Rules and Regression Discontinuities in Asset Markets

Yen-Cheng Chang

Shanghai Advanced Institute for Finance

Harrison Hong

Princeton University and NBER

First Draft: March 25th, 2011

This Draft: Sep 10th, 2011

Abstract: We show that institutional portfolio rules induce regression discontinuity experiments in asset markets using the popular Russell stock market indices. Stocks are ranked each June on their market capitalization from 1 (largest) to 3000 (smallest) and those ranked just above the 1000 cut-off are in the Russell 1000, while those just below are in the Russell 2000. Since the indices are value-weighted, smaller stocks just below the 1000 cut-off are heavily weighted in Russell 2000 and receive forced index buying. Larger ones just above the cut-off have negligible weight in Russell 1000 and are neglected by institutions. Smaller just-included stocks have discontinuously and significantly higher institutional ownership, price, liquidity, short interest, and market co-movement compared to just-excluded larger ones (see Figure 1).

Hong acknowledges support from an NSF grant. We thank David Lee and Jeremy Stein, seminar participants and discussants at the Princeton-Lausanne EPFL Conference, National Taiwan University and the Swedish Institute for Finance Research (SIFR) for very helpful comments. We also thank Russell Investments for providing data and He Ai for her research assistance.

I. Introduction

Understanding how exogenous demand shocks due to forced buying or selling impact asset markets is a key issue in financial economics. The simplest form of the efficient markets hypothesis posits that price equals fundamental value, which is defined as the expected cash flow of the asset discounted at an appropriate risk-adjusted rate. In contrast, important recent work on asset price bubbles and crashes, financial market contagion or excessive asset price co-movement, and liquidity and fire-sales are built on the notion that exogenous demand shocks such as speculative frenzies or crash fears result in a deviation of price from fundamental value (see Gromb and Vayanos (2010)). Recent evidence, such as Long-Term Capital Asset Management during the 1998 Asian crisis and the investment banks during the Great Financial Crisis of 2008, point to the pivotal role of institutional trades in moving prices.

In this paper, we propose a new set of regression discontinuity experiments associated with quantitative portfolio rules and institutional investors in financial markets that fundamentally expand and at the same time deepen our understanding of how demand shocks influence asset markets. Quantitative portfolio rules, defined as the buying or selling of assets based mechanically or transparently on some price or fundamental signals, have proliferated in the last thirty years with the rise of institutional investors. Such rules include the exclusion of bonds from the portfolios of many institutions, such as insurance companies and pensions who control trillions of dollars, if their credit ratings are below investment grade. Another is the popular Russell stock market indices that are constructed based on a firm's market capitalization and stocks in these indices are widely held by institutions including equity index funds and active managers who have to track them for benchmarking purposes.

Our basic idea is that popular quantitative portfolio rules adopted by institutional investors induce regression discontinuity experiments since these rules typically have discontinuities in terms of what gets bought and what gets sold. While the regression discontinuity approach has been widely used in other areas of economics, particularly in labor economics or program evaluation, the use of this approach in financial markets and in particular to look at asset prices is novel (see Lee and Lemieux (2010)).

We show that institutional portfolio rules induce regression discontinuities in asset markets using the popular Russell equity indices. Stocks are ranked each June on their (float-adjusted) market capitalization from 1 (largest) to 3000 (smallest) and those ranked just above the 1000 cut-off are in the Russell 1000, while those just below are in the Russell 2000. The indices are value-weighted. As a result, smaller stocks just below the 1000 cut-off receive significant weight in the Russell 2000 index and hence forced institutional buying. In contrast, those just above the 1000 cut-off receive negligible weight in the Russell 1000 index and hence are neglected by institutions.

Our exclusion restriction is that where stocks land in terms of their market capitalization around the 1000 cut-off is random and as such we can employ a regression discontinuity approach to investigate the impact of having been just included in the Russell 2000 (e.g. stocks ranked 1001 to 1010) versus stocks just excluded from the Russell 2000 and in the bottom end of the Russell 1000 (e.g. stocks 990 to 1000). Demand is proxied by a stock's institutional ownership on and subsequent to the reconstitution date. To the extent that our exclusion restriction is valid, we can then see how other variables of interest such as price, stock price volatility or correlation with the stock market index behave around this 1000 cut-off. We then can make causative inferences and calculate elasticities related to how dependent variables of interest respond to a given demand shock.

Indeed, we show that smaller just-included stocks have discontinuously higher institutional ownership, price, liquidity, short interest, and market co-movement compared to just-excluded larger stocks. A simple eyeball of the plots of even the raw data of institutional ownership (Panel A) and returns (Panel B) during the reconstitution month of June given in Figure 1 show the dramatic discontinuity. The difference in institutional ownership, a proxy for demand by institutions between the just-included versus the just-excluded for instance is around 32%. The mean institutional ownership percentage in our sample is 60%. So the difference is about one-half of this mean, which is a sizeable increase. This finding verifies the premise of our experiment that there is a significant difference in demand coming from the neglect of the bottom of the Russell 1000 stocks and the tracking of the top of the Russell 2000 index stocks. Moreover, consistent with this demand pressure view, stocks to the right of the 1000 cut-off have significantly higher returns just in June, the month of the reconstitution compared to those just to the left of this cut-off, but in no other months. Depending on the estimation method, the economic effect is on the order of 20%.

Discontinuity plots with some data smoothing and break tests proposed by Lee and Lemieux (2010) all paint the same picture of an economically and statistically significant jump in variables of interest for just included stocks compared to just-excluded stocks. One also see discontinuities and find similarly large estimates for market-to-book ratios, liquidity as measured by trading activity and price impact, short interest ratio defined as shares shorted to shares outstanding, and the correlation or co-movement with the market portfolio. The one variable that does not seem to display a discontinuity is stock price volatility.

Our results are striking for a few reasons. First, the regression discontinuity approach yields better-identified estimates of how demand shocks impact asset markets than previously available. This stands in contrast to earlier studies of whether demand shocks due to portfolio indexing impact stock prices. These studies, originating with Shleifer (1986) and Harris and Guerel (1986) which remain among the most insightful natural experiments in finance and economics, typically find that stocks which get added to the Standard and Poors (S&P) 500 stock index, a universally followed gauge of the stock market, or some other widely tracked index, experience a positive, risk-adjusted announcement day return due to the forced buying associated with passive index funds. The exclusion restriction behind this experiment is the probability of inclusion is uncorrelated with the fundamentals of the company. The black box nature of the selections opens up questions about the validity of the exclusion restriction. The alternative hypothesis is that stocks chosen are "leaders" or special on some fundamental grounds and hence the price increase is capturing this unobserved heterogeneity.[1]

Second and in a similar vein, our identification also comes from smaller stocks having higher prices and liquidity, which is contrary to conventional wisdom that larger stocks tend to be more widely held by institutions. The fact that smaller stocks around the 1000 cut-off actually are much more widely held than their larger counterparts and have higher prices and a better trading environment reinforce the clean and large estimates we obtain. The magnitudes are much larger than in earlier studies. Our inclusion effect on return is around 20% during the month June of reconstitution in contrast to the typical 9% found in earlier S&P 500 inclusion studies. The comparison of neglected or orphan large stocks to included and tracked small stocks is key. We show that it is as if a fraction of the demand from institutions for stocks 990 to 1000 has been moved and given to stocks 1001 to 1100. The regression discontinuity is quite striking and sizeable. There is little effect when we look at additions of smaller stocks into the bottom end of the Russell 2000 since these stocks receive negligible weight in the Russell 2000 index.

---

[1] These experiments have been a source of vibrant area of research over the last twenty years. While there is no affirmative evidence that S&P 500 stocks are special on fundamental grounds, the exclusion restriction is not testable and hence there is scope for cleaner experiments. Earlier studies have compared the returns of any stock (and not those just around the 1000 cut-off) included in the Russell indices to stocks not included but this comparison has the same identification challenges as the S&P 500 experiment (see Wurgler (2010) for a review). A few recent works have used alternative identification strategies to deal with some of these empirical challenges. Boyer (2011) exploits an institutional feature of S&P/Barra in which stocks may move from the Value Index to the Growth Index even when their book-to-market ratio increases, and vice versa. Greenwood (2008) shows that the overweighted stocks in the Nikkei 225 have excessive comovements. Hau, Massa, and Peress (2010) show that a major MSCI Global Equity Index redefinition induces appreciation and comovement in the currency market. But our regression discontinuity experiment is unique even relative to these clever recent tests since smaller stocks are actually getting higher prices than larger stocks and it covers a large basket of stocks in which the experiment regularly repeats every June. Our experiment also touches on more variables beyond price and covariance.

Third, the magnitudes obtained are not only larger but more diverse than existing studies. To the best of our knowledge, our results on volatility, short interest, and liquidity are the first formal tests for these variables. Importantly, as we discuss below, the very clear findings, particularly in regards to excessive co-movement and lack of excessive volatility provide important stylized facts to inform recent theory work (see Barberis, Shleifer and Wurgler (2005), Basak and Pavlova (2011), Vayanos and Woolley (2010)). We will further discuss the implications of these findings in Section IV.

Fourth, we conclude by examining the extent to which the demand shifts induced by the Russell Index reconstitution each year is forecastable. This additional analysis is not possible in most other indexing experiments such as the S&P 500 ones because the inclusion is a guarded secret and not predictable.

Our paper proceeds as follows. We discuss the background on the Russell 2000 index and the validity of the regression discontinuity approach in Section II. Data and variables are described in Section III. The results are presented in Section III. We draw implications for various theories related to pricing and liquidity, shorting, excessive volatility or co-movement in Section V. We discuss the potential for arbitrage in Section VI. We conclude with thoughts about further research in Section VII.

II. Data and Validity of Regression Discontinuity Approach

The business model of the Russell Company is that their indices are transparent and easy for managers to construct themselves, in contrast to the black box approach of the S&P 500 index. This transparency has resulted in its popularity among a significant fraction of mutual fund managers. We obtain data from Russell Investments on the constituents list for the Russell 1000 and Russell 2000 for the sample period of 1991 to 2008. The broadest Russell index, Russell3000E, that enables us to identify the lower end of Russell 2000 is available from 2005-2008.

Panel A Table I reports the estimated amount of institutional money that is indexed or benchmarked to the Russell 1000 and 2000 from 2002 to 2008. We only start having formal figures in 2002 though there has always been considerable money indexed to the Russell 2000. Russell 2000 is more popular than the Russell 1000. Around 200 to 3000 billion dollars track the Russell 2000, but the Russell 1000 has increased in appeal recently.

In Panel B, we report the weights for stocks #1001-1050, which is the top of the Russell 2000 and stocks #951-1000 which is the bottom of the Russell 1000. Recall these indices are value weighted. So the portfolio index weights for the stocks in the top of the Russell 2000 are much larger than that of the Russell 1000. The average weight of the top 50 stocks in the Russell 2000 is 0.17% in contrast to the comparable figure of .005%. The average weight is about 34 times as big for the just-included Russell 2000 stocks compared to the just-excluded Russell 2000 stocks (or the bottom end of the Russell 1000 index). Since the Russell 2000 has much more money tracking it than the Russell 1000, this will yield a large demand shift if a stock went from just-excluded to just-included.

The validity of our experiment relies on the randomness of where stocks land around the rank 1000 cut-off. What is interesting about our experiment is that stocks with smaller market capitalizations will get included at the top of the Russell 2000. As such, there is little room for manipulation since investors would have to try to short a large stock that is on the cusp of being included to get its price and market capitalization lower. But there is little incentive to do this since inclusion will lead to a higher price and a loss for the short position. The stocks near the 1000 cut-off are fairly large and not easy to manipulate. As such, we can view where stocks land near the 1000 cut-off as driven by randomness in stock prices.

It makes more sense to do this at the bottom end of the Russell 2000 but as we show, the price effects are negligible---this might perhaps be because of this pre-emptive buying or manipulation. We cannot rule this out. But even if there is scope for potential manipulation, it does not invalidate our experiment as long as there is still some randomness on whether they can succeed in doing this. Lee and Lemieux (2010) make this point in their discussion of the validity of regression discontinuity experiments when there is a potential for manipulation.

The Russell rankings are generated using the following procedure. At the last trading day of each May, they rank firms by market capitalization (with no adjustment). They then perform a float adjustment (cross-ownership by other index firms, private holdings, government holdings, etc). But Russell would not provide the details of how they calculate the float nor provide their float market capitalization data. They also state that they have a qualitative procedure with which they keep firms in their indices if their market values have not changed too much. On the last Friday of June, reconstitution takes place using the weights determined by the adjusted market capitalization. These weights are used by the investment community to determine their tracking portfolios. Since they use the float-adjusted market capitalization to determine their Russell rankings, stocks with a lower Russell rank in terms of float adjusted market capitalization may actually have a higher raw market capitalization.

We have looked at how the raw market capitalization moves with the Russell rankings. Indeed, stocks in the bottom of the Russell 1000 actually have a higher raw market capitalization then those in the top of the Russell 1000. It appears that Russell tries to keep the larger stocks regardless of their float in the bottom of the Russell 1000 index.

All stock variables are from the Center for Research in Security Prices (CRSP) and Compustat. The independent variable of interest is simply the rankings of stocks on the Russell constituent list on the last Friday in June of each year. Our dependent variables of interest include the following: *IO* is institutional ownership and is observed only quarterly. *RET* is the raw monthly stock return. *MtB* is the market to book ratio. *TURN* is stock turnover, calculated as trading volume divided by shares outstanding. Trading volume on NASDAQ is adjusted using the Gao-Ritter (2010) procedure. *ILLIQ* is the Amihud (2002) illiquidity measure, measured in percentage per million-dollar volume. *SR* is the fraction of outstanding shares being shorted. *VOL* is the standard deviation in daily stock returns in a given month. *CORR2000* (*CORR1000*) is the correlation coefficient between daily stock returns and Russell2000 (Russell1000) index returns in a given month. All variables are of monthly frequency unless otherwise stated.

Table II reports the summary statistics for the 1000 firms above and below the Russell 1000 cutoff. Our analysis will focus on the behavior of these stock variables around the 1000 cut-off. As such, we ignore in our analysis very small firms in the bottom end of the Russell 2000 index.

Panel A reports the summary statistics for these 2000 firms for all periods and various sub-periods. *IO* has a median of 0.626 and has also increased significantly over time from a low of 0.505 in 1991-1996 to 0.774 in 2003-2008. The median monthly *RET* is 0.008 and the median *MtB* is 2.451. The median of *TURN* is 0.094 a month or roughly 120% a year. This figure has been steadily rising over our sample period, from 0.048 a month in the 1991-1996 sub-period to 0.174 in the 2003-2008 sub-period. The median *ILLIQ* is 0.299 and has fallen over the sample period as the stock market has become progressively more liquid. It falls from 1.243 in the 1991-1996 sub-period to 0.098 in the 2003-2008 sub-period. The median *SR* is 0.018 (i.e. shares shorted amounts to around 2% of total shares outstanding). *SR* has increased significantly over time from 0.007 in the 1991-1996 sub-period to 0.039 in the 2003-2008 sub-period. The median of daily *VOL* is 0.021 or around 30% a year. VOL does not change significantly across sub-periods. The median of *CORR2000* is 0.416 and has increased significantly from 0.258 in the 1991-1996 sub-period to 0.567 in the 2003-2008 sub-period. The figures for *CORR1000* are almost identical. This is important to note that the Russell 1000 and Russell 2000 indices are highly correlated, with a correlation coefficient of 0.86.

Panel B reports the summary statistics by the Russell 1000 and the Russell 2000 constituents. Note that we are only still reporting summary statistics for the top 1000 of the Russell 2000 firms. Interestingly, the medians and means of these two sub-samples do not differ much except for *IO* and *ILLIQ*. The medians of *IO* are similar for these two sub-samples, but the standard deviation is much higher for the Russell 2000 constituents.

Notice that *ILLIQ* for the larger Russell 1000 firms has a median of 0.091 in contrast to the median for the top 1000 of the Russell 2000, which is 0.895. Moreover, the standard deviations of these two sub-samples also differ dramatically, 10.3 for the former and 64.2 for the latter, respectively. This statistic really suggests that there is much less liquidity in the smaller firms.

III. Empirical Results

We begin our empirical analysis in Figure 2. This figure plots the dependent variables of interest against the Russell size rankings. For each of the 1000 ranking positions above and below the Russell 2000 cutoff, we compute the means of each of these variables of interest across the entire sample period. There are visible discontinuities at the 1000 cut-off even in the raw (un-smoothed) data for *IO* (Panel A), *TURN* (Panel D), *ILLIQ* (Panel E), *CORR2000* and *CORR1000* (Panels H and I).

Here, Panel A on institutional ownership speaks to the premise of our natural experiment. We expect that the just-included in the Russell 2000 index stocks (those to the right of the 1000 cut-off) to have higher institutional ownership than those just-excluded (those to the left of the 1000 cut-off). This is precisely what we see. Indeed, the economic effects for institutional ownership are dramatic. An eyeball estimate suggests a difference of around 15-20% for institutional ownership in levels.

Panel D and E also shows that there is also a dramatic increase in the liquidity as measured by *TURN* and *ILLIQ* measure for just-excluded stocks compared to just-included stocks. This is consistent with the presence of passive indexers increasing the liquidity of stocks in the index.

Panels H and I on *CORR2000* and *CORR1000* show also dramatic effects on the correlations of the just-included versus the just-excluded Russell 2000 stocks. Stocks just-excluded show dramatically lower correlations with the Russell 1000 and 2000 indices, which are essentially proxies for the market portfolio. The difference in correlation is roughly 0.15 to 0.20 lower for the just-excluded stocks compared to the just-included stocks. This finding is consistent with the view that passive indexing generates higher correlations of member stocks than other stocks not in the index.

The raw data plots of Figure 2 also show somewhat less visible discontinuities in *RET* (Panel B), *MtB* (Panel C) and *SR* (Panel F). There also seems to be no such discontinuities in *VOL* (Panel G).

To see if indeed there are such discontinuities in these other dependent variables, in Figure 3, we plot out the same dependent variables as in Figure 2 against Russell size rankings, except that results from adjacent ranking positions are then grouped into a total of 100 bins. The bin averages are then plotted with the larger firms on the LHS and smaller firms on the RHS. The vertical line at the middle of each panel indicates the index cutoff. For each panel we fit two quadratic functions using data from each side of the cutoff separately.

We begin with Panel A in Figure 3, which shows the smoothed analog of the Panel A in Figure 2. There is an obvious jump in institutional ownership for the just-included compared to the just-excluded stocks. The economic magnitudes as before appear quite substantial. The mean of the *IO* variable is around 0.63. So this discontinuity contributes to nearly 25% of the mean of *IO*. There is also an obvious jump in *TURN*. One could interpret *TURN* as a proxy for demand along the lines of institutional ownership. A standard deviation of turnover in our Summary Statistics Table II is 0.175. So this discontinuity contributes to nearly 60% of a standard deviation of this dependent variable. Needless to say, these are substantial magnitudes and speak very much to the premise of our experiment that there is a jump in demand at the Russell rank 1000 cut-off. The difference in *ILLIQ* is around 3 between just-excluded versus just-included firms, which is around 75% of the mean of the *ILLIQ* measure. It is however a smaller fraction relative to a standard deviation of *ILLIQ* since many of the small firms have extremely high *ILLIQ*.

There are also substantial discontinuities for *CORR1000* and *CORR2000* as reported in Panels H and I. The magnitudes are similar to what we eyeballed in Figure 2. A standard deviation of *CORR1000* and *CORR 2000* is around 0.3. This discontinuity contributes a jump of around 0.2, which is roughly 66% of a standard deviation of these two dependent variables of interest.

These findings are not surprising given the stark discontinuities apparent even in the raw un-smoothed data. Where this smoothed version of Figure 3 is useful is to get a sense of the discontinuity for the other dependent variables in which the jump is not as obvious. The smoothed plots show visible discontinuities in *RET* (Panel B), *MtB* (Panel C) and *SR* (Panel F). The most prominent is in *SR*, while the pricing effects seem to be modest. The jump in *SR* looks to be around a couple of percent, which is quite a significant economic effect given the small amount of short interest typically in stocks. Interestingly, notice that the *SR* of the stocks in the Russell 2000 are uniformly higher than those in the Russell 1000. As we show below, this is consistent with the rise of short interest in small stocks in the latter periods of our sample. Our analysis suggests perhaps that the popularity of the Russell 2000 over the last ten years contributed to this rise in shorting among small stocks.

There are also differences between the just-included versus the just-excluded stocks in terms of price effects, measured either through *RET* (Panel B) or *MtB* (Panel C). Just-included stocks experience higher returns and a higher price. The return difference appears to be on the order of a couple percent, which is a fairly sizeable effect. However, there seems to be no visible discontinuities in *VOL* (Panel G).

We then formally conduct a regression discontinuity test in Table III. This table reports the regression discontinuity test results. For each variable of interest and each Russell ranking position, we compute the average across the twelve months between each index reconstitution. The test results are from estimating the following linear regression around the vicinity of the Russell 1000 and Russell 2000 cutoff c:

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

where $c - h_L \leq X \leq c + h_R$, Y is the dependent variable of interest, and X is the stock's Russell size rank, D is an indicator variable for stocks in Russell 2000. The subscripts L and R correspond with the directions in Figure 2, i.e., left (right) indicates stocks in Russell 1000 (Russell 2000). Year dummies are included in all regressions.

The bandwidth h is estimated using two alternative procedures: (1) Cross-validation (CV) bandwidth, and (2) Rule-of-thumb (ROT) bandwidth. We report estimates of τ and the t-stats (in parentheses). τ then measures the estimate of the discontinuity in the dependent variable of interest around the cut-off.

The optimal cross-validation bandwidth is chosen following the "leave one out" procedure proposed by Ludwig and Miller (2007) and Imbens and Lemieux (2007). Given a bandwidth *h*, for each observation X_i on the RHS of the index cutoff, we run a linear regression using the observations $X_i < X \leq X_i + h$. If X_i is on the LHS of the index cutoff, then the regression is run using observations $X_i - h \leq X < X_i$. The mean squared difference between the observed dependant variable Y_i and the predicted value from the above regression is recorded. This procedure is repeated for each X_i and the cross-validation criterion is the summed mean squared error. Formally,

$$CV_Y(h) = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}(X_i)\right)^2 .$$

The optimal CV bandwidth is then the value of h that minimizes the above expression. Equivalently,

$$h_{CV} = \arg\min_{h} CV_Y(h) .$$

Following Fan and Gijbels (1996) and Lee and Lemieux (2010), the rule-of-thumb (ROT) bandwidth is computed using the following formula:

$$h_{ROT} = 2.702\left(\frac{\hat{\sigma}^2 R}{\sum_{i=1}^{N}\tilde{m}''(X_i)^2}\right),$$

where σ is the standard error of an regression of Y on X, R is the range of the independent variable, m'' is the second derivative of the regression, and 2.702 is a constant specific to the rectangular kernel. Following Lee and Lemieux (2010), we use a quartic specification for this regression. The bandwidth is computed separately for those above and below the index cutoff.

The discontinuity is robust to using either CV or ROT, both in statistical and economic significance. Table III reports the results using data from the whole sample and also sub-periods of 1991-1996, 1997-2002, and 2003-2008. Our theory is silent regarding the implications of how these magnitudes should change over time. On the one hand, the amount of money tracking the Russell 2000 index has risen over time substantially. So all else equal one might expect stronger effects. On the other hand, the stock market has also gotten significantly larger and more liquid over time as well. So such indexing pressures might be attenuated later in our sample. Also, there is no way to tell a strong causal story with looking at time changes. So we view the sub-period analyses as affording us a measure of the robustness of our discontinuity effects. Note that the economic magnitudes here can potentially be different from our eyeball estimates from the previous figures because we are performing a local linear kernel fit of the regression discontinuity. We want to use these alternative methods to gauge the robustness of our effects.

Looking at the first column under CV bandwidth, we see that there are significant jumps in *IO*, *RET*, *MtB*, *TURN*, *ILLIQ*, *SR*, *CORR2000*, and *CORR1000*. The jump in *VOL* is only marginally significant. Looking at the first column of ROT bandwidth, we see very similar results in terms of both economic and statistical significance except for

VOL, which is statistically insignificant. Notice, however, that the economic estimates using ROT are typically smaller in absolute magnitudes than for CV Bandwith. The difference in magnitudes varies depending on the dependent variable of interest. For *IO*, *RET*, *MtB*, *SR*, *CORR2000*, and *CORR1000*, the differences are small. But for *TURN* and *ILLIQ*, the difference is larger. Finally, note that the economic and statistical significance is present for most of the sub-periods. Our effects are really in each sub-period and not driven by a particular year or time span. This greatly increases our confidence in the results. As we alluded to earlier, we hesitate to make a strong causal interpretation since our theory is silent on these time changes unless we have much more precise estimates of how demand elasticity has changed.

Finally, in Table III, we also report the bandwidths chosen using these two methods of CV versus ROT. The bandwidths are optimally chosen to better fit a linear regression over two sides of the cutoff separately. The methods will choose a smaller number (smaller bandwidth) if the data is highly nonlinear. For example, one can see from the Figure 3 that *ILLIQ* is much more linear on the Russell 2000 side, hence the larger CV bandwidth on the right of the cutoff, h_R=141, compared to on the left of the cutoff, h_L=31. One sees a similar result in this instance using ROT. The same principles apply for the other dependent variable of interest.

To see this more clearly, in Figure 4, we plot out the optimal CV bandwidths for a dependent variable of interest *MtB* as a case in point. The CV in effect measures how well local linear regressions can fit the data on each side of the data cutoff. In this example, the CV value for data on the right of the cutoff decreases in value as we increase the bandwidth before sharply increasing. On the left side of the data CV is generally decreasing as we increase bandwidth. One can see this analogy in Figure 3. In Panel D, the fitted quartic function on the right of the cutoff show a somewhat linear pattern for a short range, while that on the left has a linear patter for a much larger range. One point worth mentioning here is previous RD literature generally suggests limiting the data on either side of the cutoff for CV computation since the RD design is inherently a local estimation. Imbens and Lemiux (2007) suggest using 50% of data on either side of the cutoff, while Ludwig and Miller (2007) suggest 5%. Thus, in this paper we restrict our CV bandwidth search to a maximum of 150.

Up to this point, we have pooled all the months of the year together in presenting our results in Figures 2 and 3 and Table III. In Figures 5 and 6 and in Table IV, we focus on the first three months following the index reconstitution in June of each year, i.e. June, July and August. The idea here is that some of these discontinuity effects are likely to be larger closest to the reconstitution month each year. For instance, while a variable like *CORR2000* might display discontinuities through out the year since passive funds hold the stock for the entire year, a variable like *RET* is likely to exhibit a big effect only on the reconstitution month when much of the buying is likely to happen. The analysis of *IO* is excluded from this month-by-month analysis since we only have quarterly data for *IO*.

Figure 5 plots the unsmoothed analogs of Figure 2 for the months of June, July and August separately. The patterns are very similar to Figure 2. The one interesting thing to note is that there is a striking discontinuity effect in *RET* in the month of June but none in July and August. This speaks very much to our prior that the price effects associated with the index addition is likely to be during the month of the reconstitution.

The other variables, excluding *VOL*, exhibit a persistent discontinuity pattern across the three months.

Figure 6 plots the smoother analogs of Figure 3 for the months of June, July and August separately. The patterns are very similar to Figure 3. Again, *RET* in June exhibits a striking discontinuity consistent with the prediction that just being added to the Russell 2000 yields a jump in returns as indexers buy these stocks to track. The economic effect appears to be about 20%. This economic effect is much larger than what has been observed for the S&P 500 index inclusion effect, which is on the order of around 5%. There are a couple of potential reasons for this difference in magnitudes. The first is that the Russell 2000 stocks are smaller than S&P 500 stocks and hence a given demand shock is likely to affect them more than big stocks. Moreover, part of this big spread is coming from a comparison of the stocks just at the edge---i.e. a local estimate at the discontinuity. The plots of the other variables yield a similar conclusion as Figure 5.

In Table IV, we then conduct the same formal regression discontinuity tests as in Table III, except that we run these month-by-month for each year. Rather than report all the results, we report only the results for the months of June, July and August using the CV method. In addition, we report the number of months in which the regression discontinuity test yields a statistically significant break at the cut-off. For most of the variables, this analysis suggests that the discontinuity is present in almost all of the months of the year. The two exceptions are *VOL* and *RET*. There appears to be a very small statistically jump in *VOL* for the just-included compared to the just-excluded in June. In contrast, we see formally now that the just-included exhibits a big jump in *RET* in June compared to the just-excluded. The economic estimate is around 20% with a t-statistic of 10, consistent with our eyeball estimates in Figure 6.

Finally, we repeat our analyses for the stocks around the lower end of Russell 2000; i.e., stocks ranked above and below 3000. Here our sample period is restricted to 2005-2008 because Russell3000E, which includes roughly 4000 stocks in the U.S. market and allows us to identify the firms around the lower end of Russell2000, is not available until June 2005. In Figure 7 and 8 we plot the raw and smoothed versions of the variables of interest for stocks just-included in the lower end of Russell 2000 and those just-excluded. These plots are analogs of Figure 2 and 3. As discussed before, we expect to find negligible discontinuities at this cutoff due to value-weighting of Russell indexes. Indeed, there are no visible discontinuities in the unsmoothed plots in Figure 7. The visible discontinuities in Figure 8 are marginal compared to those in Figure 3. Table V reports the formal regression discontinuity tests around the 3000 cutoff. The results are generally statistically and economically insignificant compared to those in Table III.

Overall, the economic magnitude of our results for the top-end of Russell 2000 are typically larger than those using traditional methods in the existing literature since RD design is essentially a local estimation. In this framework, when firms have imprecise control around the index cutoff, the local variation is as good as a random experiment and it is not necessary to control for other baseline covariates and fixed effects. Furthermore, Lee and Lemieux (2010) show that RD estimates can be interpreted as an average treatment effect weighted by the probability that a firm falls around the cutoff. In our context, local treatment effect will also be mechanically larger since firms at the top-end of Russell 2000 receive higher index weights.

On the contrary, typical index inclusion effect literature computes the average (abnormal) return of all firms that moved into and out of an index. Over a similar subsample of 1996-2002, Madhavan (2003) reports the Russell 2000 average addition and deletion portfolio return in June to be 9.82% and -5.11%, respectively. The RD estimate in June for *RET* using a wider ROT bandwidths is 16% (unreported), lower than the 21% using narrower CV bandwidths and more comparable with existing literature.

An alternative methodology used in this line of research is by constructing a control sample by matching on observables. Morck and Yang (2002) use firm size and industry to match up S&P 500 index member and non-member firms. They find that the average Tobin's Q for index firms is 0.74 to 1.04 higher than those non-index firms around the turn of the century. Our RD results using *MtB* are 2.48 and 1.14 using CV and ROT bandwidths. It is worth noting that researchers using observables matching (or equivalently regression control) face the usual challenge of the choice of variables to match by. Furthermore, in the limiting case where the collection of all matched observables perfectly predict treatment, it leaves the researcher with no control-treatment comparisons.

Finally, existing literature compares the change in beta for firms added into an index to gauge the effect on comovement (see, for example, Barberis, Shleifer, and Wurgler (2005) and Boyer (2011)). Our RD estimates on *CORR2000* is 0.228 for the last sample period, which roughly translates into a change in beta of 0.107[2]. Barberis, Shleifer, and Wurgler's (2005) estimate of the S&P 500 effect on comovement is a comparable 0.151.

[2] The calculation here assumes there is no change in the volatility of firms around the cutoff. This assumption is consistent with our findings and Wurgler (2010).

IV. Implications of Our Findings

Our experiments allow us to draw a number of interesting lessons, which we elaborate on below.

A. Pricing and Liquidity

The increase in price and liquidity is consistent with earlier findings, though our estimates are larger because the discontinuity method focuses on local comparisons. One interpretation here is that our estimates might be quite a bit larger than what is found in earlier studies because it is better identified.

B. Sources of Shorting Interest

Increased demand from portfolio indexing also brings with it more shorting, presumably because shares are easier to be lent and perhaps more hedging demand for shorts as the just-included stocks, while they have a higher price to fundamental, do not experience subsequently low returns following the reconstitution month of June. Our experiment provides a causal analysis that has also been missing in this literature.

C. Sources of Excessive Asset Price Co-Movement

The increased demand due to portfolio indexing not only significantly affects the return of the just-included stocks but also their covariance, though not their volatility. That demand shocks emanating from institutions might be a source of excessive co-movement fits with recent behavioral theories such as Barberis, Shelifer and Wurgler (2005) and institutional investor flows such as Basak and Pavlova (2011) and Vayanos and Woolley (2010). However, all these theories also predict excessive volatility, which we do not find. This suggests that there is are other effects associated with institutional investors. Notably, if there are other noise traders in the market, then institutions can perhaps be stabilizing as is the case in the noise traders of Delong, Shleifer, Summers and Waldmann (1990). It is interesting to observe that an experiment from Foucault, Sraer and Thesmar (2009) in the French stock market in which the entrance of retail investors due to a reform lead to higher stock price volatility but not excessive co-movement.

V. Forecastability of the Demand Shifts Associated with Russell Reconstitution

To assess the extent to which Russell reconstitution is forecastable, we conduct the following trading strategy by first forming "guess lists" of firms on each side of the index cut-off. Specifically, following Russell methodology, all firms are first ranked by their end-of-May total market capitalization and firms ranked #1-#1000 are in Russell 1000 and those ranked #1001-#3000 are in Russell 2000. The guess list for the bottom end (top end) of Russell 1000 (Russell 2000) in the June reconstitution consist of stocks in the bottom (top) 100 ranked by either the end-of-May total market capitalization or the Russell index weights[3]. The trading strategy then enters into long positions for stocks that were in last June's bottom end of Russell 1000 but no longer in the Russell 1000 guess list, and short positions for those that were in last June's top end of Russell 2000 but no longer in the Russell 2000 guess list. To the extent the guess lists mimics the

[3] As discussed earlier, Russell uses adjusted market float and prices on the last Friday of June for the annual reconstitution. The adjusted market float considers Russell index member cross-holdings, private holdings, etc.

actual rankings, the long-short portfolio would capture return reversal as demand increases for the neglect stocks and decreases for the heavily bought stocks.

Table VI reports the monthly long-short portfolio returns for the 6 months following reconstitution. To form the guess lists, Panel A uses market capitalization and Panel B uses Russell index weights. Over the whole sample period, the long-short portfolio has a marginally significant 1.30% in June using Russell index weights while that using total market capitalization is weak. This is expected since index weights are based on adjusted market float, which serves as a better predictor. The subsample analyses show that the profitability in June is concentrated only in 1997-2002. The returns from other subsequent months are mostly not significantly different from zero. In unreported tests, we use an alternative strategy by simply buying the Russell 2000 and selling the Russell 1000 guess lists. The results are qualitatively similar.

Overall, the results suggest that the precise ranking upon reconstitution is not easily forecastable. This may reflect the difficulty to predict one-month-ahead stock prices and how Russell adjust market float. In addition, this further confirms the validity of our RD design such that where firms rank around the index cutoff is close to random.

VI. Conclusion

In this paper, we show that portfolio indexing induces regression discontinuity experiments, which deepen our understanding of how forced buying or demand shocks influence asset markets. In contrast to earlier experiments, which focused on natural experiments such as inclusion into S&P500 index, our use of the quantitative and transparent nature of the Russell 2000 index is what allows for our regression discontinuity approach. This approach can be applied to other popular quantitative portfolio rules in financial markets. Indeed, the use of such rules has exploded in recent years with the advent of sophisticated trading technologies and ever deeper and more liquid financial markets. As a result, our regression discontinuity approach opens up a set of new experiments that utilize quantitative portfolio rules to gain insights into how demand shocks impact affect financial markets.

References

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283-317.

Basak, Suleyman, and Anna Pavlova, 2011, Asset prices and institutional investors, London Business School Working Paper.

Boyer, Brian H., 2011, Style-related comovement: fundamentals or labels? *Journal of Finance* 66, 307-332.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

Fan, Jianqing, and Irene Gijbels, 1996, *Local Polynomial Modelling and Its Applications* (Chapman and Hall, London; New York and Melbourne).

Foucault, Thierry, David Sraer, and David Thesmar, 2009, Individual investors and volatility, *Journal of Finance,* forthcoming.

Gao, Xiaohui, and Jay R. Ritter, 2010, The marketing of seasoned equity offerings, *Journal of Financial Economics* 97, 33-52.

Gromb, Denis and Dimitri Vayanos, 2010, Limits of arbitrage: The state of the theory, *Annual Review of Financial Economics* 2, 251-275.

Harris, Lawrence, and Eitan Guerel, 1986, Price and volume effects associated with the new S&P 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41, 815-829.

Hau, Harald, Massimo Massa, and Joel Peress, 2010, Do demand curves for currencies slope down? Evidence from the MSCI Global Index change, *Review of Financial Studies* 23, 1681-1717.

Hong, Harrison and Jeremy C. Stein, 2007, Disagreement and the Stock Market, *Journal of Economic Perspectives* 21, 109-128.

Imbens, Guido W., and Thomas Lemieux, 2007, Regression discontinuity designs: A guide to practice, *Journal of Econometrics* 142, 615-635.

Lee, David S., and Thomas Lemieux, 2010, Regression discontinuity designs in economics, *Journal of Economic Literature* 48, 281-355.

Ludwig, Jens, and Douglas L. Miller, 2007, Does head start improve children's life chances? Evidence from a regression discontinuity design, *Quarterly Journal of Economics* 122, 159-208.

Madhavan, Ananth, 2003, The Russell reconstitution effect, *Financial Analysts Journal* 59, 51-64.

Morck, Randall, and Fan Yang, 2002, The mysterious growing value of S&P 500 membership, Working Paper.

Petajisto, 2011, The index premium and its hidden cost for index funds, *Journal of Empirical Finance* 18, 271-288.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down? *Journal of Finance* 41, 579-590.

Vayanos, Dimitri and Paul Woolley, 2010, An institutional theory of momentum and reversal, London School of Economics Working Paper.

Vijh, Anand M., 1994, S&P 500 trading strategies and stock betas, *Review of Financial Studies* 7, 215-251.

Wurgler, Jeffrey, 2010, On the economic consequences of index-linked investing, in W.T Allen, R. Khurana, J. Lorsch, and G. Rosenfeld, ed.: *Challenges to Business in the Twenty-First Century: The Way Forward*, forthcoming.

Table I: Assets Benchmarked to Russell Indexes and Index Weights Around Cutoff

Panel A reports the amount of institutional assets, in billions, benchmarked to Russell 1000 and Russell 2000 each year. The products surveyed are primarily mutual funds, separate accounts, and commingled funds. Panel B reports the average index member weights around the Russell 2000 cutoff after each year's reconstitution. Firms ranked #1001-#1050 are in the top-end of Russell 2000 and firms ranked #951-1000 are in the bottom-end of Russell 1000.

Panel A							
Year	2002	2003	2004	2005	2006	2007	2008
Russell 2000	198.20	140.70	162.50	201.40	221.10	291.39	263.70
Russell 1000	47.60	37.30	66.90	90.00	146.10	172.67	168.56
Panel B							
Average Index Weight #1001-#1050	0.1667%	0.1624%	0.1572%	0.1538%	0.1469%	0.1731%	0.2131%
Average Index Weight #951-#1000	0.0049%	0.0058%	0.0065%	0.0070%	0.0076%	0.0076%	0.0061%

Table II: Summary Statistics

This table reports summary statistics of the variables in this paper. The sample period is from 1991 to 2008, covering all member stocks in the annual Russell 1000 reconstitution, and the same number of stocks from the top of Russell 2000. All variables are of monthly frequency unless otherwise stated. *IO* is quarterly institutional ownership. *TURN* is stock turnover, calculated as trading volume divided by shares outstanding. Trading volume on NASDAQ is adjusted using the Gao-Ritter procedure. *RET* is monthly stock return. *MtB* is the market to book ratio. *ILLIQ* is the Amihud (2002) illiquidity measure, measured in percentage per million dollar volume. *SR* is the fraction of outstanding shares being shorted. *VOL* is the standard deviation in daily stock returns. *CORR2000* (*CORR1000*) is the correlation coefficient between daily stock returns and Russell 2000 (Russell 1000) index returns. Panel A shows the mean, median, and standard deviation of these variables for the whole sample and three different subperiods. Panel B splits the whole sample into those in Russell 1000 and those in Russell 2000.

Panel A: Summary Statistics by Period

	Summary Statistics, All Periods						Medians (Means), Subperiods		
	25%	Median	75%	Mean	Stdev	No. Obs.	1991-1996	1997-2002	2003-2008
IO	0.444	0.626	0.776	0.599	0.227	76,192	0.505	0.620	0.774
							(0.488)	(0.591)	(0.730)
RET	-0.051	0.008	0.068	0.010	0.135	356,439	0.011	0.006	0.007
							(0.015)	(0.010)	(0.006)
MtB	1.629	2.451	4.034	3.881	5.645	300,886	2.319	2.572	2.469
							(3.656)	(4.391)	(3.595)
TURN	0.048	0.094	0.183	0.147	0.175	357,888	0.048	0.083	0.174
							(0.070)	(0.119)	(0.235)
ILLIQ	0.077	0.299	1.226	4.549	46.405	355,035	1.243	0.398	0.098
							(12.969)	(1.513)	(0.324)
SR	0.007	0.018	0.044	0.035	0.047	283,436	0.007	0.017	0.039
							(0.019)	(0.031)	(0.056)
VOL	0.014	0.021	0.031	0.025	0.018	356,352	0.018	0.025	0.020
							(0.002)	(0.030)	(0.025)
CORR2000	0.210	0.416	0.604	0.395	0.271	356,419	0.258	0.404	0.567
							(0.246)	(0.380)	(0.531)
CORR1000	0.193	0.404	0.594	0.382	0.276	356,419	0.241	0.389	0.550
							(0.233)	(0.369)	(0.516)

Panel B: Summary Statistics by Index, All Periods

	Russell 1000				Russell 2000			
	Median	Mean	Stdev	No. Obs.	Median	Mean	Stdev	No. Obs.
IO	0.649	0.621	0.207	39,421	0.592	0.574	0.244	36,771
RET	0.009	0.009	0.120	176,784	0.008	0.011	0.148	179,655
MtB	2.742	4.215	5.757	155,805	2.188	3.522	5.500	145,081
TURN	0.095	0.148	0.170	177,383	0.093	0.146	0.179	180,505
ILLIQ	0.091	0.719	10.309	175,356	0.895	8.287	64.212	179,679
SR	0.017	0.028	0.035	141,549	0.021	0.042	0.056	141,887
VOL	0.018	0.023	0.017	176,740	0.023	0.028	0.020	179,612
CORR2000	0.415	0.394	0.266	176,772	0.417	0.396	0.277	179,647
CORR1000	0.440	0.415	0.267	176,772	0.367	0.349	0.281	179,647

Table III: Regression Discontinuity Tests

This table reports the regression discontinuity test results. For each variable of interest and each Russell ranking position, we compute the average across the twelve months between each index reconstitution. The test results are from estimating the following linear regression around the vicinity of the Russell 1000 and Russell 2000 cutoff *c*:

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

where $c - h_L \leq X \leq c + h_R$, *Y* is the variable of interest, and *X* is the stock's Russell size rank, *D* is an indicator variable for stocks in Russell2000. The subscripts *L* and *R* correspond to the directions in Figure 2, i.e., *L* (*R*) indicates stocks in Russell1000 (Russell2000). Year dummies are included in all regressions. The bandwidth *h* is estimated using two alternative procedures: (1) Cross-validation (CV) bandwidth by Ludwig and Miller (2007) and Imbens and Lemieux (2007), and (2) Rule-of-thumb (ROT) bandwidth by Fan and Gijbels (1996). We report estimates of τ and the t-stats (in parentheses). The sample period is 1991 to 2008.

	CV Bandwidth						ROT Bandwidth					
	Whole Sample	h_L	h_R	1991-1996	1997-2002	2003-2008	Whole Sample	h_L	h_R	1991-1996	1997-2002	2003-2008
IO	0.323	36	142	0.331	0.311	0.331	0.299	99	455	0.305	0.280	0.308
	(12.94)			(8.93)	(10.46)	(9.65)	(18.33)			(13.29)	(10.87)	(12.31)
RET	0.033	97	68	0.016	0.083	0.008	0.020	272	184	0.015	0.040	0.005
	(3.07)			(4.46)	(2.04)	(1.66)	(4.79)			(6.22)	(3.44)	(2.02)
MtB	2.674	86	48	0.059	2.482	2.618	0.641	206	233	-0.720	1.137	1.761
	(3.38)			(0.07)	(3.12)	(3.62)	(1.83)			(-1.43)	(2.06)	(3.39)
TURN	0.139	48	40	0.052	0.065	0.188	0.089	154	191	0.039	0.055	0.116
	(8.42)			(6.07)	(4.51)	(9.11)	(9.88)			(7.42)	(4.79)	(6.57)
ILLIQ	-14.65	31	141	-39.210	-4.069	-0.909	-5.419	130	241	-13.094	-2.668	-0.662
	(-3.98)			(-3.71)	(-3.61)	(-6.60)	(-4.31)			(-3.72)	(-5.13)	(-6.77)
SR	0.027	81	132	0.009	0.018	0.056	0.020	161	235	0.004	0.011	0.035
	(8.19)			(2.35)	(3.53)	(7.75)	(7.65)			(1.34)	(2.87)	(7.41)
VOL	0.007	71	30	-0.001	0.010	-0.002	-0.001	189	181	-0.001	0.001	-0.002
	(1.71)			(-0.93)	(0.90)	(-1.37)	(-0.87)			(-1.91)	(0.20)	(-1.36)
CORR2000	0.154	81	149	0.051	0.147	0.228	0.141	151	474	0.061	0.119	0.207
	(17.15)			(4.34)	(10.36)	(14.67)	(17.70)			(6.31)	(10.03)	(15.80)
CORR1000	0.143	71	146	0.043	0.121	0.211	0.122	163	356	0.041	0.089	0.185
	(15.39)			(3.17)	(8.52)	(14.26)	(14.61)			(4.73)	(7.46)	(15.18)

Table IV: Monthly Regression Discontinuity Tests

This table reports the regression discontinuity tests by month. For each month after the annual index reconstitution, and for each Russell ranking position, we compute the average of each variable of interest across the sample years. The regressions used here are the same as in Table III, using CV bandwidths. For brevity, only results from the first three months after index reconstitution are presented. The last column reports the number of months with significant t-stats. The sample period is from 1991 to 2008.

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

Table V: Regression Discontinuity Tests for Russell 2000 Lower End Cutoff

This table reports the regression discontinuity test results. For each variable of interest and each Russell ranking position, we compute the average across the twelve months between each index reconstitution. The test results are from estimating the following linear regression around the Russell2000 lower-end cutoff *c*:

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

where $c - h_L \leq X \leq c + h_R$, *Y* is the variable of interest, and *X* is the stock's Russell size rank, *D* is an indicator variable for stocks out of (i.e. smaller than) Russell 2000. The subscripts *L* and *R* correspond to the directions in Figure 7, i.e., *L* indicates stocks in Russell 2000 and *R* for stocks smaller than Russell 2000. Year dummies are included in all regressions. The bandwidth *h* is estimated using two alternative procedures: (1) Cross-validation (CV) bandwidth by Ludwig and Miller (2007) and Imbens and Lemieux (2007), and (2) Rule-of-thumb (ROT) bandwidth by Fan and Gijbels (1996). We report estimates of τ and the t-stats (in parentheses). The sample period is 2005 to 2008.

	CV Bandwidth			ROT Bandwidth		
	2005-2008	h_L	h_R	2005-2008	h_L	h_R
IO	-0.065	150	148	-0.037	290	203
	(-1.90)			(-1.56)		
RET	0.008	132	105	-0.001	405	390
	(0.82)			(-0.17)		
MtB	0.087	123	149	0.469	351	276
	(0.16)			(1.08)		
TURN	-0.037	98	150	-0.008	772	324
	(-1.62)			(-0.66)		
ILLIQ	-7.267	147	78	0.018	348	345
	(-0.85)			(0.00)		
SR	-0.004	62	139	-0.006	364	228
	(-0.47)			(-1.31)		
VOL	-0.004	121	150	-0.000	729	453
	(-0.63)			(-0.04)		
CORR2000	-0.006	112	148	-0.018	234	453
	(-0.20)			(-1.05)		

Table VI: Demand Forecastability

This table reports monthly long-short portfolio returns, in percentages. Each year in May, firms are ranked using their end-of-month market capitalization (Panel A) or index weights (Panel B) to determine "guess lists" of the 100 positions in the top end of Russell 2000 and bottom end of Russell 1000 for the following months' reconstitution. The long positions consists of stocks that were in last June's bottom 100 of Russell 1000 but not in the "guess list". The short positions are the stocks that were in last June's top 100 of Russell 2000 but not in the "guess list". All returns are equal weighted. For each long-short portfolio the monthly returns are reported for the 6 months following index reconstitution. The sample period is from 1991 to 2008. The t-stats are in parentheses.

Panel A						
	Jun	Jul	Aug	Sep	Oct	Nov
Whole Sample	0.78	-0.11	-0.03	-0.05	-0.39	0.95
	(1.47)	(-0.15)	(-0.05)	(-0.11)	(-0.88)	(1.36)
1991-1996	0.01	-0.17	-0.07	-0.00	0.22	0.73
	(0.01)	(-0.12)	(-0.09)	(-0.01)	(0.38)	(0.64)
1997-2002	2.44	-0.42	0.55	-0.86	-0.91	1.33
	(2.54)	(-0.29)	(0.39)	(-1.49)	(-0.92)	(1.60)
2003-2008	-0.25	0.26	-0.57	0.74	-0.38	0.75
	(-0.45)	(0.23)	(-1.00))	(1.03)	(-0.57)	(0.45)
Panel B						
	Jun	Jul	Aug	Sep	Oct	Nov
Whole Sample	1.30	-0.50	0.07	-0.80	-0.17	0.99
	(1.79)	(-0.47)	(0.09)	(-1.32)	(-0.29)	(1.02)
1991-1996	0.80	0.89	-0.99	-0.08	0.82	2.54
	(1.13)	(0.47)	(-1.01)	(-0.06)	(0.98)	(1.54)
1997-2002	3.25	-0.86	0.77	-0.23	-0.57	1.02
	(2.21)	(-0.36)	(0.59)	(-2.86)	(-0.38)	(0.98)
2003-2008	0.03	-0.51	0.09	-0.20	0.19	-1.42
	(0.03)	(-0.34)	(0.06)	(-0.29)	(-0.38)	(0.09)

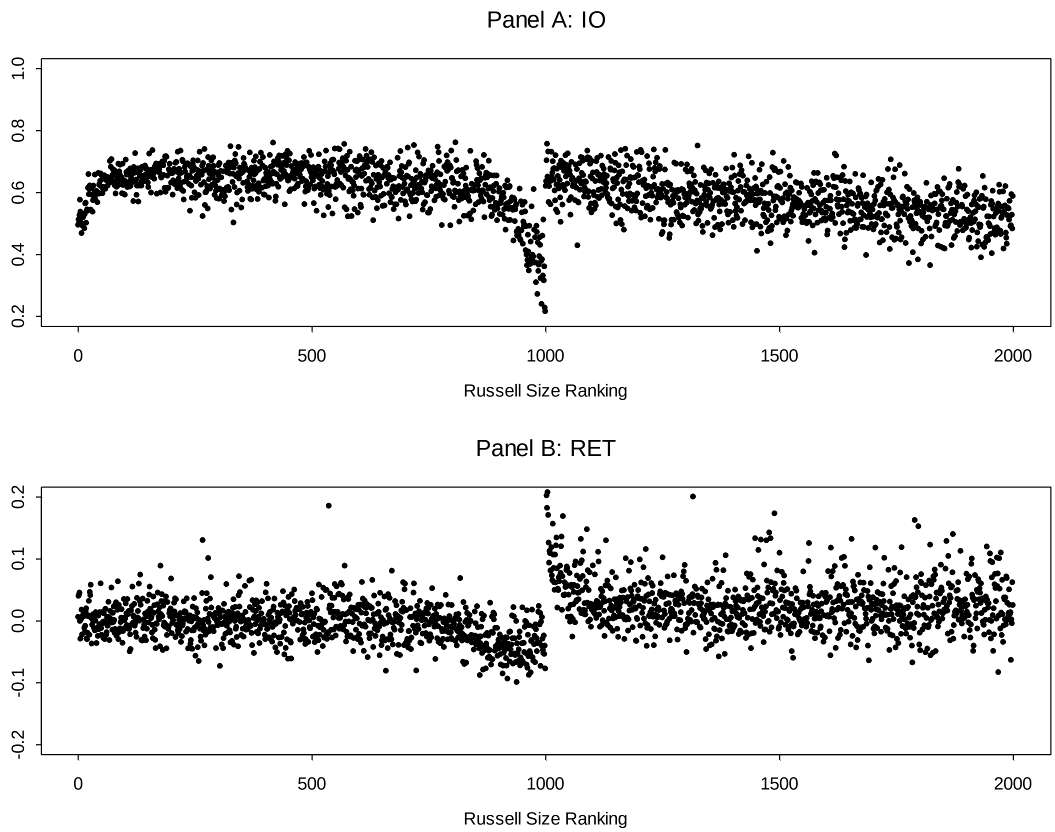
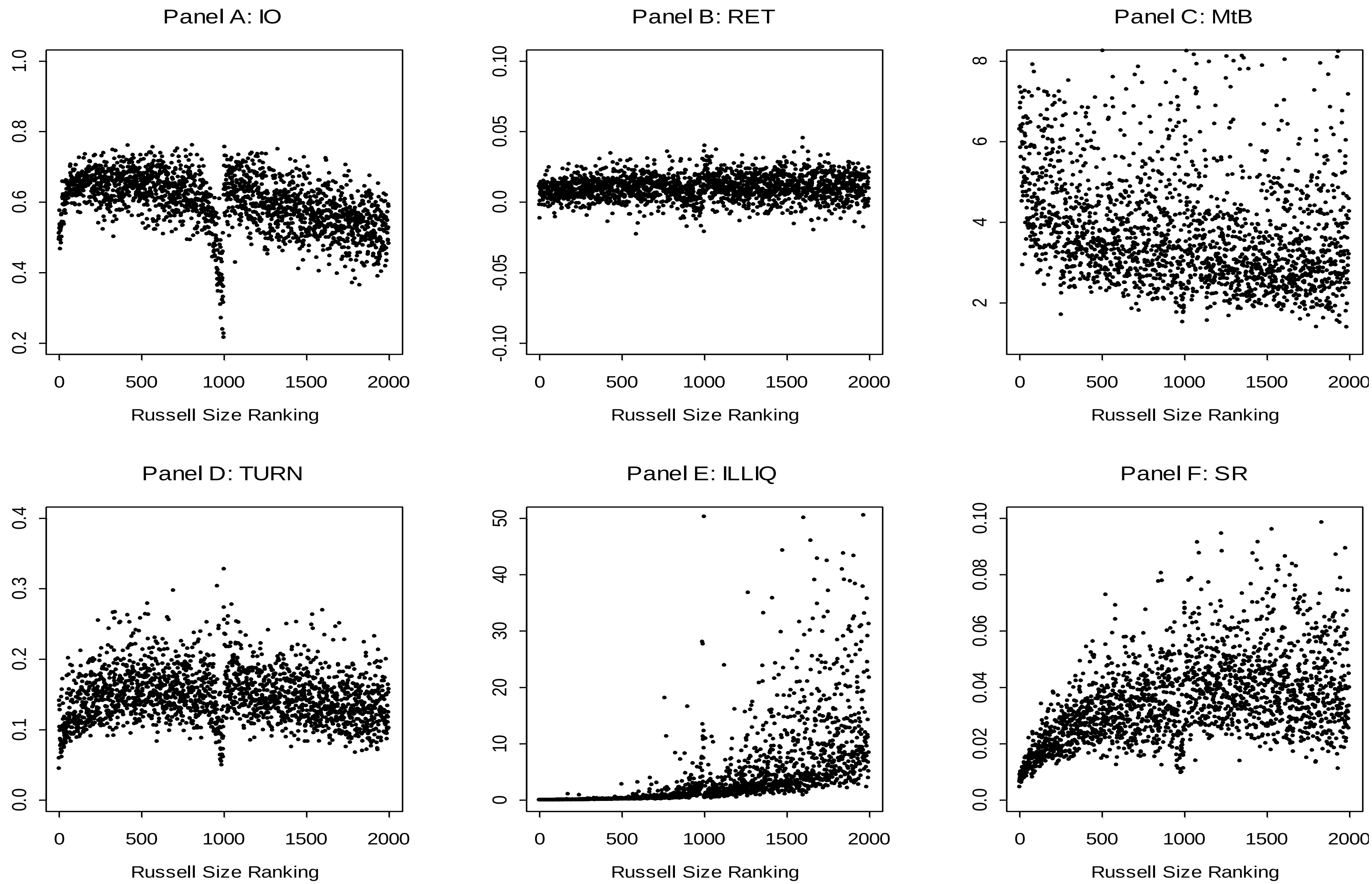


Figure 1. Discontinuity in Around Index Cutoff. This figure plots the June return and institutional ownership of Russell member stocks against size rankings. We compute the means of the institutional ownership (Panel A) and monthly return in June (Panel B) across the sample years for each of the 1000 ranking positions above and below the Russell 1000 & Russell 2000 cutoff. The sample period is from 1991 to 2008.


Panel A: IO
Panel B: RET
Panel C: MtB
Panel D: TURN
Panel E: ILLIQ
Panel F: SR
Russell Size Ranking
0
500
1000
1500
2000
0.2
0.4
0.6
0.8
1.0
-0.10
-0.05
0.0
0.05
0.10
2
4
6
8
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
0.0
0.02
0.04
0.06
0.08
0.10

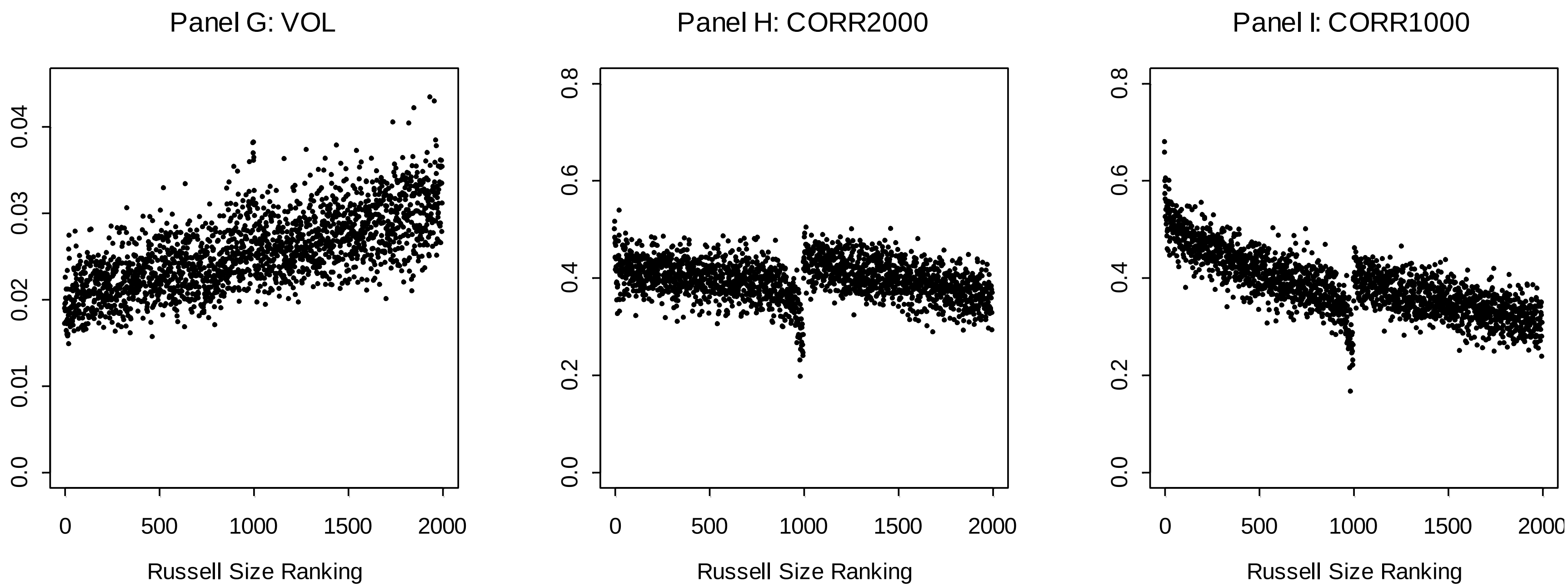
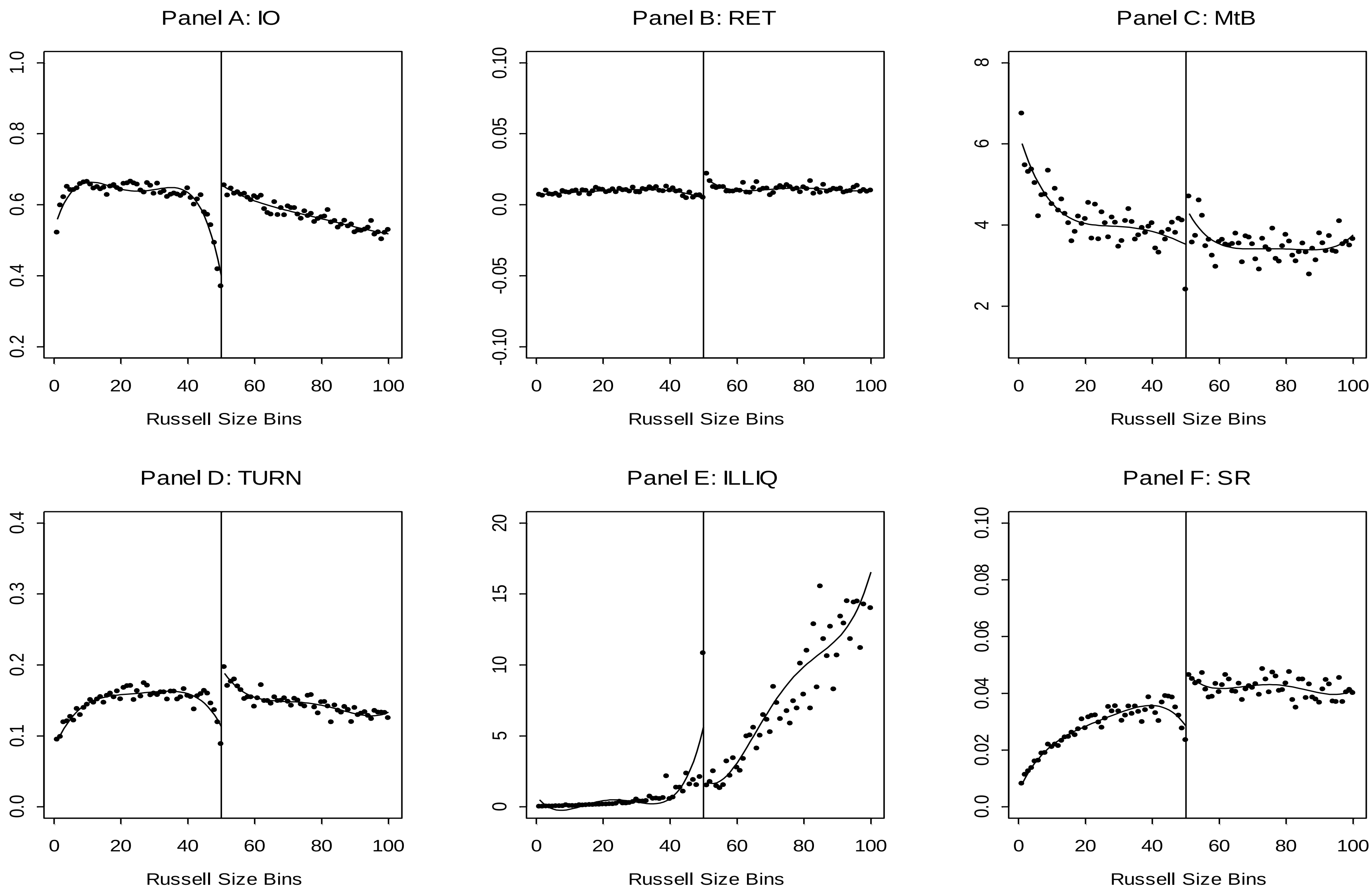


Figure 2. Discontinuity around Russell1000 & Russell2000 cutoff. This figure plots the test variables against Russell size rankings. For each of the 1000 ranking positions above and below the Russell 1000 & Russell 2000 cutoff, we compute the means of each variable of interest across the sample years and the twelve months following each annual index reconstitution. The data is arranged with larger stocks on the LHS of each plot. The sample period is from 1991 to 2008.


Panel A: IO
Panel B: RET
Panel C: MtB
Panel D: TURN
Panel E: ILLIQ
Panel F: SR
Russell Size Bins
0
20
40
60
80
100

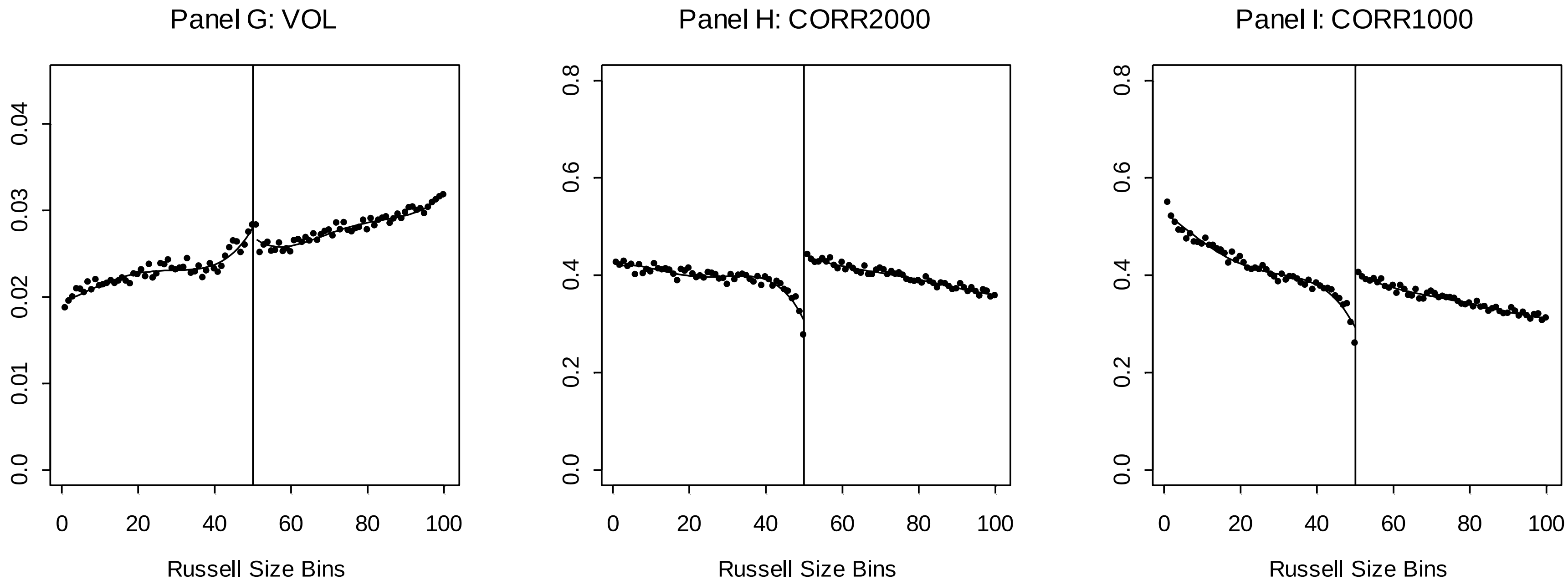


Figure 3. Discontinuity around Russell1000 & Russell2000 cutoff. This figure plots the test variables against Russell size rankings. For each of the 1000 ranking positions above and below the Russell 1000 & Russell 2000 cutoff, we compute the means of each variable of interest across the sample years and the twelve months following each annual index reconstitution. Results from adjacent ranking positions are then grouped into a total of 100 bins. The bin averages are then plotted with the larger firms on the LHS and smaller firms on the RHS. The vertical line at the middle of each panel indicates the index cutoff. For each panel we overlay two quartic functions from estimating the following regressions: $Y_B = \alpha_0 + \alpha_1 X_B + \alpha_2 X_B^2 + \alpha_3 X_B^3 + \alpha_4 X_B^4 + \varepsilon$, where Y_B is the bin average, and X_B is the bin number. The quartic functions are estimated using data from each side of the cutoff separately. The sample period is from 1991 to 2008.

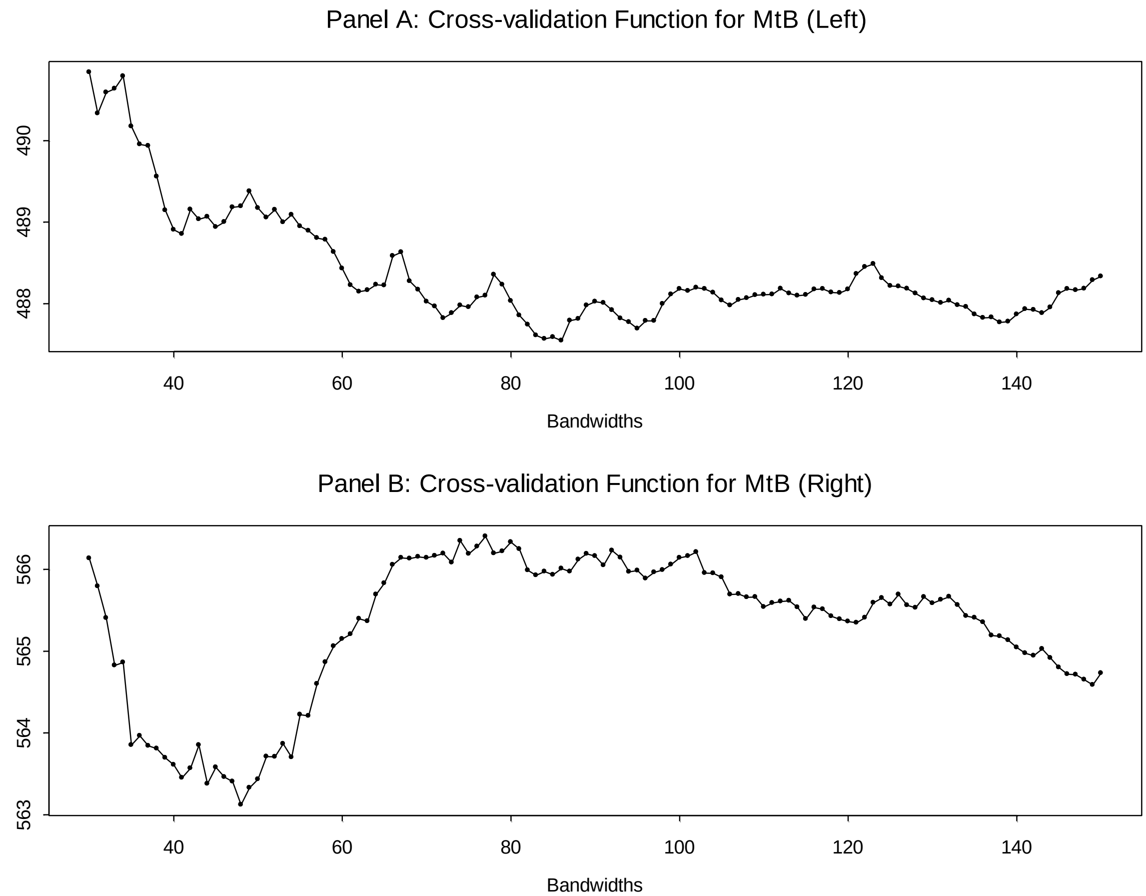
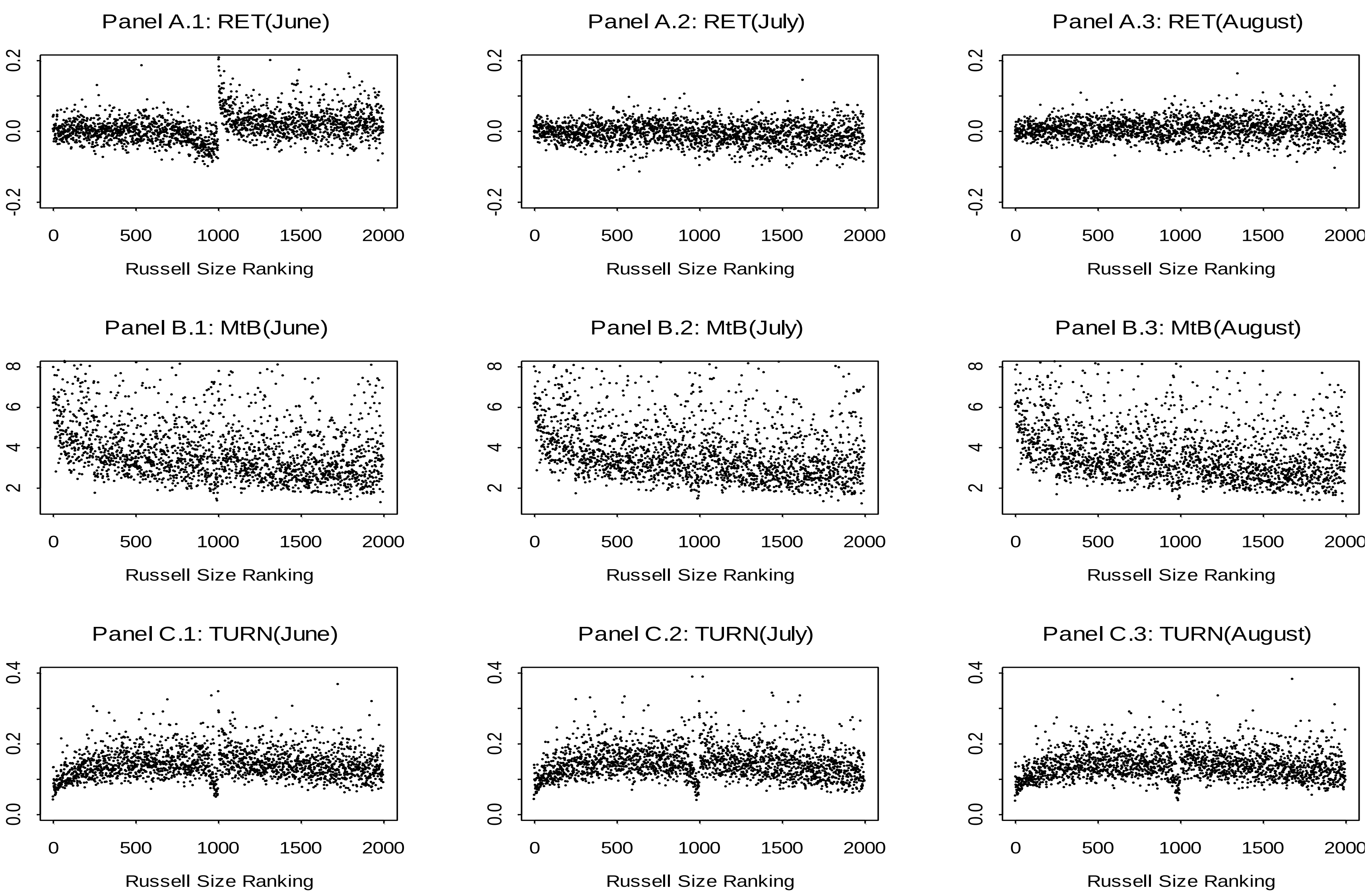
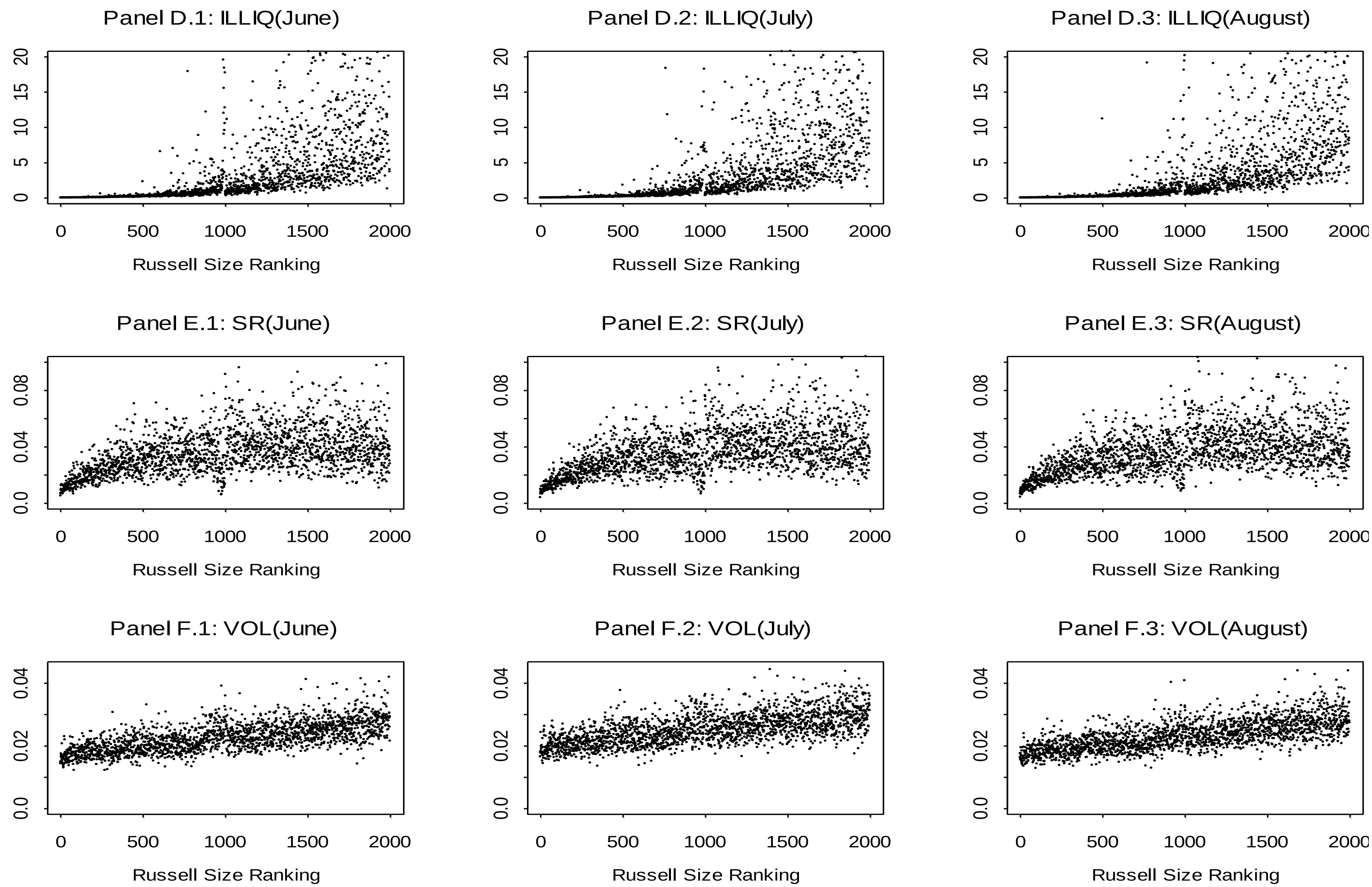


Figure 4. Optimal cross-validation bandwidths. This figure plots the cross-validation value given different bandwidths, using pooled market-to-book over the whole sample period as in Table II. The search procedure is described in the text, following Ludwig and Miller (2007) and Imbens and Lemieux (2007).


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)
Russell Size Ranking


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)
Russell Size Ranking
0 500 1000 1500 2000
0 5 10 15 20
0.0 0.04 0.08
0.0 0.02 0.04

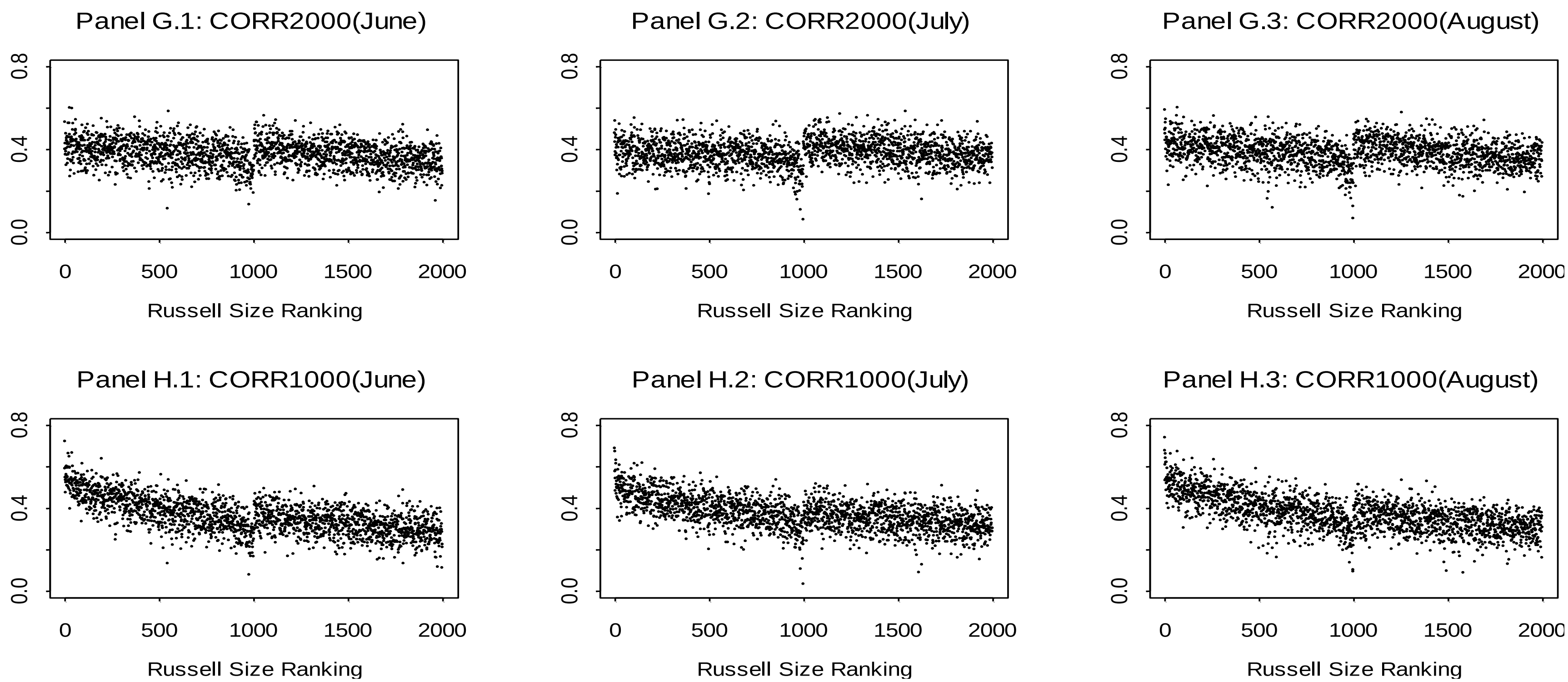
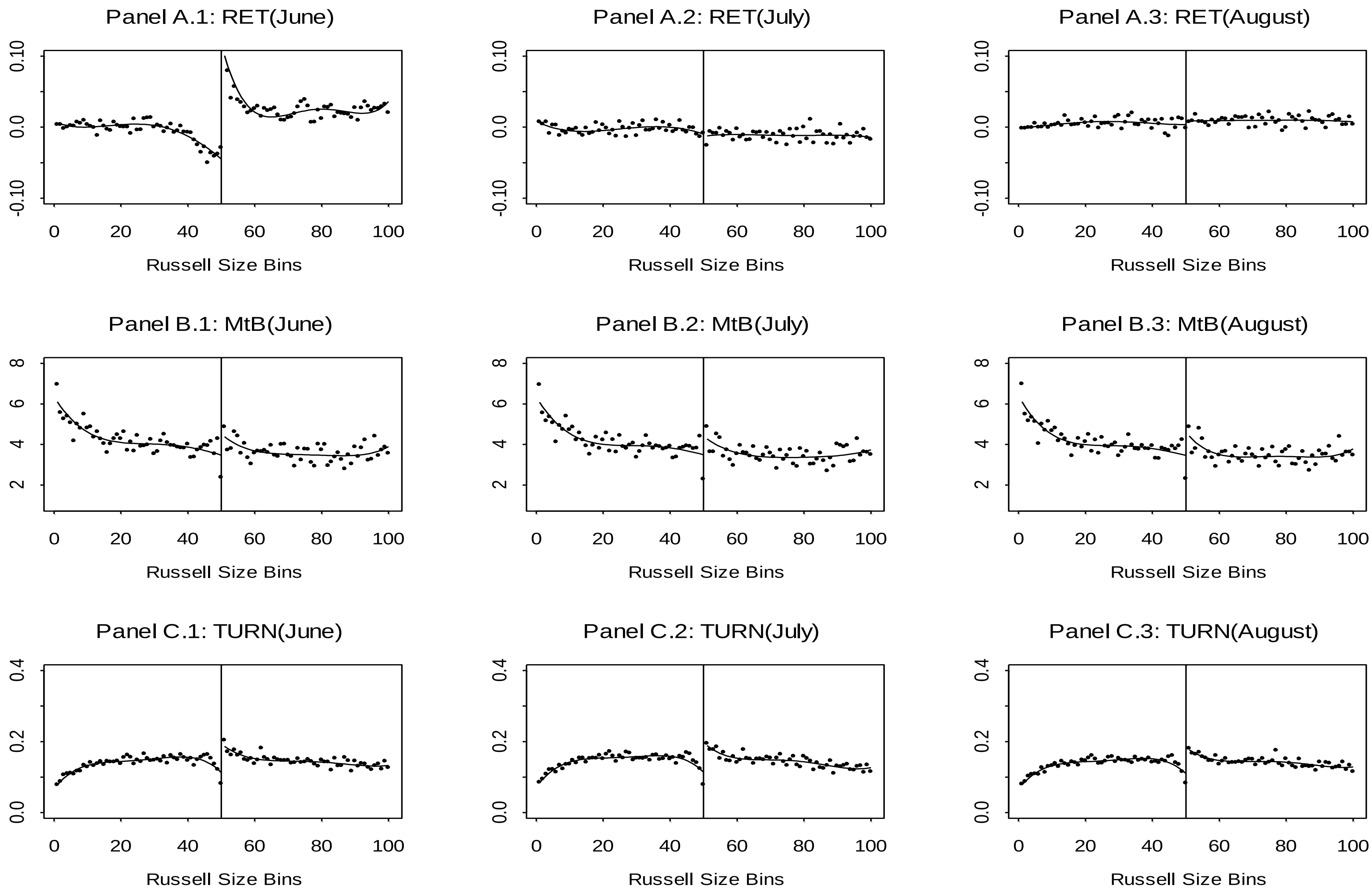
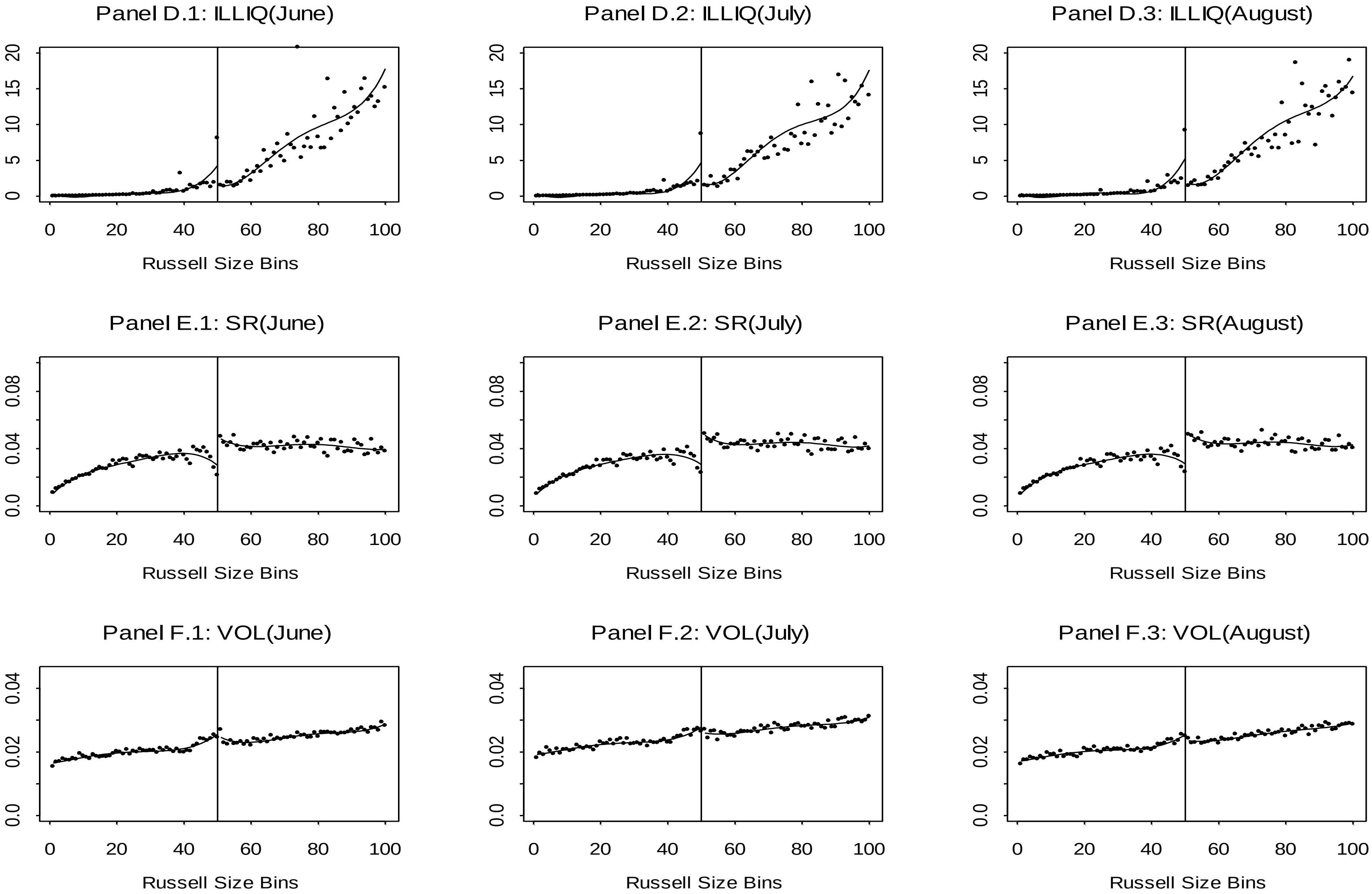


Figure 5. Monthly discontinuity around Russell1000 & Russell2000 cutoff. This figure plots the test variables against Russell size rankings. For each of the 1000 ranking positions above and below the Russell 1000 & Russell 2000 cutoff, we compute the means of each variables of interest across the sample years for each month. The sample period is from 1991 to 2008.



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)
Russell Size Bins
-0.10
0.0
0.10
2
4
6
8
0.0
0.2
0.4
0
20
40
60
80
100


Panel D.1: ILLIQ(June)
Panel D.2: ILLIQ(July)
Panel D.3: ILLIQ(August)
0
5
10
15
20
Panel E.1: SR(June)
Panel E.2: SR(July)
Panel E.3: SR(August)
0.0
0.04
0.08
Panel F.1: VOL(June)
Panel F.2: VOL(July)
Panel F.3: VOL(August)
0.0
0.02
0.04
0
20
40
60
80
100
Russell Size Bins

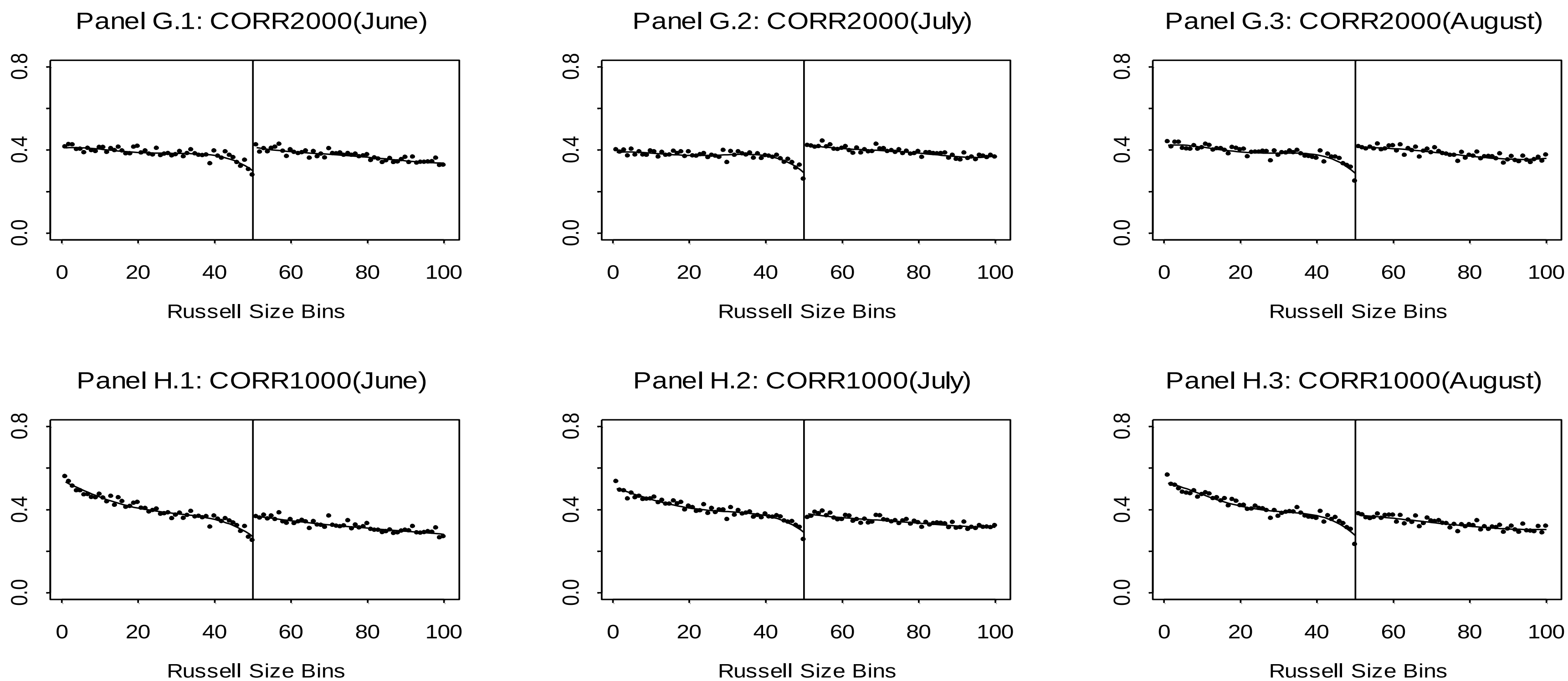
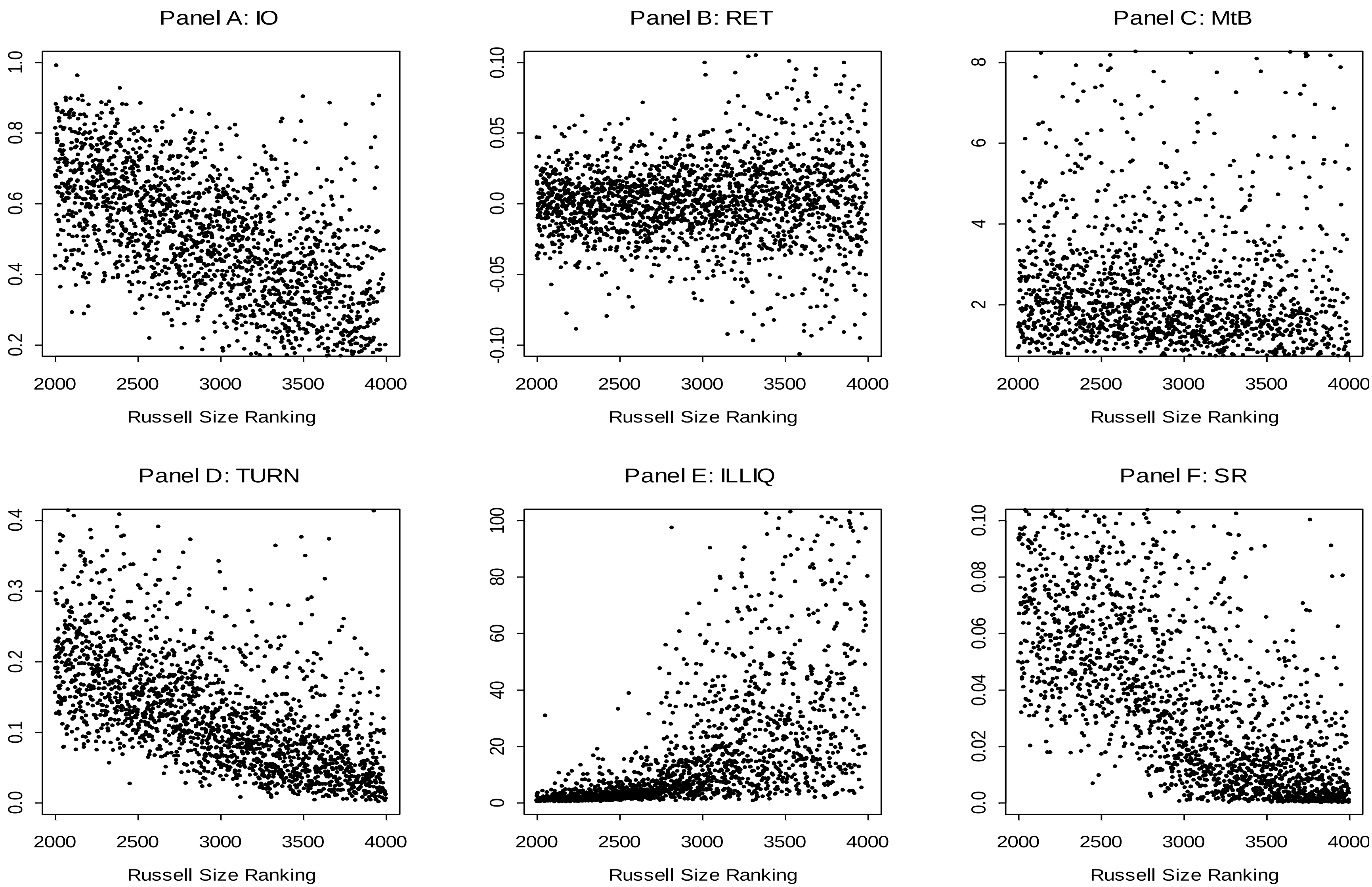


Figure 6. Monthly discontinuity around Russell1000 & Russell2000 cutoff. This figure plots the test variables against Russell size rankings. For each of the 1000 ranking positions above and below the Russell 1000 & Russell 2000 cutoff, we compute the means of each variables of interest across the sample years for each month. Results from adjacent ranking positions are then grouped into a total of 100 bins. The bin averages are then plotted with the larger firms on the LHS and smaller firms on the RHS. The vertical line at the middle of each panel indicates the index cutoff. For each panel we overlay two quartic functions from estimating the following regressions: $Y_B= \alpha_0+\alpha_1 X_B + \alpha_2 X_B^2+ \alpha_3 X_B^3+ \alpha_4 X_B^4+\varepsilon$, where Y_B is the bin average, and X_B is the bin number. The quartic functions are estimated using data from each side of the cutoff separately. The sample period is from 1991 to 2008.


Panel A: IO
1.0
0.8
0.6
0.4
0.2
2000
2500
3000
3500
4000
Russell Size Ranking
Panel B: RET
0.10
0.05
0.0
-0.05
-0.10
2000
2500
3000
3500
4000
Russell Size Ranking
Panel C: MtB
8
6
4
2
2000
2500
3000
3500
4000
Russell Size Ranking
Panel D: TURN
0.4
0.3
0.2
0.1
0.0
2000
2500
3000
3500
4000
Russell Size Ranking
Panel E: ILLIQ
100
80
60
40
20
0
2000
2500
3000
3500
4000
Russell Size Ranking
Panel F: SR
0.10
0.08
0.06
0.04
0.02
0.0
2000
2500
3000
3500
4000
Russell Size Ranking

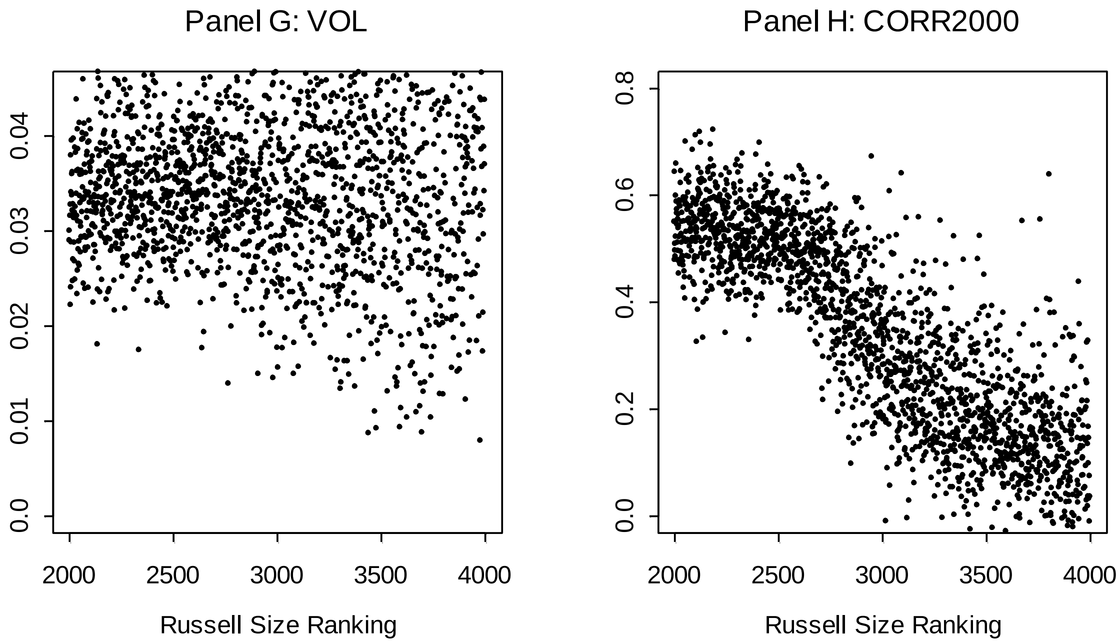
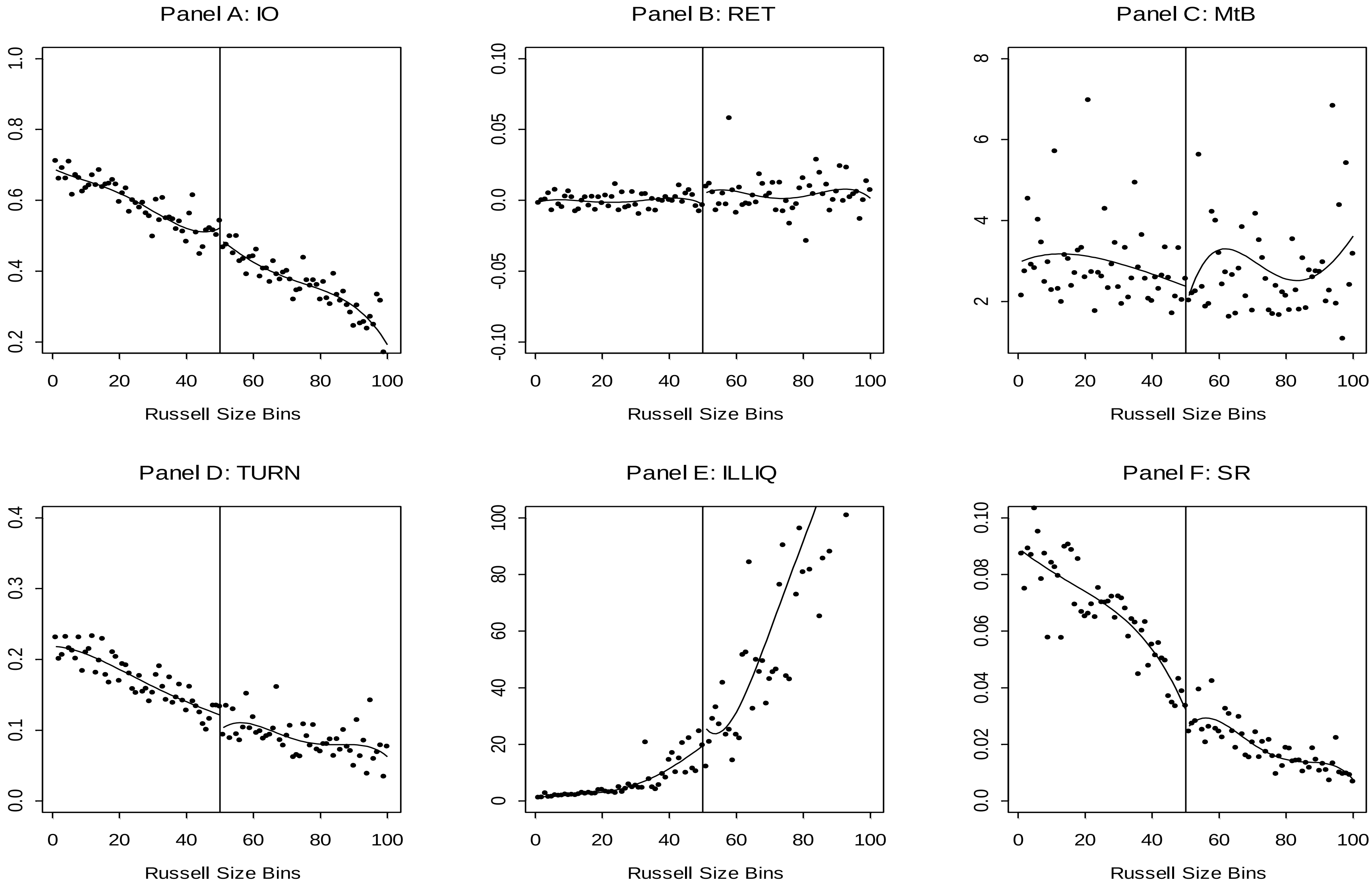


Figure 7. Discontinuity around Russell 2000 lower end cutoff. This figure plots the test variables against Russell size rankings. For each of the 1000 ranking positions above and below the Russell 2000 lower end cutoff, we compute the means of each variable of interest across the sample years and the twelve months following each annual index reconstitution. The data is arranged with larger stocks on the LHS of each plot. The sample period is from 2005 to 2008.


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

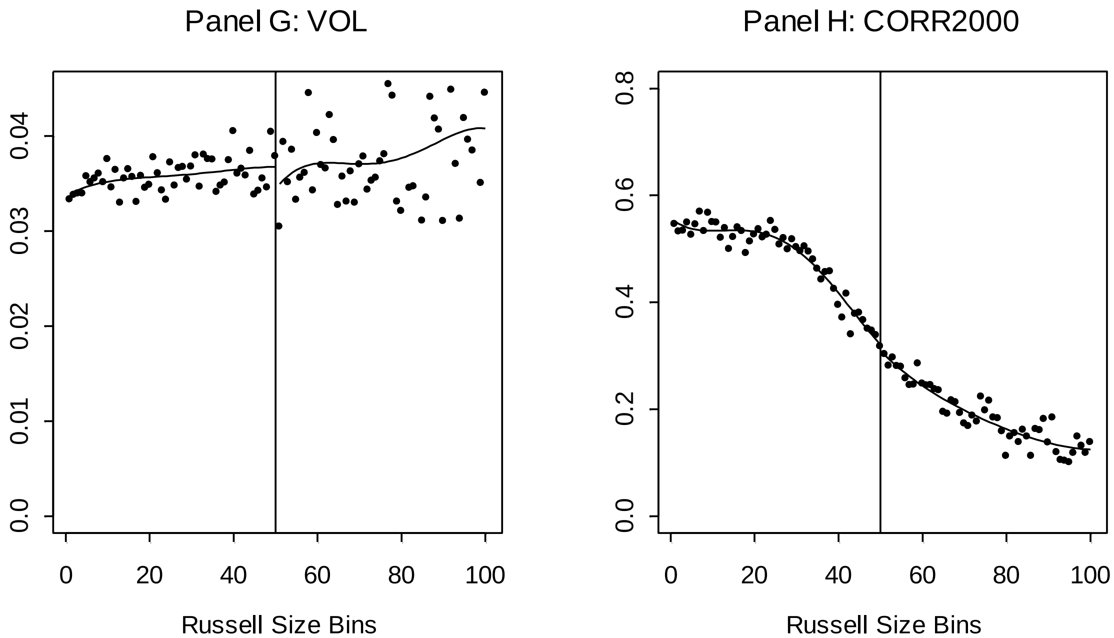


Figure 8. Discontinuity around Russell 2000 lower-end cutoff. This figure plots the test variables against Russell size rankings. For each of the 1000 ranking positions above and below the Russell 2000 lower-end cutoff, we compute the means of each variable of interest across the sample years and the twelve months following each annual index reconstitution. Results from adjacent ranking positions are then grouped into a total of 100 bins. The bin averages are then plotted with the larger firms on the LHS and smaller firms on the RHS. The vertical line at the middle of each panel indicates the index cutoff. For each panel we overlay two quartic functions from estimating the following regressions: $Y_B = \alpha_0 + \alpha_1 X_B + \alpha_2 X_B^2 + \alpha_3 X_B^3 + \alpha_4 X_B^4 + \varepsilon$, where Y_B is the bin average, and X_B is the bin number. The quartic functions are estimated using data from each side of the cutoff separately. The sample period is from 2005 to 2008.