INTERNET APPENDIX

Russell Index Reconstitutions, Institutional Investors, and Corporate Social Responsibility

Simon Glossner
Catholic University of Eichstätt-Ingolstadt

October 24, 2019

Table IA1: Instrumental variable approach (incorrectly) based on sharp RD

Description: This table estimates an IV approach based on sharp RD (which uses float-adjusted June ranks). Formally, the first-stage regression of the IV approach is specified by

$$IO_{i,t} = \alpha_0 + \tau_0 \text{R2000}_{i,t} + \sum\nolimits_n \delta_n (\text{Rank}_{i,t}^{Jun})^n + \sum\nolimits_n \gamma_n \text{R2000}_{i,t} (\text{Rank}_{i,t}^{Jun})^n + v_t + u_{i,t},$$

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 in year t, $\text{Rank}_{i,t}^{Jun}$ is the rank of firm i during the index reconstitution of year t, v_t are year dummies, and $u_{i,t}$ is the error term. I construct variable $\text{Rank}_{i,t}^{Jun}$ based on Russell's float-adjusted end-of-June ranks. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: The IV approach based on sharp RD shows that firms at the top of the Russell 2000 have 12–27 percentage points higher institutional ownership than firms at the bottom of the Russell 1000.

Dependent	Independent	(1)	(2)	(3)	(4)
Total institutional ownership	R2000	0.120***	0.177***	0.226***	0.265***
	(T)	(10.60)	(10.68)	(9.43)	(8.54)
	Polynomial(n)	1	2	3	4
	Observations	26629	26629	26629	26629
		(5)	(6)	(7)	(8)
Ownership by quasi-index investors	R2000	0.107***	0.126***	0.174***	0.215***
	(T)	(14.24)	(11.85)	(12.15)	(12.31)
	Polynomial(n)	1	2	3	4
	Observations	26629	26629	26629	26629

Table IA2: First-stage of fuzzy regression discontinuity approaches

Description: This table estimates first-stage regressions of a fuzzy RD approach specified by

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

where $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 after the annual index reconstitution in June of year t, $\text{Rank}_{i,t}^{May}$ is the end-of-May rank of firm i at year t, $\text{PredictR2000}_{i,t}$ is a dummy indicating whether $\text{Rank}_{i,t}^{May}$ predicts membership in the Russell 2000, v_t are year dummies, and $u_{i,t}$ is the error term. Variable $\text{Rank}_{i,t}^{May}$ is centered around the cutoff. Panel A uses data from Compustat and CRSP to construct the May ranks (see Appendix A for details), and Panel B uses data only from CRSP to construct the May ranks. F-statistic indicates the instrument strength. The regressions are estimated only on those observations that lie within a bandwidth close to the threshold. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: CRSP/Compustat May rankings are a better predictor of actual index assignment than CRSP May rankings.

Panel A: May ranks are constructed with data from CRSP and Compustat

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

Panel B: May ranks are constructed with data only from CRSP

Dependent:	R2000			
	(1)	(2)	(3)	(4)
PredictR2000	−0.129***	0.243***	0.433***	0.536***
(T)	(−4.30)	(10.09)	(20.98)	(29.58)
Bandwidth	100	200	300	400
Observations	1790	3569	5343	7117
F-Statistic	298.0	317.7	519.5	686.2
Adj. R^2	0.55	0.66	0.70	0.75

Table IA3: Instrumental variable approach by Appel, Gormley, and Keim (2016)

Description: This table estimates an IV approach by Appel, Gormley, and Keim (2016). The first-stage regression of this approach is specified by

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

where $\mathrm{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 index at time t, $\mathrm{Mktcap}_{i,t}$ is the logarithm of the May market cap of firm i in year t, $\mathrm{Float}_{i,t}$ is the logarithm of the float-adjusted June market cap in year t, v_t are year dummies, and $u_{i,t}$ is the error term. Panel A presents the original approach by Appel, Gormley, and Keim (2016), which calculates variable $\mathrm{Mktcap}_{i,t}$ by using data from CRSP and selects the bandwidth based on float-adjusted June ranks. Panel B presents a modified version, which selects the bandwidth based on unadjusted May market caps. Panel C shows a modified version, which selects the bandwidth based on May market caps *and* calculates variable $\mathrm{Mktcap}_{i,t}$ by multiplying stock prices from CRSP by outstanding shares from Compustat. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: This IV approach shows a lower difference in quasi-index investors between the firms close to the threshold when using the modified approaches.

Panel A: Original IV approach by Appel, Gormley, and Keim (2016)

Dependent:	Ownership of quasi-index investors			
	(1)	(2)	(3)	(4)
R2000	0.011	0.017**	0.023***	0.027***
(T)	(0.93)	(2.16)	(3.16)	(3.73)
Bandwidth	100	200	300	400
Observations	1784	3563	5332	7105

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

Dependent:	Ownership of quasi-index investors			
	(1)	(2)	(3)	(4)
R2000	0.004	0.010	0.016**	0.018**
(T)	(0.35)	(1.05)	(2.03)	(2.28)
Bandwidth	100	200	300	400
Observations	1790	3569	5343	7117

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

Dependent:	Ownership of quasi-index investors			
	(1)	(2)	(3)	(4)
R2000	0.007	0.012	0.016*	0.019**
(T)	(0.50)	(1.12)	(1.81)	(2.37)
Bandwidth	100	200	300	400
Observations	1794	3567	5341	7117

Table IA4: Replication of Rubio and Vazquez (2018)

Description: This table replicates the results by Rubio and Vazquez (2018). It estimates an IV approach that is specified by

$$IO_{i,t} = \alpha_0 + \tau_0 \mathrm{R2000}_{i,t} + \sum\nolimits_n l_n (\mathrm{Mktcap}_{i,t})^n + \rho_0 \mathrm{Float}_{i,t} + v_t + u_{i,t}$$
$$Y_{i,t+1} = \alpha_1 + \tau_1 \widehat{IO_{i,t}} + \sum\nolimits_n \lambda_n (\mathrm{Mktcap}_{i,t})^n + \rho_1 \mathrm{Float}_{i,t} + v_{t+1} + \epsilon_{i,t+1},$$

where $IO_{i,t}$ is ownership of institutional investors, $\mathrm{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 index at time t, $\mathrm{Mktcap}_{i,t}$ is the logarithm of the unadjusted end-of-May CRSP market capitalization of firm i in year t, $\mathrm{Float}_{i,t}$ is the logarithm of the float-adjusted end-of-June market capitalization of firm i in year t, v_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t+1}$ are the error terms. Panel A estimates the first-stage regressions of the IV approach. Panel B shows the second-stage regressions. The regressions are estimated only on those observations that lie within a bandwidth close to the threshold (based on float-adjusted end-of-June ranks). Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: This replication shows that institutional investors have no significant effect on CSR, contrary to the findings by Rubio and Vazquez (2018).

Panel A: First-stage regressions

Dependent:	Institutional ownership (1)	(2)	(3)
R2000	0.038	0.068***	0.066***
(T)	(1.22)	(2.69)	(2.98)
Polynomial(n)	2	2	2
Bandwidth	300	500	700
Observations	903	1544	2200

Panel B: Second-stage regressions

Dependent	Independent	(1)	(2)	(3)
Strengths-only CSR score	$\widehat{IO}$	−12.981	−7.559	−2.952
	(T)	(−0.88)	(−1.29)	(−0.96)
	Polynomial(n)	2	2	2
	Bandwidth	300	500	700
	Observations	1616	2703	3790
		(4)	(5)	(6)
Concerns-only CSR score	$\widehat{IO}$	−6.208	−6.853	−5.318
	(T)	(−0.80)	(−1.32)	(−1.39)
	Polynomial(n)	2	2	2
	Bandwidth	300	500	700
	Observations	1616	2703	3790

Table IA5: Replication of Chen, Dong, and Lin (2018)

Description: This table replicates the results by Chen, Dong, and Lin (2018). It estimates an IV approach that is specified by

$$IO_{i,t} = \alpha_0 + \tau_0 \mathrm{R2000}_{i,t} + \sum_n \delta_n (\mathrm{Rank}_{i,t})^n + \sum_n \gamma_n \mathrm{R2000}_{i,t} (\mathrm{Rank}_{i,t})^n + \xi_0 \mathrm{FloatAdj}_{i,t} + \beta_0 X_{i,t} + \eta_j + v_t + u_{i,t}$$

$$Y_{i,t} = \alpha_1 + \tau_1 \widehat{IO_{i,t}} + \sum_n \lambda_n (\mathrm{Rank}_{i,t})^n + \sum_n l_n \mathrm{R2000}_{i,t} (\mathrm{Rank}_{i,t})^n + \xi_1 \mathrm{FloatAdj}_{i,t} + \beta_1 X_{i,t} + \eta_j + v_t + \epsilon_{i,t},$$

where $IO_{i,t}$ is ownership of institutional investors, $Y_{i,t}$ is the net CSR score, $\mathrm{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 in year t, $\mathrm{Rank}_{i,t}$ is the rank of firm i during the index reconstitution of year t, $X_{i,t}$ is a vector of control variables (size, leverage, return on assets, market-to-book, cash holdings, advertising, R&D intensity, sales growth, dividends), η_j are industry (sic2) dummies, v_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t}$ are the error terms. Variable $\mathrm{FloatAdj}_{i,t}$ is the difference between the rank implied by the end-of-May market capitalization and the actual rank assigned by Russell in June. Both panels show the second-stage regressions. Panel A shows the original approach, which constructs rank $\mathrm{Rank}_{i,t}$ based on Russell's float-adjusted end-of-June ranks. Panel B shows the modified approach, which constructs rank $\mathrm{Rank}_{i,t}$ based on the unadjusted end-of-May ranks (see Appendix A). Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: This replication shows that institutional investors have no significant effect on CSR when using unadjusted May rankings in the approach.

Panel A: Original approach (using float-adjusted June ranks)

Dependent:	Net CSR score		
	(1)	(2)	(3)
$\widehat{IO}$	4.217***	2.750**	0.739
(T)	(2.61)	(2.25)	(0.67)
Polynomial(n)	3	3	3
Bandwidth	50	150	250
Observations	474	1517	2573

Panel B: Modified approach (using unadjusted May ranks)

Dependent:	Net CSR score		
	(1)	(2)	(3)
$\widehat{IO}$	−1.357	−0.325	0.616
(T)	(−0.44)	(−0.25)	(0.49)
Polynomial(n)	3	3	3
Bandwidth	50	150	250
Observations	483	1536	2601

Table IA6: Replication of Hou and Zhang (2017)

Description: This table replicates the results by Hou and Zhang (2017). It estimates an IV approach that is specified by

$$IO_{i,t} = \alpha_0 + \tau_0 \text{R2000}_{i,t} + \sum\nolimits_n l_n(\text{Mktcap}_{i,t})^n + \rho_0 \text{Float}_{i,t} + \beta_0 X_{i,t} + \eta_j + v_t + u_{i,t}$$

$$Y_{i,t+1} = \alpha_1 + \tau_1 \widehat{IO_{i,t}} + \sum\nolimits_n \lambda_n(\text{Mktcap}_{i,t})^n + \rho_1 \text{Float}_{i,t} + \beta_1 X_{i,t} + \eta_j + v_{t+1} + \epsilon_{i,t+1},$$

where $Y_{i,t+1}$ is the net CSR score, $IO_{i,t}$ is ownership of passive funds in percentage, $\text{R2000}_{i,t}$ is a dummy indicating whether firm i is a member of the Russell 2000 index at time t, $\text{Mktcap}_{i,t}$ is the logarithm of the end-of-May CRSP market capitalization of firm i in year t, $\text{Float}_{i,t}$ is the logarithm of the float-adjusted end-of-June market capitalization of firm i in year t, $X_{i,t}$ is a vector of control variables (total assets, return on assets, market-to-book, tangibility, cash holdings, and dividends), η_j are industry (sic2) dummies, v_t are year dummies, and $u_{i,t}$ and $\epsilon_{i,t+1}$ are the error terms. Both panels show the second-stage regressions. Panel A shows the original approach, which uses CRSP May market caps and selects the bandwidth based on float-adjusted end-of-June rankings. Panel B shows the modified approach, which uses CRSP/Compustat May market caps and selects the bandwidth based on unadjusted end-of-May ranks. Standard errors are clustered on the firm level. The number in parenthesis is the t-statistic of the estimate.

Interpretation: This replication shows that passive mutual funds have no significant effect on CSR when using the modified approach instead of the original approach.

Panel A: Original approach (CRSP mcaps, June bandwidth)

Dependent:	Net CSR score		
	(1)	(2)	(3)
$\widehat{IO}$	-0.240^{**}	-0.248^{**}	-0.253^{**}
(T)	(-2.38)	(-2.28)	(-2.13)
Polynomial(n)	1	2	3
Bandwidth	250	250	250
Observations	1677	1677	1677

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

Dependent:	Net CSR score		
	(1)	(2)	(3)
$\widehat{IO}$	-0.430	-0.327	-0.258
(T)	(-1.50)	(-1.64)	(-1.43)
Polynomial(n)	3	3	3
Bandwidth	150	250	350
Observations	1013	1692	2369

Simulation R Code

```
library(AER)
library(Runuran)
library(lmtest)

RussellSim = function(noiseFunction, noisePar, outLoopN){

  # preallocate output and run loop
  out_c = matrix(NA, nrow=outLoopN, ncol=6)
  out_t = matrix(NA, nrow=outLoopN, ncol=6)
  for (j in 1:outLoopN) {

    # create dataframe with mcaps, ranks, and index labels
    data = data.frame(n=1:3000)
    data$mcaps = sort(rlnorm(3000, meanlog=7.0, sdlog=1.4), decreasing=TRUE)
    data$index = c(rep(1000, 1000), rep(2000, 2000))
    data$r2000 = ifelse(data$index == 2000, 1, 0)
    data$rank  = rank(-data$mcaps) - 1000

    # calculate free float and float-adjusted June ranks
    data$float = data$mcaps * (1-urexp(3000, rate=3.5, ub=1))
    data[data$index==1000, "adjrank"] = rank(-data[data$index==1000, "float"]) - 1000
    data[data$index==2000, "adjrank"] = rank(-data[data$index==2000, "float"])

    # calculate institutional ownership of period t
    data$io_error = rnorm(3000, 0.35, 0.23)
    data$io = (-0.16)*log(data$mcaps) + 0.20*log(data$float) + data$io_error
    data$io = replace(data$io, data$io > 1, 1)
    data$io = replace(data$io, data$io < 0, 0)

    # calculate mcaps and ranks of period t+1 (only for switcher approach)
    c = rlnorm(3000, 0, 0.25)
    data$mcaps_t1 = data$mcaps*c
    data$rank_t1  = rank(-data$mcaps_t1) - 1000
    data$index_t1 = ifelse(data$rank_t1 <= 0, 1000, 2000)
    data$toR2000  = ifelse(data$index == 1000 & data$index_t1 == 2000, 1, 0)
    data$toR1000  = ifelse(data$index == 2000 & data$index_t1 == 1000, 1, 0)

    # calculate institutional ownership of t+1 (only for switcher approach)
    data$io_t1  = (-0.16)*log(data$mcaps*c) + 0.20*log(data$float*c) +
      0.9*data$io_error + 0.1*rnorm(3000, 0.35, 0.23)
    data$io_t1 = replace(data$io_t1, data$io_t1 > 1, 1)
    data$io_t1 = replace(data$io_t1, data$io_t1 < 0, 0)

    # create noisy CRSP market caps
    if (noiseFunction == "uniform"){
      data$crsp    = data$mcaps    * runif(3000, min=1-noisePar, max=1+noisePar)
      data$crsp_t1 = data$mcaps_t1 * runif(3000, min=1-noisePar, max=1+noisePar)
```

```
} else if (noiseFunction == "normal"){
  data$crsp    = data$mcaps    * rnorm(3000, 1, noisePar)
  data$crsp_t1 = data$mcaps_t1 * rnorm(3000, 1, noisePar)
} else if (noiseFunction == "triangle") {
  data$crsp    = data$mcaps    * urtriang(3000, a=1-noisePar, b=1+noisePar, m=1)
  data$crsp_t1 = data$mcaps_t1 * urtriang(3000, a=1-noisePar, b=1+noisePar, m=1)
} else if (noiseFunction == "laplace") {
  data$crsp    = data$mcaps    * urlaplace(3000, location=1, scale=noisePar, lb=0.1)
  data$crsp_t1 = data$mcaps_t1 * urlaplace(3000, location=1, scale=noisePar, lb=0.1)
}

# run fuzzy RD approach
data$crsprank = rank(-data$crsp) - 1000
data$r2000treat = ifelse(data$crsprank > 0, 1, 0)
data$crsprank_r2000 = data$crsprank * data$r2000
data$crsprank_r2000treat = data$crsprank * data$r2000treat

mod4stage1 = lm(r2000 ~ crsprank + r2000treat + crsprank_r2000treat,
                data=data, subset=data$crsprank %in% seq(-200,+200))
coeftest_mod4stage1 = coeftest(mod4stage1)
out_c[j, 1] = coeftest_mod4stage1[3,1]
out_t[j, 1] = coeftest_mod4stage1[3,3]

mod4 = ivreg(io ~ crsprank + r2000 + crsprank_r2000 |
               crsprank + r2000treat + crsprank_r2000treat,
             data=data, subset=data$crsprank %in% seq(-200,+200))
coeftest_mod4 = coeftest(mod4)
out_c[j, 2] = coeftest_mod4[3,1]
out_t[j, 2] = coeftest_mod4[3,3]

# run IV approach by AGK 2016 with May bandwidth
mod3b = lm(io ~ r2000+log(crsp)+I(log(crsp)^2)+I(log(crsp)^3)+log(float),
           data=data, subset=data$crsprank %in% seq(-200,+200))
coeftest_mod3b = coeftest(mod3b)
out_c[j, 3] = coeftest_mod3b[2,1]
out_t[j, 3] = coeftest_mod3b[2,3]

# run IV approach by AGK 2016 with June bandwidth
mod3c = lm(io ~ r2000+log(crsp)+I(log(crsp)^2)+I(log(crsp)^3)+log(float),
           data=data, subset=data$adjrank %in% seq(-200,+200))
coeftest_mod3c = coeftest(mod3c)
out_c[j, 4] = coeftest_mod3c[2,1]
out_t[j, 4] = coeftest_mod3c[2,3]

# run IV based on index switchers
data$crsprank_t1 = rank(-data$crsp_t1) - 1000
data$crsprank_diff = data$crsprank_t1 - data$crsprank
data$io_change = data$io_t1 - data$io

mod5 = lm(io_change ~ toR1000 + toR2000 + crsprank_diff, data=data)
```

```r
    coeftest_mod5 = coeftest(mod5)
    out_c[j, 5] = coeftest_mod5[2,1]
    out_t[j, 5] = coeftest_mod5[2,3]
    out_c[j, 6] = coeftest_mod5[3,1]
    out_t[j, 6] = coeftest_mod5[3,3]
  }

  # return output
  coef    = apply(out_c, 2, function(x) mean(x))
  ols_t   = apply(out_t, 2, function(x) mean(x))
  ols_sig = apply(out_t, 2, function(x) sum(abs(x)>=1.64)/length(x))
  output  = rbind(coef, ols_t, ols_sig)
  return(output)

}

outLoopN = 100000

normal_mod1   = RussellSim("normal", 0.05, outLoopN)
normal_mod2   = RussellSim("normal", 0.09, outLoopN)
normal_mod3   = RussellSim("normal", 0.13, outLoopN)
normal_mod4   = RussellSim("normal", 0.17, outLoopN)
uniform_mod1  = RussellSim("uniform", 0.08, outLoopN)
uniform_mod2  = RussellSim("uniform", 0.13, outLoopN)
uniform_mod3  = RussellSim("uniform", 0.18, outLoopN)
uniform_mod4  = RussellSim("uniform", 0.23, outLoopN)
triangle_mod1 = RussellSim("triangle", 0.12, outLoopN)
triangle_mod2 = RussellSim("triangle", 0.22, outLoopN)
triangle_mod3 = RussellSim("triangle", 0.32, outLoopN)
laplace_mod1  = RussellSim("laplace", 0.04, outLoopN)
laplace_mod2  = RussellSim("laplace", 0.08, outLoopN)
laplace_mod3  = RussellSim("laplace", 0.12, outLoopN)
```