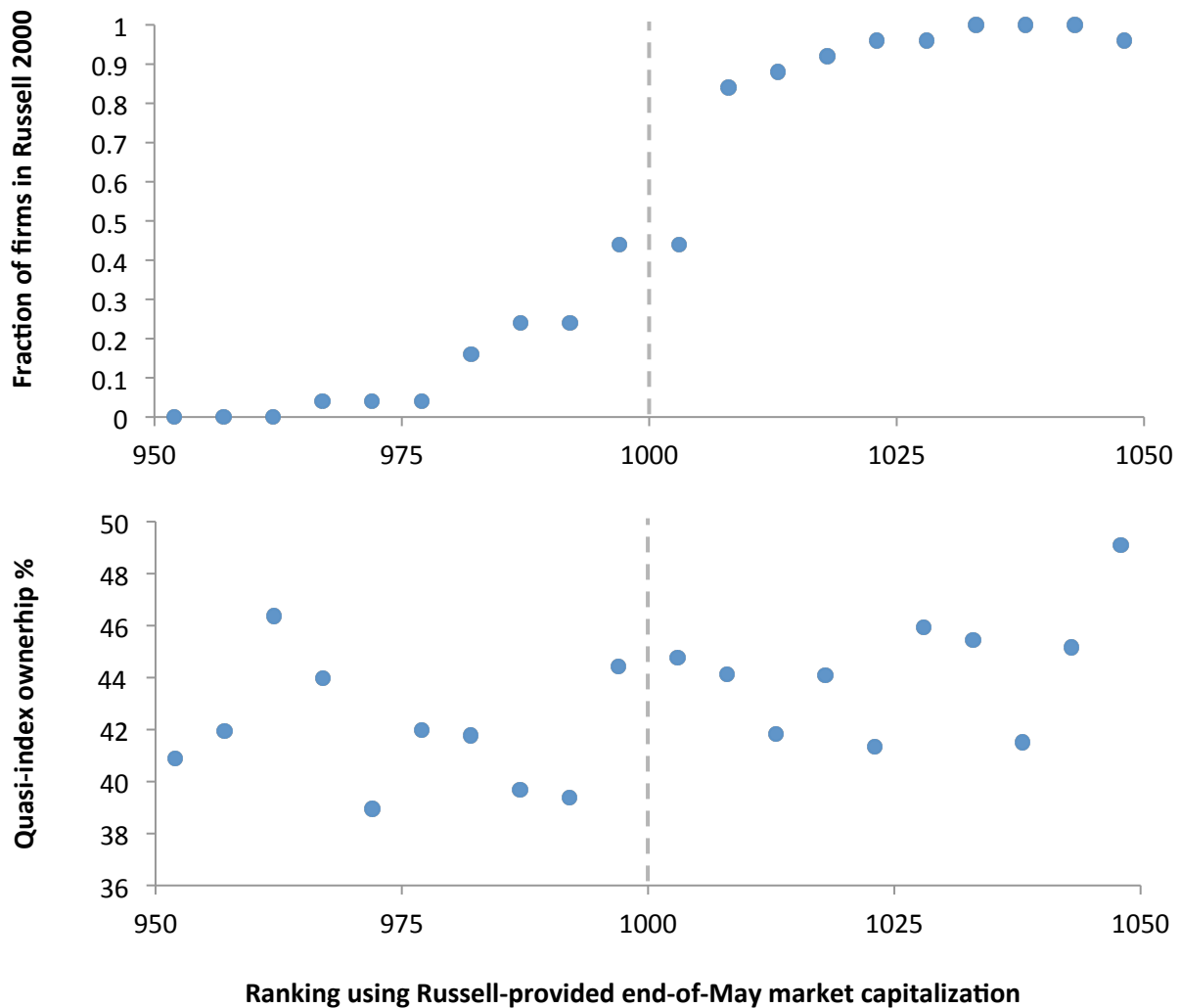
**Figure 1**
**Probability of treatment and quasi-index ownership by ranking near the Russell 1000/2000 threshold using Russell-provided market capitalizations**
This figure plots the average fraction of firm-year observations in the Russell 2000 and percent quasi-index ownership by size ranking for the 950th to 1050th largest firms, where ranking is determined using end-of-May market capitalization numbers provided directly by Russell Investments for firms in the Russell 1000/2000 indices between 2002 and 2006. Averages are calculated using bins of five rankings and data from 2002-2006.
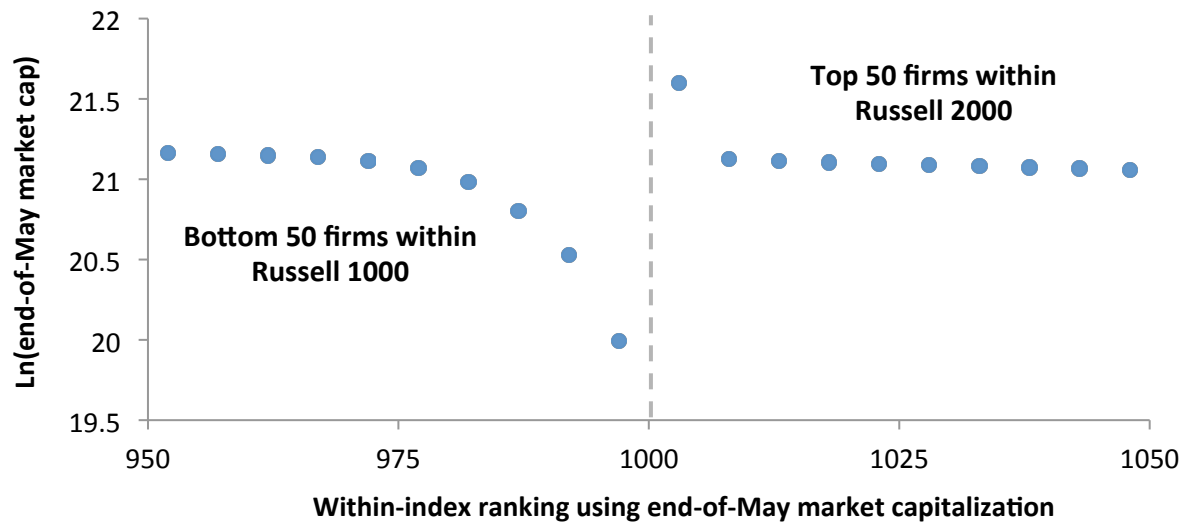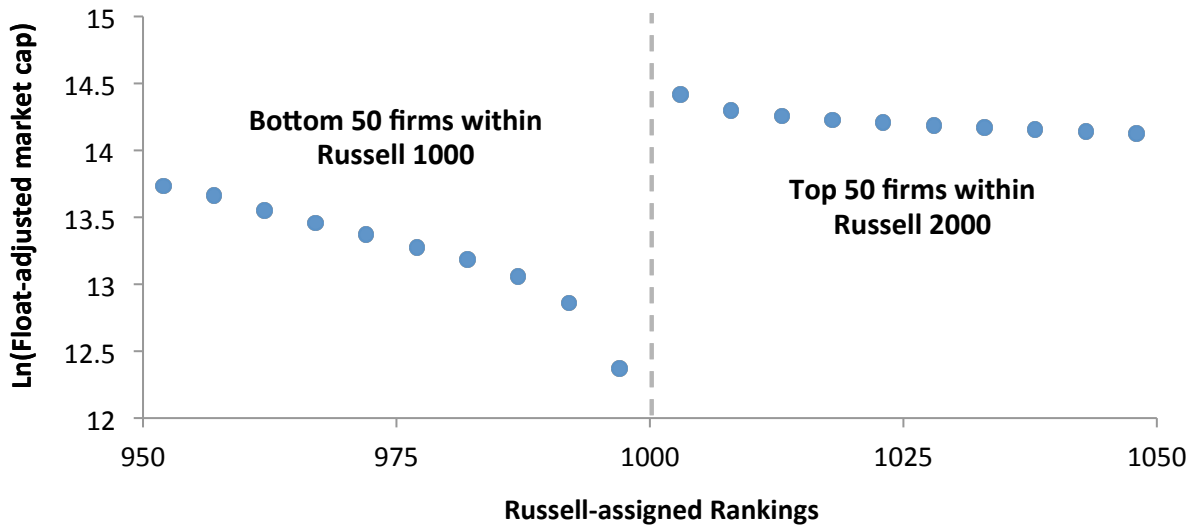
**Figure 2**

**Average Ln(*Float*) and Ln(*Mktcap*) by ranking, where ranking is calculated using either float-adjusted portfolio weights assigned by Russell or within-index rankings based on end-of-May market capitalizations**

The top panel of this figure plots the average Ln(float-adjusted market cap) by Russell-determined rankings for the bottom 50 firms in the Russell 1000 index and the top 50 firms in the Russell 2000 index for the years 1998-2006.  A ranking of 1000 reflects the firm with the lowest portfolio weight in the Russell 1000 index, while a ranking of 1001 reflects the firm with the highest portfolio weight in the Russell 2000 index.  The bottom panel of this figure plots the average Ln(end-of-May CRSP market cap) by size ranking for firms ranked between 950 and 1050, where ranking is determined using within-index end-of-May CRSP market caps.   A ranking of 1000 reflects the firm with the lowest end-of-May market cap in the Russell 1000 index, while a ranking of 1001 reflects the firm with the highest end-of-May market cap in the Russell 2000 index.  Averages are calculated using bins of five rankings for the years 1998-2006.