Rules and Regression Discontinuities in Asset Markets

Yen-Cheng Chang
Shanghai Advanced Institute of Finance

Harrison Hong
Princeton University
Quantitative portfolio rules induce natural experiments in financial markets that help us understand how demand shocks influence asset prices.

- Popular over the last 30 years and now widely adopted by the investment community.

- Mechanically based on price or fundamental signals: momentum, value, index funds, etc. Specific rules on what gets bought and what gets sold, which leads to discontinuities in demand.

This paper: portfolio indexing to the Russell 2000 induce regression discontinuity (RD) experiments.
Russell’s indices are based on simple rules of firm size:

- Russell 1000: #1-1000
- Russell 2000: #1001-3000
- Russell 3000: #1-3000
- Russell 3000E: #1-4000

Annual Index reconstitution on the last Friday of June. Constituent firms are ranked according to their float-adjusted market capitalization.

Value weighted indices

Russell 2000 has more than $500b benchmarked to it.
Russell 2000 induces a regression discontinuity experiment at the #1000 cutoff

- Stocks just-included in the Russell 2000 (e.g. #1001-1100) creates forced buying and hedging demands compared to the just-missed (e.g. #901-1000).

- Whether stocks land on the left or on the right of the cutoff is random.

- Regression discontinuity (RD) design allows for estimation of the impact around the vicinity of the cutoff.
  - Fund managers do try to manipulate stock prices.
  - Lee and Lemieux (2010): imprecise control around cutoff validates the RD design

- Further investigate other variables of interest: return, valuation, liquidity, short interest, volatility, comovement
Stocks just-included vs. just-excluded around the #1000 cutoff have:

- Higher institutional ownership
- Higher return (only in June) and valuation
- Higher turnover and liquidity
- Higher short interest
- Higher comovement with Russell 1000 & Russell 2000

Negligible effect around the #3000 cutoff.
How do demand shocks due to forced buying/selling influence stock prices?

- Classical efficient market hypothesis
- Alternative: Prices deviate from fundamental value due to exogenous demand shocks
  - Limits to arbitrage
  - Price comovement/detachment
  - Bubbles and Crashes
  - Hong and Stein (2007); Gromb and Vayanos (2010)
Literature Review

➢ Stock market indexing
  – Shleifer (1986); Harris and Gurel (1986)
  – Barberis, Shleifer, and Wurgler (2005); Greenwood (2008)

➢ Regression Discontinuity Design/Applications
  – Ludwig and Miller (2007); Lee (2008)
  – Imbens and Lemieux (2007)
The RD design can potentially be applied to other quantitative portfolio rules.

S&P 500’s choice of constituents is not based on transparent rules and the present challenges to properly identify the index effect.

- Traditional methodologies face similar endogeneity challenges.

The RD design is a local estimation which is better identified
  - Deepens our understanding of the effect of demand shocks in asset markets.
Russell constituents data
   – Holdings data for Russell 1000, 2000 and 3000E

Variables of interest
   – \textit{IO}: institutional ownership
   – \textit{RET}: monthly stock return
   – \textit{MtB}: market-to-book ratio
   – \textit{TURN}: stock turnover; NASDAQ volume is adjusted
   – \textit{SR}: fraction of outstanding shares shorted
   – \textit{VOL}: return volatility within a month
   – \textit{CORR2000}; \textit{CORR1000}: return correlation with Russell 2000 or Russell 1000
### Summary Statistics—By Period

<table>
<thead>
<tr>
<th></th>
<th>Summary Statistics, All Periods</th>
<th>Medians (Means), Subperiods</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IO$</td>
<td>0.444</td>
<td>0.626</td>
</tr>
<tr>
<td>$RET$</td>
<td>-0.051</td>
<td>0.008</td>
</tr>
<tr>
<td>$MtB$</td>
<td>1.629</td>
<td>2.451</td>
</tr>
<tr>
<td>$TURN$</td>
<td>0.048</td>
<td>0.094</td>
</tr>
<tr>
<td>$ILLIQ$</td>
<td>0.077</td>
<td>0.299</td>
</tr>
<tr>
<td>$SR$</td>
<td>0.007</td>
<td>0.018</td>
</tr>
<tr>
<td>$VOL$</td>
<td>0.014</td>
<td>0.021</td>
</tr>
<tr>
<td>$CORR2000$</td>
<td>0.210</td>
<td>0.416</td>
</tr>
<tr>
<td>$CORR1000$</td>
<td>0.193</td>
<td>0.404</td>
</tr>
</tbody>
</table>

---

Yen-Cheng Chang & Harrison Hong  
Rules and Regression Discontinuities in Asset Markets
## Summary Statistics—By Index

<table>
<thead>
<tr>
<th></th>
<th>Russell 1000</th>
<th></th>
<th>Russell 2000</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>Stdev</td>
<td>No. Obs.</td>
</tr>
<tr>
<td><strong>IO</strong></td>
<td>0.649</td>
<td>0.621</td>
<td>0.207</td>
<td>39,421</td>
</tr>
<tr>
<td><strong>RET</strong></td>
<td>0.009</td>
<td>0.009</td>
<td>0.120</td>
<td>176,784</td>
</tr>
<tr>
<td><strong>MtB</strong></td>
<td>2.742</td>
<td>4.215</td>
<td>5.757</td>
<td>155,805</td>
</tr>
<tr>
<td><strong>TURN</strong></td>
<td>0.095</td>
<td>0.148</td>
<td>0.170</td>
<td>177,383</td>
</tr>
<tr>
<td><strong>ILLIQ</strong></td>
<td>0.091</td>
<td>0.719</td>
<td>10.309</td>
<td>175,356</td>
</tr>
<tr>
<td><strong>SR</strong></td>
<td>0.017</td>
<td>0.028</td>
<td>0.035</td>
<td>141,549</td>
</tr>
<tr>
<td><strong>VOL</strong></td>
<td>0.018</td>
<td>0.023</td>
<td>0.017</td>
<td>176,740</td>
</tr>
<tr>
<td><strong>CORR2000</strong></td>
<td>0.415</td>
<td>0.394</td>
<td>0.266</td>
<td>176,772</td>
</tr>
<tr>
<td><strong>CORR1000</strong></td>
<td>0.440</td>
<td>0.415</td>
<td>0.267</td>
<td>176,772</td>
</tr>
</tbody>
</table>
For each of the 1000 ranking positions above and below the Russell 2000 cutoff

- Compute the variables of interest

- Average across the sample years: 1991-2008

- Average across 12 months of each reconstitution (June-May)
  • Monthly plots
Raw (unsmoothed) Plots

Panel A: IO

Panel B: RET

Panel C: MtB

Panel D: TURN

Panel E: ILLIQ

Panel F: SR

Yen-Cheng Chang & Harrison Hong  Rules and Regression Discontinuities in Asset Markets
Raw (unsmoothed) Plots

Panel G: VOL

Panel H: CORR2000

Panel I: CORR1000
RD (smoothed) Plots

Panel A: IO

Panel B: RET

Panel C: MtB

Panel D: TURN

Panel E: ILLIQ

Panel F: SR

INTRODUCTION

EMPIRICAL RESULTS

CONCLUSION
RD (smoothed) Plots

Panel G: VOL

Panel H: CORR2000

Panel I: CORR1000
RD Tests: Local Linear Regression Approach

- For each variable of interest and each Russell ranking position, compute the average across the 12-months between each reconstitution.

- Local Linear Regression:

\[ Y = \alpha_l + \tau \cdot D + \beta_l \cdot (X - c) + (\beta_r - \beta_l) \cdot D \cdot (X - c) + \varepsilon, \]

where \( c - h_L \leq X \leq c + h_R \).
Bandwidth choice:

CV bandwidth: “Leave-one-out” procedure
- For each bandwidth $h$ and each observation $X_i$
  - If $X$ is on the RHS of the cutoff, run a linear regression with observations $X_i < X \leq X_i + h$.
  - If $X$ is on the LHS of the cutoff, run a linear regression with observations $X_i - h \leq X < X_i$.
- The CV is the summed mean squared error

$$CV_y(h) = \frac{1}{N} \sum_{i=1}^{N} \left( Y_i - \hat{Y}(X_i) \right)^2.$$
– The optimal CV bandwidth

\[ h_{CV} = \arg \min_{h} CV_{Y}(h) \]

– The amount of data to use:
  • Imbens and Lemieux (2007): 50% of data on either side
  • Ludwig and Miller (2007): 5% of data on either side
  • This paper: Restrict to 150.

➢ The ROT bandwidth:

\[ h_{ROT} = 2.702 \left( \frac{\hat{\sigma}^2 R}{\sum_{i=1}^{N} \hat{m}''(X_i)^2} \right) \]
### EMPIRICAL RESULTS

- **Preview**
  - Data and Summary Statistics
  - Aggregate Results
  - Results by Month
  - Lower-end Cutoff

#### RD Tests: Results

<table>
<thead>
<tr>
<th></th>
<th>CV Bandwidth</th>
<th>ROT Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IO</strong></td>
<td>0.323 (12.94)</td>
<td>36</td>
</tr>
<tr>
<td><strong>RET</strong></td>
<td>0.033 (3.07)</td>
<td>97</td>
</tr>
<tr>
<td><strong>MiB</strong></td>
<td>2.674 (3.38)</td>
<td>86</td>
</tr>
<tr>
<td><strong>TURN</strong></td>
<td>0.139 (8.42)</td>
<td>48</td>
</tr>
<tr>
<td><strong>ILLIQ</strong></td>
<td>-14.65 (-3.98)</td>
<td>31</td>
</tr>
<tr>
<td><strong>SR</strong></td>
<td>0.027 (8.19)</td>
<td>81</td>
</tr>
<tr>
<td><strong>VOL</strong></td>
<td>0.007 (1.71)</td>
<td>71</td>
</tr>
<tr>
<td><strong>CORR2000</strong></td>
<td>0.154 (17.15)</td>
<td>81</td>
</tr>
<tr>
<td><strong>CORR1000</strong></td>
<td>0.143 (15.39)</td>
<td>71</td>
</tr>
</tbody>
</table>
Statistically significant discontinuities in $IO$, $RET$, $MtB$, $TURN$, $ILLIQ$, $SR$, $CORR2000$, and $CORR1000$.

- Robust using either CV or ROT bandwidth.
- Robust across sub-sample periods.

The result on $VOL$ is only marginally significant and not robust across bandwidth choice.

Do the patterns hold across months within the year?
Monthly RD Plots

Panel A.1: RET(June)

Panel A.2: RET(July)

Panel A.3: RET(August)

Panel B.1: MtB(June)

Panel B.2: MtB(July)

Panel B.3: MtB(August)

Panel C.1: TURN(June)

Panel C.2: TURN(July)

Panel C.3: TURN(August)
Monthly RD Plots

Panel D.1: ILLIQ(June)

Panel D.2: ILLIQ(July)

Panel D.3: ILLIQ(August)

Panel E.1: SR(June)

Panel E.2: SR(July)

Panel E.3: SR(August)

Panel F.1: VOL(June)

Panel F.2: VOL(July)

Panel F.3: VOL(August)
Rules and Regression Discontinuities in Asset Markets

Yen-Cheng Chang & Harrison Hong

Monthly RD Plots

INTRODUCTION

EMPIRICAL RESULTS

CONCLUSION

Preview

Data and Summary Statistics

Aggregate Results

Results by Month

Lower-end Cutoff

INTRODUCTION
### INTRODUCTION

### EMPIRICAL RESULTS

#### Data and Summary Statistics

#### Aggregate Results

#### Results by Month

#### Lower-end Cutoff

<table>
<thead>
<tr>
<th></th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>No. Mons with $t &gt; 1.96$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RET</td>
<td>0.210</td>
<td>-0.010</td>
<td>0.004</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(10.43)</td>
<td>(-1.18)</td>
<td>(0.47)</td>
<td></td>
</tr>
<tr>
<td>MtB</td>
<td>2.210</td>
<td>1.847</td>
<td>1.775</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(3.51)</td>
<td>(2.75)</td>
<td>(2.42)</td>
<td></td>
</tr>
<tr>
<td>TURN</td>
<td>0.151</td>
<td>0.122</td>
<td>0.099</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>(6.96)</td>
<td>(9.19)</td>
<td>(8.30)</td>
<td></td>
</tr>
<tr>
<td>ILLIQ</td>
<td>-10.941</td>
<td>-9.328</td>
<td>-10.132</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>(-4.81)</td>
<td>(-4.15)</td>
<td>(-4.12)</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>0.030</td>
<td>0.032</td>
<td>0.031</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>(8.06)</td>
<td>(7.90)</td>
<td>(7.72)</td>
<td></td>
</tr>
<tr>
<td>VOL</td>
<td>0.009</td>
<td>-0.000</td>
<td>-0.002</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(2.12)</td>
<td>(-0.12)</td>
<td>(-1.46)</td>
<td></td>
</tr>
<tr>
<td>CORR2000</td>
<td>0.121</td>
<td>0.153</td>
<td>0.175</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>(7.39)</td>
<td>(8.25)</td>
<td>(8.33)</td>
<td></td>
</tr>
<tr>
<td>CORR1000</td>
<td>0.127</td>
<td>0.090</td>
<td>0.133</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>(6.14)</td>
<td>(4.57)</td>
<td>(6.98)</td>
<td></td>
</tr>
</tbody>
</table>
Raw (unsmoothed) Plots

Panel A: IO

Panel B: RET

Panel C: MtB

Panel D: TURN

Panel E: ILLIQ

Panel F: SR
Raw (unsmoothed) Plots

Panel G: VOL

Panel H: CORR2000
INTRODUCTION
EMPIRICAL RESULTS
CONCLUSION

RD (smoothed) Plots

Panel A: IO
Panel B: RET
Panel C: MtB
Panel D: TURN
Panel E: ILLIQ
Panel F: SR

Yen-Cheng Chang & Harrison Hong  Rules and Regression Discontinuities in Asset Markets
## RD Tests

<table>
<thead>
<tr>
<th></th>
<th>CV Bandwidth</th>
<th>ROT Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2005-2008</td>
<td>$h_L$</td>
</tr>
<tr>
<td><strong>IO</strong></td>
<td>-0.065</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>(-1.90)</td>
<td></td>
</tr>
<tr>
<td><strong>RET</strong></td>
<td>0.008</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>(0.82)</td>
<td></td>
</tr>
<tr>
<td><strong>MtB</strong></td>
<td>0.087</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td><strong>TURN</strong></td>
<td>-0.037</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>(-1.62)</td>
<td></td>
</tr>
<tr>
<td><strong>ILLIQ</strong></td>
<td>-7.267</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>(-0.85)</td>
<td></td>
</tr>
<tr>
<td><strong>SR</strong></td>
<td>-0.004</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>(-0.47)</td>
<td></td>
</tr>
<tr>
<td><strong>VOL</strong></td>
<td>-0.004</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>(-0.63)</td>
<td></td>
</tr>
<tr>
<td><strong>CORR2000</strong></td>
<td>-0.006</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>(-0.20)</td>
<td></td>
</tr>
</tbody>
</table>
Implications

- **Pricing and Liquidity**
  - We provide a cleaner test with results consistent with prior studies.

- **Sources of short interest**
  - Stocks with higher SR and valuation do not experience low returns following reconstitution.

- **Excessive asset price comovement**
  - Consistent with Barberis, Shleifer, and Wurgler (2005) and Wurgler (2005).
    - Foucault, Sraer, and Thesmar (2009)

- **Methodology**
Average SR and IO

**INTRODUCTION**

**EMPIRICAL RESULTS**

**CONCLUSION**
Average SR and IO

**INTRODUCTION**

**EMPIRICAL RESULTS**

**CONCLUSION**

Average SR and Size Decile

Average IO and Size Decile
Size Portfolio Performance

Size Portfolio Performance

Russell Index Performance

Cumulative Return

0 10 20 30 40 50 60 70 80 90


Cumulative Return

Russell 2000
Russell 1000

0 1 2 3 4 5 6 7 8


Yen-Cheng Chang & Harrison Hong  Rules and Regression Discontinuities in Asset Markets
Average VOL and CORR

INTRODUCTION

EMPIRICAL RESULTS

CONCLUSION

Yen-Cheng Chang & Harrison Hong  Rules and Regression Discontinuities in Asset Markets
Average CORR

Average CORR2000

Average CORR3000
Indexing induces RD experiments that deepen our understanding of how forced buying/demand shocks influence asset markets.

Compared with previous studies, the quantitative and transparent nature of Russell 2000 allows for the RD approach.

This approach may be applied to different quantitative portfolio rules.