Internet Appendix accompanying

Flying Under the Radar: The Effects of Short-Sale Disclosure Rules on Investor Behavior and Stock Prices

October 21, 2019

This appendix contains supplemental material for the paper "Flying Under the Radar: The Effects of Short-Sale Disclosure Rules on Investor Behavior and Stock Prices." We provide additional descriptive statistics, results of robustness tests, and analyses corroborating the findings in the main text.

The roadmap is as follows: In Sections IA.1 and IA.2, we show additional figures describing the stock universe underlying our analyses. Section IA.3 provides further evidence of investors' avoidance to disclose their position publicly using alternative estimation procedures. Section IA.4 reports additional tests related to the notion of reverse engineering. Section IA.5 contains additional descriptive statistics and robustness tests related to the documented return predictability using the calendar-time portfolio approach.

Contents

IA.1 Additional sample characteristics

As described in the main text (see page 12), the vast majority (73.5%) of stocks with short positions are in the highest market capitalization quartile, and almost all stocks with short position notifications (95.8%) appear above the median value of the market capitalization distribution. Figure IA.1(a) provides further details by plotting the frequency of stocks with an open short position across the four quartiles of market capitalization. The quartile breakpoints are defined according to the overall sample of stocks in the German regulated stock market. As is evident from the figure, the distribution is highly right-skewed irrespective of whether stocks have confidential or public short position notifications. Similar findings emerge when plotting the distribution across the four quartiles of institutional ownership, as in Figure IA.1(b). Also, stocks with short position notifications are highly liquid as evident from Figure IA.1(c) (Amihud illiquidity measure) and Figure IA.1(d) (bid-ask spread). Finally, stocks with short position notifications are held by a considerable share of institutional investors indicating a sufficiently large supply of stocks to borrow (Nagel, 2005).

The descriptive statistics in Table 2 of the main text show that 22% of all German stocks in the regulated market have at least one short position notification over the sample period. Figure IA.2(a) illustrates this share measured at each point in time. As is evident from the blue line, the share of stocks with at least one short position notification increases from 15% at the beginning of the sample to around 25% at the end of the sample period. We also observe a slight increase of the stocks with at least one *public* short position notification from around 7% to 10% of the overall stock universe (red line). As suggested by Figure IA.2(b), the relative increase of stocks with short position notifications is driven by both an increase in the absolute number of short position notifications and a decline of the overall number of stocks in the regulated market.

IA.2 Distribution of short position notifications

As discussed in the main text, we sort the short positions into bins of 10 basis points (bps) each. The reason is that the exact (two-digit precision) value of the short position is only recorded on the day at which the position crosses the respective thresholds at 0.20, 0.30, 0.40%, ... This feature of the reporting rule mechanically generates a pattern in the data, as illustrated in Figure IA.3. In this histogram we report the frequency of the short position notifications with a bin width of 0.01%, i.e. at two-digit precision. The histogram shows frequency spikes occurring at regular distance always at low and high second decimal places.

For a better understanding as to why this pattern emerges, we report the frequency of decimal places occurring in the short position notifications in Figure IA.4(a). It shows that the u-shaped pattern within the reporting bins originates from two processes overlaying each other: Positions that come from below are more likely to enter the reporting bin at low second decimal places. Positions that come from above are more likely to enter the bin at high second decimal places. The slightly asymmetric u-shape is due to a lower number of decreases in the sample. Figure IA.4(b) displays the two *conditional* distributions of second decimal places, showing a very symmetrical picture. Conditional on the positions coming from below (increase), the likelihood of reporting a zero as a second decimal is 35.4%, whereas the corresponding likelihood of a nine as a second decimal place is only 2.3%. The mirror image emerges when conditioning on the positions coming from above (decrease): The likelihood of reporting a short position with a zero as a second decimal place is 2.7%, in contrast to a probability of 37.9% for a nine as a second decimal place.

The pattern around high and low second decimal places of short sale notifications, as shown in Figures IA.3, IA.4(a), and IA.4(b), is the main reason why we collect positions into reporting bins. Because of these "spikes", it is unfeasible to compare density forecasts of the true underlying distribution around the disclosure threshold. In the way the reporting is set up, the share of positions with a reported value of 0.49 does not represent the true fraction of positions with that value. In other words, investors' short position is essentially measured at one-digit precision.

IA.3 Alternative specifications for avoidance tests

IA.3.1 Bunching estimation

Table IA.1 shows bunching estimates for several orders of the polynomial regression model described in Section 4. For comparison, we also present the bunching estimates for the full sample and the sample of positions below their record high. The estimates reported in Section 4 correspond to positions at their record high and the polynomial regression model of order $q = 6$. The estimated excess mass of 0.928 or 92.8% relative to the counterfactual and the corresponding standard error vary only slightly for orders of $q = 5$ or higher. Consistent with the avoidance hypothesis, for positions below the record high, we do not find any evidence that point toward a positive excess mass in the 0.4 bin. If anything, for this subsample we observe even slightly negative though economically small excess mass just below the publication threshold.

IA.3.2 Duration in reporting bins

Table IA.2 re-conducts the duration analysis of Table 4 by using the the median duration spent in each reporting bin. Table IA.3 re-conducts the analysis of Table 4 without winsorizing the upper tail at 99%, showing that the duration results are not sensitive to this outlier correction. In line with our main findings, the two table shows that the abnormal long duration in the 0.4 bin is unique to the sample of "at-the-record-high" positions while avoidance is not evident for the positions "below-the-record-high".

IA.3.3 Strategic behavior when crossing the disclosure threshold

Table IA.4 re-conducts the duration analysis of Table 6 by using the the median duration spent in each reporting bin. In line with our main findings, the table shows that there is no abnormally long duration in the 0.4 bin for positions that eventually cross the disclosure threshold. This finding applies to positions both at the record high and below the record high.

Table 7 in the main text reports the average magnitude of the position increase across different reporting bins. Now, we look at various percentiles of position increases for different reporting bins in Table IA.5. As apparent from Table IA.5, the median increase is 0.1 and identical across all starting bins including the 0.4 bin. Also, when looking at higher percentiles, we find no abnormal patterns for the bin just below the disclosure threshold even in the tails of the distribution. No other insights arise from the sample split of positions at or below their record high. Overall, the results do not support the conjecture that investors immediately jump to very high positions once they decided to cross the disclosure threshold.

IA.3.4 Alternative sample split

To uncover avoidance of investors to cross the disclosure threshold, we rely in our main analyses on a sample split into positions at and positions below their record high. As discussed in Section 4, the avoidance to pass the disclosure threshold should be particularly pronounced for positions at their record high relative to positions below their record high. We repeat our main empirical tests from Figure 1(b), and those in Table 4 and Table 5 of the main text with a slightly different definition for the sample split. Specifically, we divide all observations into position-days for which the last change was an increase, and positions previously decreased.

The two measures relate to two extremes in terms of considered data history: The measure in the main text is determined by the running maximum of the entire history of the short position. The alternative measure is determined by the shortest possible window of past position values, namely the maximum of the current and previous position value. Related to the 0.4 bin, "previously increased" positions include position days, which were never public in the past but also positions which were public prior to the last increase. Because of the latter positions, we expect that the alternative sample split results in a weaker avoidance effect relative to the "at-the-record-high" subsample from the main text.

Figure IA.5 plots the frequency of days with open short positions for the two subsamples. Similar to the findings in the main text in Figure 1(b), our alternative sample split also provides descriptive evidence for investors' avoidance to cross the disclosure threshold. Only for the "previously-increased" subsample we observe an accumulation of observations in the 0.4 bin.

Table IA.6 re-conducts the duration analysis of Table 4, but uses the alternative sample split. In line with our main findings, the table shows that the abnormal long duration in the 0.4 bin is unique to the sample of previously increased positions while avoidance is not evident for the positions previously decreased (and public). In relative terms, the mean duration in the bin just below the disclosure threshold is 27% and 55% longer than in the two neighboring intervals. The corresponding reduction in Table 4 was 22% to 55%. Results for median durations in Panel B are comparable.

Table IA.7 re-conducts the analysis of Table 5 by using the alternative sample split. In line with our main findings, Table IA.7 shows that the low probability of an increase in the 0.4 bin is entirely driven by positions approaching the threshold from below. In relative terms, for the subsample of previously increased positions, in the bin just below the disclosure threshold it is 14% and 26% less likely to increase a short position than in the two neighboring intervals. Comparing these results with findings of Table 5 with a reduction between 20% and 34%, we observe that the avoidance for previously increased positions is only slightly lower than for positions at the record high, corroborating the robustness of our results.

Overall, we confirm our main findings from Section 4.2 and Sections 4.3 by using only the last position change instead of the entire history of the short position evolution as the sample split variable.

IA.3.5 Accounting for (un)observed stock and investor characteristics

Our findings that there is both an abnormally low probability of increase and an abnormally long duration in the bin just below the publication threshold indicate that a large share of investors avoid increasing their positions because of the publication threshold. However, this lower probability of increase might stem from stock or investor characteristics that correlate with a position being just below this threshold. To rule out alternative mechanisms, we conduct panel regressions with a battery of control variables associated with short-selling constraints as well as progressively saturated fixed-effects specifications in the spirit of Jiménez, Ongena, Peydró, and Saurina (2014).

We start with a standard binary outcome model and construct a dependent variable equal to 1 on day t if a short position moves up at least one bin on the next trading day, and equal to zero otherwise: $y_{i,j,t+1} = \mathbb{1}\left(bin_{i,j,t+1} > bin_{i,j,t}\right)$. To identify whether short sellers avoid crossing the disclosure threshold, we specify the following regression models:

$$y_{i,j,t+1} = \alpha_0 + \beta_1 \; \textit{Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + \boldsymbol{\gamma'} \, \boldsymbol{X_{i,t}} + \alpha_t + \varepsilon_{i,j,t+1}, \quad \text{(IA.1a)}$$

$$y_{i,j,t+1} = \alpha_0 + \beta_1 \; \textit{Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + \alpha_{i,t} + \alpha_{j,t} + \varepsilon_{i,j,t+1}, \quad \text{(IA.1b)}$$

where *Just below threshold* is a dummy variable, indicating the reporting bin just below the publication threshold ($bin = 0.4$). The indicator $k \; bin_{i,j,t}$ is equal to 1 if investor j has a short position in stock i in bin k at time t. In our specifications the omitted benchmark bin is 0.3, i.e. $k \in \{0.2, 0.5, 0.6^{+}\}$, where 0.6^{+} indicates position bins greater than or equal to 0.6. Importantly, we not only draw our inference from the coefficient of the 0.4 bin (relative to the omitted 0.3 bin) but also estimate the effect relative to *all* reporting bins around the 0.4 bin.[1] $\boldsymbol{X_{i,t}}$ represents a vector of control variables related to short selling activity and α_t, $\alpha_{i,t}$, $\alpha_{j,t}$ denote day, stock-day, and investor-day fixed effects, respectively. If investors are on average avoiding crossing the publication threshold we would expect the coefficient of the *Just below threshold* dummy β_1 to be significantly negative relative to the omitted neighboring 0.3 bin and significantly lower than the regression coefficient of the 0.5 bin dummy.

Our first benchmark specification includes daily time-fixed effects in addition to the reporting bin indicators. Column (1) of Panel A in Table IA.8 reports the estimates of this specification. The results are in line with our previous findings. The probability of an increase conditional on being in the bin just below the disclosure threshold is significantly lower than in the 0.3 bin. In economic terms, when a short seller is in the 0.4 bin, the position is 0.54 percentage points less likely to be increased than if the short seller were in the 0.3 bin. This effect is economically significant, given that the baseline unconditional probability of an increase on the next day is slightly below 2.0%. In Panel B we observe that the probability of increase in the bin just below the threshold is also lower than in the neighboring 0.5 public bin.

The second specification includes various control variables that might be endogenously related to different dimensions of short-sale constraints. Most importantly, the coefficients of the bins around the disclosure threshold remain virtually unchanged (see Column (2) of Panel A and B). Although our main interest is still the coefficient of the 0.4 bin, this subsection also sheds some light on other drivers of short-selling activity. For example, the bid–ask spread is negatively associated with the probability of position increase, suggesting that investors are more reluctant to build up a short position in illiquid stocks. We find similar but statistically weaker evidence for the Amihud (2002) price impact measure. Consistent with Cohen, Diether, and Malloy (2007) and Kaplan, Moskowitz, and Sensoy (2013), the positive and statistically significant coefficient on the percentage of lendable

[1] Our results are robust to using a more detailed specification that includes all reporting bins up to 1.0 into the regression separately. These results are discussed on page 10.

stocks suggests that there are more short position increases if there is larger supply of stocks to borrow. In line with Prado, Saffi, and Sturgess (2016), we find a negative relation between inventory concentration and the probability of a short position increase. Although in our sample the fee itself is not associated with the probability of a short position increase, we observe that the second moment of the securities lending fee is negatively associated with short position increases. This finding is consistent with the notion that investors are concerned with high variation in fees when establishing their position (D'Avolio, 2002; Engelberg, Reed, and Ringgenberg, 2018). Furthermore, consistent with the idea that it is difficult to establish the same relative position in larger stocks, market capitalization relates negatively to the probability of an increase. As described in Section 2.1 of the main text, the reported net short positions also include equivalent derivative positions, which are accounted for on a delta-adjusted basis. Thus, if futures or listed options exist for an underlying stock, it is presumably easier to establish and increase a net short position. In keeping with this notion, the dummy indicating the existence of futures or listed options is associated with a positive coefficient. Interestingly, when we control for the above-mentioned short-selling frictions, we observe that investors tend to increase their short position more aggressively for stocks with higher return volatility. Such an effect is consistent with the notion that low-volatility stocks are insufficient to provide short sellers with the expected level of profit, one which exceeds their transaction costs. As a consequence, they demonstrate a preference for high-volatility stocks (Angel, Christophe, and Ferri, 2003).

In Column (3), we exploit the three-dimensional structure of our data set by including stock-day and investor-day fixed effects to suppress the influence of stock and investor unobservables on the decision to increase a short position. The fixed effects control for characteristics, such as the overall borrowing demand for the stock, the supply of stocks to borrow, borrowing fees, or liquidity, but they also control for any time-varying investor characteristic such as the investor's size or leverage, or any sort of funding constraint that may vary at investor level. The identification of investors' avoidance to cross the publication threshold stems from comparing the probability of an position increase on the *same* day by different short sellers for the *same* stock, controlling for any time-varying investor characteristic. The estimated regression results in Column (3) indicate that neither unobserved stock-specific nor investor-specific characteristics explain the abnormally low probability in the 0.4 bin. That is, the probability of an increase when the investor is in the bin just below the publication threshold is on average significantly lower than in the next lower bin (Panel A) but also than in *any* other reporting bin (Panel B).[2] Taking the results in Table IA.8 together, the avoidance of a position increase in the bin just below the threshold is robust to any time-varying stock or investor characteristic.

Alternative clustering of standard errors. Table IA.9 repeats the analysis of Table IA.8 using an alternative clustering of standard errors. Instead of clustering standard errors at the investor-stock and time level we cluster standard errors at the stock and time level.

Controlling for additional reporting bins separately. The results from the linear probability model (LPM) from Table IA.8 document an abnormally low probability of increase in the bin just below the publication threshold. In the regression specification of the main text, we collect all position days with a position size of 0.6 and above in one reporting bin to ensure a sufficient number of observations for each dummy variable. In this section, we extend the number of reporting bins and include the intervals 0.7, 0.8, 0.9, and 1.0^+ in the LPM, where 1.0^+ denotes all positions shorting 1% or more of the stock's shares outstanding. This extension comes at the expense of a low number of observations

[2] The inclusion of the different fixed effects reduces the number of observations in some of the estimated models due to missing variation. However, this loss of observations is relatively small in this rich data set. With both stock×time and investor×time fixed effects, the number of observations with sufficient variation within stock and investor level, accounts for 73% of the overall sample size.

for the bins above 0.6. Despite this shortcoming, our main result of short sellers' avoidance to cross the publication threshold remains essentially the same.

In Table IA.10 we report regression results in line with the findings in the main text. In particular, the probability of a position increase when the position is in the bin just below the publication threshold is significantly lower relative to the corresponding probabilities in the two neighboring bins. This result holds even when we account for different dimensions of unobserved effects. For example, in the most saturated specification, controlling for both time-varying stock and time-varying investor fixed effects, we find that the probability of a position increase is the lowest among all reporting bins.

Fixed-effects regression with at-record high interactions. Throughout Section 4 we document that avoidance to cross the disclosure threshold is particularly strong for positions at their record high. To account for this difference in our fixed-effects regression model from Equation (IA.1b), we interact all bin dummies with a dummy variable indicating positions at their record high. The indicator variable *Record high* is defined according to Equation (1) of the main text. Formally, we estimate the following model:

$$\begin{aligned} y_{i,j,t+1} \quad = \quad & \beta_0 + \beta_1 \ \textit{Just below threshold}_{i,j,t} + \sum_k \beta_k \ k \ \textit{bin}_{i,j,t} \\ + \quad & \tilde{\beta}_1 \ \textit{Just below threshold}_{i,j,t} \times \textit{Record high}_{i,j,t} \\ + \quad & \sum_k \tilde{\beta}_k \ k \ \textit{bin}_{i,j,t} \times \textit{Record high}_{i,j,t} \\ + \quad & \gamma \ \textit{Record high}_{i,j,t} + u_{i,j,t+1}. \end{aligned} \tag{IA.2}$$

where we control for all reporting bins up to 1.0^+ separately, and estimate the model using different specifications of $u_{i,j,t+1}$ to account for unobserved stock and investor effects as in Table IA.10.

Similar to Panel B of Table IA.10, in Table IA.11 we report the probability of increase for the bin just below the publication threshold *relative* to all other reporting bins. Consistent with the findings presented in the main text, we find evidence on avoidance only for positions at their record high. The probability of position increase in the 0.4 bin is significantly lower than in the neighboring bins for positions at their record high across all specifications. For positions below their record high, in contrast, this difference is insignificant or significantly positive.

Comparison: "increase", "decrease" and "no change". In our analysis we focus on the probability of a position increase. Analyzing this event is a natural choice because we are interested in investors' potential avoidance to cross the disclosure threshold. Naturally, at any given day, an investor has three options: increasing, decreasing, and not changing its short position. Essentially, our analysis in the main text compares the event "increase" against the events "decrease" and "no change". In the following we shed more light on the other two events. To do so, we follow our modeling approach of Equation (IA.1b) from the main text and define our three dependent variables as follows:

$$y_{i,j,t}^{increase} = \mathbb{1}\left(bin_{i,j,t} > bin_{i,j,t-1}\right), \tag{IA.3}$$

$$y_{i,j,t}^{no\ change} = \mathbb{1}\left(bin_{i,j,t} = bin_{i,j,t-1}\right), \text{and} \tag{IA.4}$$

$$y_{i,j,t}^{decrease} = \mathbb{1}\left(bin_{i,j,t} < bin_{i,j,t-1}\right). \tag{IA.5}$$

where $bin_{i,j,t}$ denotes the reporting bins (0.2, 0.3, 0.4, 0.5, ...). In case of avoidance, we expect that the probability of not changing the short position in the bin 0.4 is higher than in the two neighboring bins 0.3 and 0.5. Moreover, given the longer duration in bin 0.4 (see Table 4 in the main text), we expect that the probability of decrease in the bin just below the publication threshold is lower than in the two neighboring bins.

The results of this extended analysis, using the most saturated fixed-effects model, are shown in Table IA.12. As in Panel B of Table IA.10, and Table IA.11, we report the estimates for the bin just below the publication threshold *relative* to all other reporting bins. Column (1) shows the relative probability of increase, which corresponds to our previous analysis shown in Table IA.10, Panel B, Column 5. Column (2) of Table IA.12 displays the results for the relative probability of not changing the bin on the next day. Coefficients of all bins are significantly positive. This result implies that the probability of staying in the same bin is significantly higher just below the threshold compared to *all* other reporting bins. The difference in the probability is economically sizable when compared to the neighboring bins, amounting to 1.51 to 2.42 percentage points. Since investors are avoidant to cross the disclosure threshold, they stay longer in the 0.4 bin. This result reflects our finding from Section 4.2 of the main text: The duration in the 0.4 bin is the longest among all reporting bins.

In Column (3) we observe that the probability of a position decrease is lower in the 0.4 bin compared to its neighboring bins. This finding can be explained by the longer duration in the 0.4 bin. In particular, days with a position decrease decline in relative terms because investors increase the days they stay in the 0.4 bin.

IA.4 Additional analyses related to the notion of reverse engineering

In this section, we provide more details on the construction of the strategy distinctiveness index (Sun, Wang, and Zheng, 2012) and the flow-performance sensitivity of hedge funds. We also examine alterations of these two measures in the comparison analysis of secretive and non-secretive positions of Table 9. In particular, we change the number of peer groups to calculate the distinctiveness of the fund's strategy and we shorten the estimation window to compute the flow-performance sensitivity.

IA.4.1 Hedge funds style categories and strategy distinctiveness

Sun et al. (2012) propose a measure of the distinctiveness of a fund's investment strategy that is based on historical fund return data. The measure is constructed using a model-free approach that does not require information about funds' individual positions but exploits the co-movement between funds' past returns. The SDI quantifies the degree to which a hedge fund follows an investment strategy that is distinct from its peers.

Estimation of style categories. To gauge the distinctiveness of a fund's investment strategy, Sun et al. (2012) first define ten different hedge fund styles to determine a fund's cohort. They rely on the original study by Brown and Goetzmann (2003), who propose a statistical clustering approach to group funds into style categories based on similarities in their past return variation. They show that, in terms of identifying similarities in fund strategies, these statistical clusters are significantly more helpful than the self-reported, static, and database-dependent style categories.

In our paper, we base our style formation on Ward's (1963) hierarchical clustering method, which seeks to create groups of funds that minimize the across-fund variation within the styles. The main advantage of hierarchical clustering is the replicability of the clusters. In other words, the algorithm returns exactly the same clusters irrespective for the starting point of the clustering procedure.[3] For each year, we take the monthly returns for the previous three years for all funds in the Lipper TASS universe to construct the clusters. As an example, Figure IA.7 illustrates how funds are assigned to styles as the number of clusters increases for the years 2014 and 2015. The y-axis shows the average return in the previous three years for each style category, and the thickness of the line indicates the size of the cluster. Note that the clustering procedure relies not only on the average returns of the funds but also minimizes the distance between funds within a cluster using *every* single monthly return.

[3]This characteristic contrasts with partitional clustering, in which the final cluster composition depends on the initial choice of clusters.

Similar to the original paper of Brown and Goetzmann (2003), we also test how well the style categories capture the similarity in fund strategies in our sample from 2012 to 2015. To do so, we regress the monthly return of a fund, $r_{j,t}$, on the contemporaneous value-weighted return of the corresponding style category, $\mu_{J,t}$, and include fund and time fixed effects:

$$r_{j,t} = a + \beta_1 \mu_{J,t} + \gamma_j + \gamma_t + \varepsilon_{j,t} \tag{IA.6}$$

Style categories are defined either as in the TASS database or as eight and ten styles based on clustering the historic returns. The idea of this exercise is to assess how much of a fund's return variation (strategy) is explained by the returns of the fund's peers. A higher (within) R^2 suggests a better grouping of funds in categories. We report the results of the regression in Table IA.13.

The within R^2 is the statistic of main interest. We observe in Column (1) that the explanatory power of the "TASS-style" return for the individual fund return is very low, with a value of 3.48%. In Columns (2) and (3), these values are almost four times higher and they increases significantly to 11.39% and 12.37% when we form eight and ten styles according to Ward's (1963) hierarchical clustering procedure. These results are broadly in line with the original paper of Brown and Goetzmann (2003) and suggest that statistical style formation dominates the static, self-reported style definition of the Lipper TASS database. The choice of eight and ten clusters is based on previous literature (Sun et al., 2012; Brown and Goetzmann, 2003). Note that the results shown in the last two columns are not mechanically driven by the clustering procedure. The statistical clustering of funds into style portfolios is based on *past* return data. Thus, there is no data overlap between the period when we construct the clusters and the period when we estimate the return co-movement of funds and their peers.

Estimation of the SDI measure. Following Sun et al. (2012), we compute a hedge fund's SDI as 1 minus the correlation between the fund's return and the return of its peers based on monthly data for the previous three years:

$$\mathrm{SDI}_j = 1 - corr(r_j, \mu_J) \tag{IA.7}$$

where r_j is fund j's return, μ_J is the mean return of all funds belonging to the same style J, and $j \in J$. The SDI ranges between 0 and 2 and represents a "distance" measure: The higher the SDI, the farther a fund is from its peer group and the more distinctive the fund's strategy is. We use the statistical clustering approach described above to group funds into style categories. We then construct two versions of the SDI measure, depending on whether the mean style return, μ_J, is computed by using equal-weighted fund returns, μ_J^{EW}, or by weighting the returns with the value of each fund's assets under management, μ_J^{VW}. Moreover, each of these versions of the SDI is constructed assuming either ten (Sun et al., 2012) or eight (Brown and Goetzmann, 2003) style categories. Following Sun et al. (2012), we exclude fund-of-funds when calculating the SDI. Lastly, to merge the strategy distinctiveness of funds with their short position disclosures at the fund company level, we compute the SDI of a hedge fund company as the AUM-weighted SDI across all the funds within a fund company and match institutions by name.

Strategy distinctiveness and disclosure avoidance. Regarding the disclosure threshold we put forward the following hypothesis. If short sellers avoid disclosing their short positions because they fear that competitors may reverse engineer their profitable strategies, we expect them to be more likely to follow distinctive investment strategies. Our results in Table IA.14 support this hypothesis. When fund returns are value-weighted within style categories, the SDI measure for funds with a secretive short position is 0.74 compared 0.56 for non-secretive short positions (Panel A). The difference of 0.17 is statistically and economically significant. In our sample, the mean of the SDI is 0.58 and its standard deviation is 0.26. Our finding is robust to different alterations in measuring the SDI, i.e., equal-weighting the returns within style clusters or using eight instead of ten clusters as style categories. The results of Panel A, Table IA.14 are reported in the main text in Panel B of Table 9.

IA.4.2 Flow-performance sensitivity and disclosure avoidance

To estimate the flow-performance sensitivity of each fund, at every year-end, we first run the following regression specification with three or two years' worth of monthly return and flow data:

$$\text{Flow}_{j,t} = a + \beta_1 \text{Performance Decile}_{j,t-1,t-12} + \beta_2 \text{Performance}_{j,t-1,t-12} + \varepsilon_{j,t}, \qquad \text{(IA.8)}$$

where

$$\text{Flow}_{j,t} = \frac{\text{AUM}_{j,t}}{\text{AUM}_{j,t-1}} - (1 + r_{j,t}).$$

$\text{Flow}_{j,t}$ refers to the cashflow of fund j in month t in terms of the percentage of the assets under management. $\text{Performance Decile}_{j,t-1,t-12}$ refers to fund's cross-sectional decile ranking, based on either raw returns or CAPM alphas for the past twelve months. The variable takes the value of 10 for the top 10% of the cross-sectional performance distribution and 1 for the bottom 10%. $\text{Performance}_{j,t-1,t-12}$ refers to the corresponding raw values. Thus, our flow-performance sensitivity measure comes in four different versions, depending on whether we define performance based on raw returns or on CAPM alphas, and on whether we estimate the sensitivity with three or two years of past data. The market beta of each fund is calculated using three years of past return data and updated at the end of every year. In addition, all variables are de-meaned within style-month to account for time-varying differences in hedge fund styles.[4] To merge the flow-performance sensitivity of funds with their short position disclosures at the fund company level, we

[4] This step corresponds to a fixed-effect estimation.

compute the flow-performance sensitivity of a hedge fund company as the AUM-weighted sensitivity across all the funds within a fund company and match institutions by name.

Hedge funds with a high flow-performance sensitivity lose more in terms of inflows if their profitable strategies are reverse-engineered by competitors and profits are competed away. Thus, these funds have a stronger incentive to hide their strategy from competitors.

In Table IA.15, we test whether investors of secretive funds are particularly sensitive to relative performance. In Panel A (Panel B) we use three (two) years' worth of monthly return and flow data. Irrespective of whether we use the relative performance measure based on raw returns or the one that is based on CAPM alphas, we find strong evidence that the flows of secretive hedge funds tend to be more sensitive to past performance compared to their peers in the control group. These results are discussed in Section 6.2 on page 28. The results of Panel A of Table IA.15 are reported in the updated Panel B of Table 9 in the main text.

IA.5 Additional analyses to the calender-time approach

In Section 7.1 of the main text we find that a portfolio consisting of stocks with secretive positions just below the publication threshold yields on average a significant abnormal return of -5.46 to -6.13 bps per day. In this section of the Appendix we briefly illustrate the composition of the portfolio over time.

As indicated on page 35 in the main text, the average number of stocks in this portfolio is 32 ranging from 21 up to 49 stocks on a given day. In Figure IA.6 we plot the number of distinct stocks in the portfolio over time. In addition, the figure includes the time series of positions in this portfolio. The number of positions is slightly higher given that there can be more than one secretive position that is just below the publication threshold for the *same* stock by different investors. Two notable findings emerge from the figure. First, the number of stocks and positions in the portfolio only slightly varies, with an increasing trend during our sample period. Second, the majority of positions come from distinct stocks, suggesting that investors' publication avoidance and its consequences for price efficiency documented in this study are prevalent effects in the financial market.

In this section we conduct a number of variations to our sample and variable definitions to test the robustness of the results from Section 7.1. Table IA.16 reports the average abnormal returns of the benchmark model from Panel A of Table 10 and those of our sensitivity analyses.

In the main analysis we include a stock into the portfolio if we observe a secretive position that is in the bin just below the publication threshold. Consequently, the returns within the portfolio are weighted by the number of *positions*. The rationale is that we expect overpricing to be positively related to investors' avoidance and to the number of positions that are possibly constrained. As a robustness test, we now use equal weighting. That is, we include a stock in the portfolio if we observe one or more secretive positions that are in the bin just below the publication threshold. As is evident from Panel B of Table IA.16, employing the alternative, less powerful, portfolio weighting yields only slightly lower returns in absolute terms relative to the benchmark result in Panel A.

In our second robustness test, we exclude all short positions for which the stock price on the day before crossing the reporting threshold of 0.2% is below 1 EUR. Such a restriction ensures that our main overpricing results are not driven by small-cap and illiquid stocks. As already noted in Section 3, the sample of stocks with open short positions consists almost entirely of stocks above the median market capitalization of the German stock universe. As a consequence, the exclusion of penny stocks only marginally affects our sample size. Using the 1 EUR filter, we drop merely four stocks from our sample. Accordingly, the alphas in Panel C are virtually the same as in the benchmark model in Panel A.

In Panel D we conduct another change in our sample by defining investors to be avoidant to cross the publication threshold if they spend at least five days in their maximum bin of 0.4. In comparison, in the main analysis we require only a minimum of two days spent in the maximum bin of 0.4. With the more conservative definition of being avoidant using restrictions on the shorting period, we expect to better identify avoidance, short-selling impediments, and, as a result, more accurate estimates. In line with this expectation, we indeed find that the standard errors of the abnormal returns decrease and the statistical significance of the effect increases across all asset pricing models compared to the baseline model in Panel A, despite a slight decrease in the economic magnitude.

Our main results on overpricing are also robust to changes in the asset pricing factors. Namely, in Panel E of Table IA.16, we use European factor portfolios, which are also obtained from Andrea Frazzini's data library, provided through AQR's website.[5] Using European factors results in economically comparable, but noisier alpha estimates. This result is in line with Griffin (2002) who documents that local factor portfolios are generally better suited for performance evaluation.

[5] See: https://www.aqr.com/Insights/Datasets/Quality-Minus-Junk-Factors-Daily

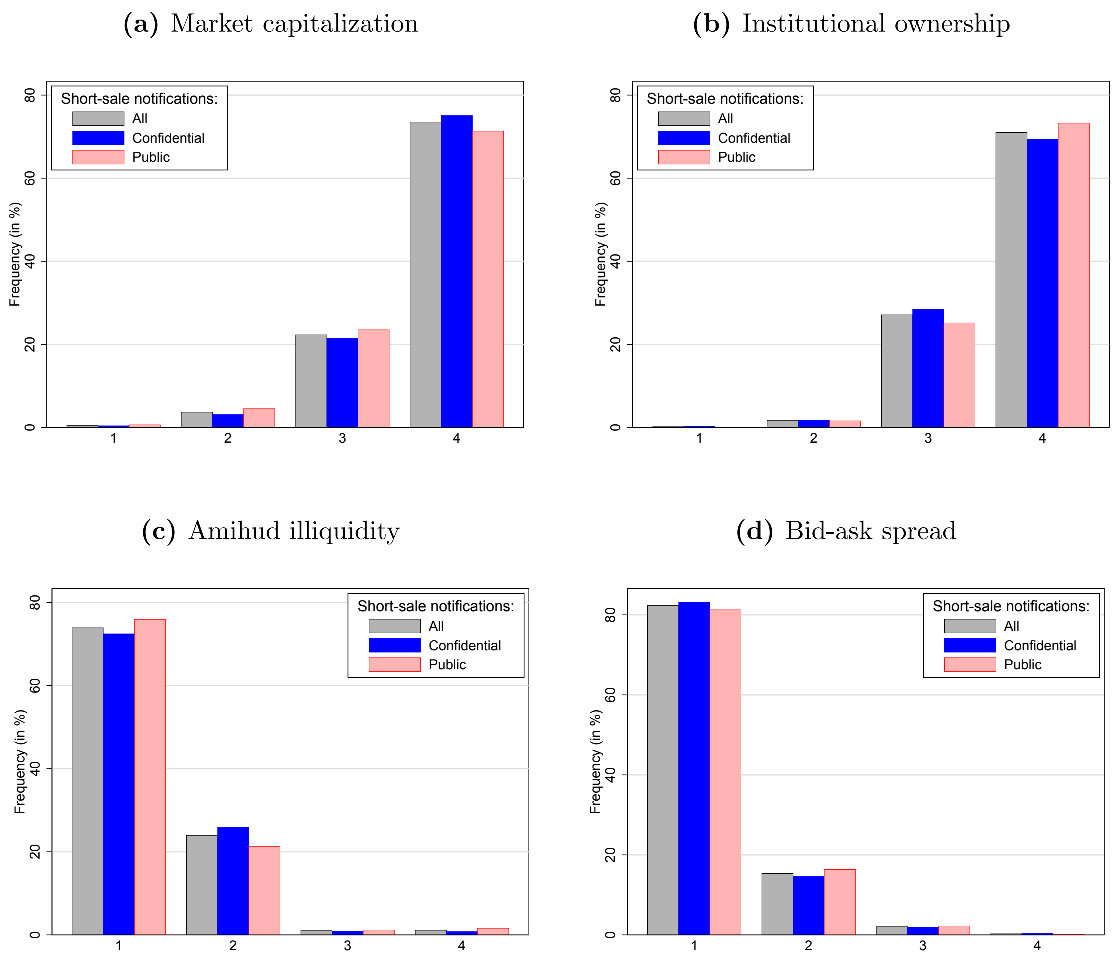
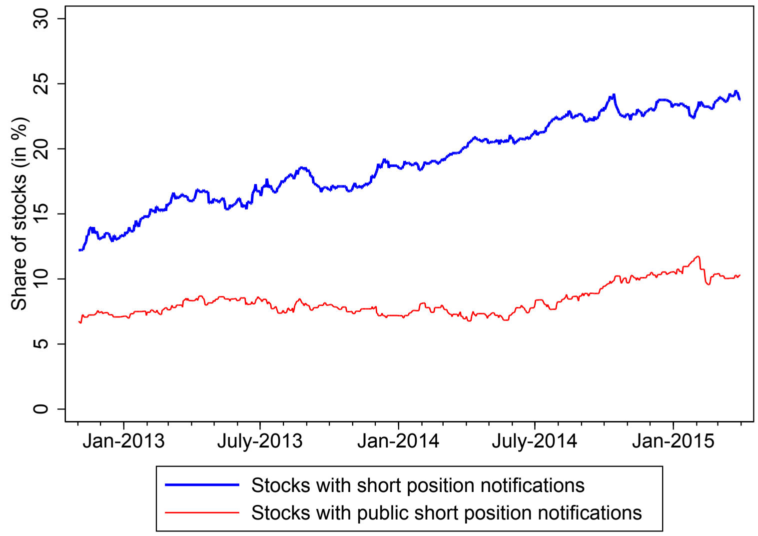


Figure IA.1:
Distribution of stocks with open short-position notifications across size, institutional ownership, and illiquidity
This figure displays the distribution of stock-days with open short-sale notifications across quartiles of market capitalization (Figure IA.1(a)), institutional ownership (Figure IA.1(b)), the Amihud illiquidity measure (Figure IA.1(c)), and the bid-ask spread (Figure IA.1(d)). The quartile breakpoints are defined according to the overall sample of stocks in the German regulated stock market. All short-sale notifications subsume stocks with either public or confidential short-sale notifications, confidential short-sale notifications consist of stocks with positions above 0.2% but below 0.5% of issued share capital, and public short-sale notifications consist of stocks with at least one short position of 0.5% or higher. The overall sample contains all German domestic equity in the regulated market from November 5, 2012, to March 31, 2015.

(a) Percentages



(b) Number of stocks

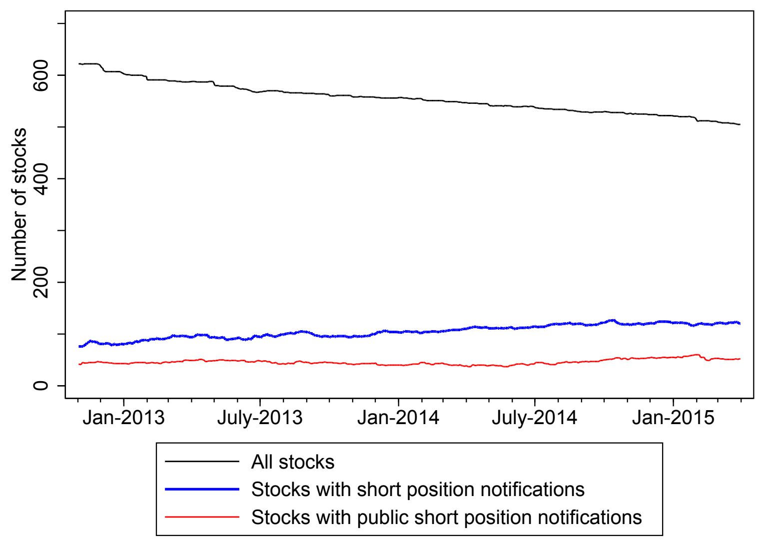


Figure IA.2:
Stocks with short position notifications over time

Figure IA.2(a) displays the percentage of stocks that have at least one short position notification below the publication threshold ($\geq 0.2\%$) and the percentage of stocks that have at least one short position notification above the publication threshold ($\geq 0.5\%$) over time, respectively. Figure IA.2(b) shows the same statistics over time in absolute terms as well as the overall number of stocks in our sample. The overall sample contains all German domestic equity in the regulated market from November 5, 2012, to March 31, 2015.

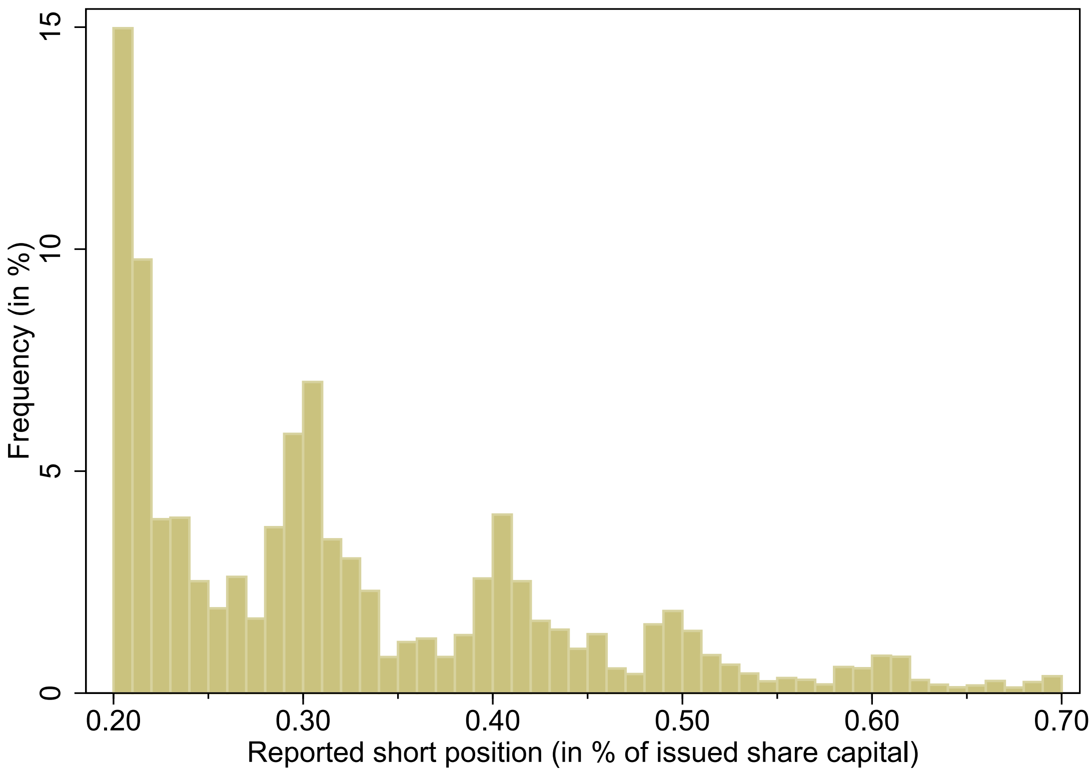
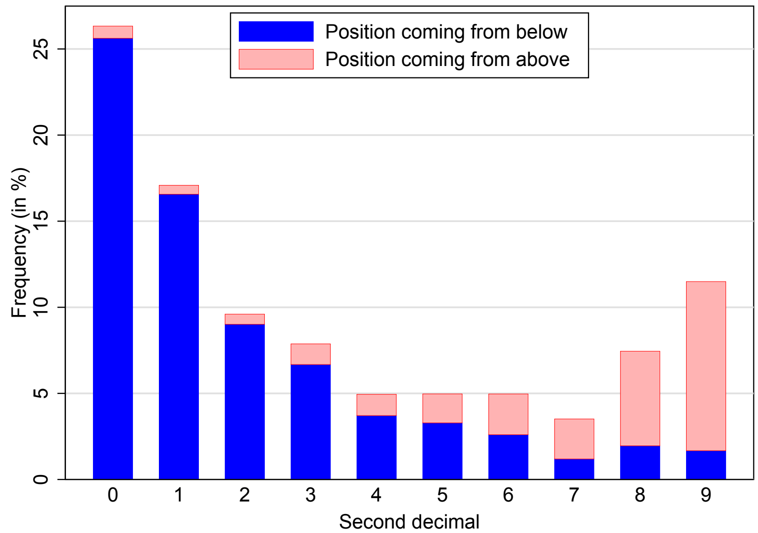


Figure IA.3:
Detailed histogram of reported short positions

This figure displays the detailed frequency histogram of reported short positions with a bin width of 0.01%. Positions greater than 0.70% are truncated for readability. The sample contains all short position notifications in the German regulated equity market from November 5, 2012, to March 31, 2015.

(a) Overall distribution



(b) Conditional distribution

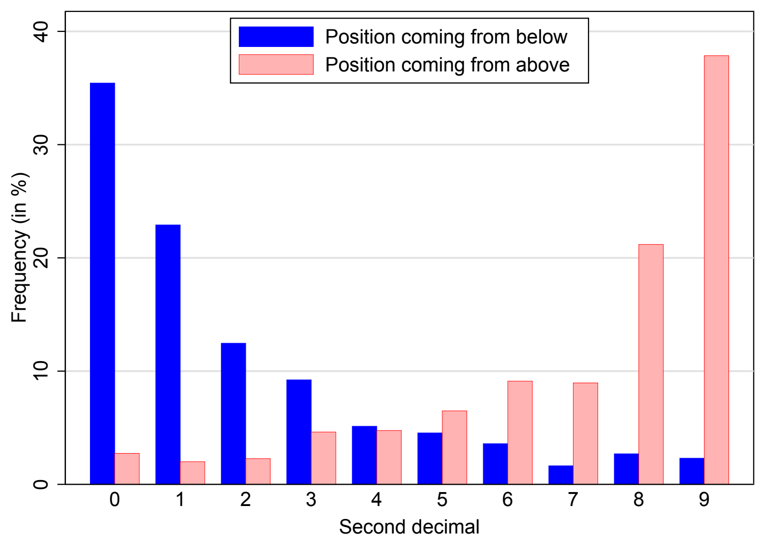


Figure IA.4:
Second decimal places of the short position notification
Figure IA.4(a) displays the relative frequency of second decimal places of short-position notifications. Figure IA.4(b) shows the frequency distribution conditional on positions coming from above or below. The sample contains all short position notifications in the German regulated equity market from November 5, 2012, to March 31, 2015.

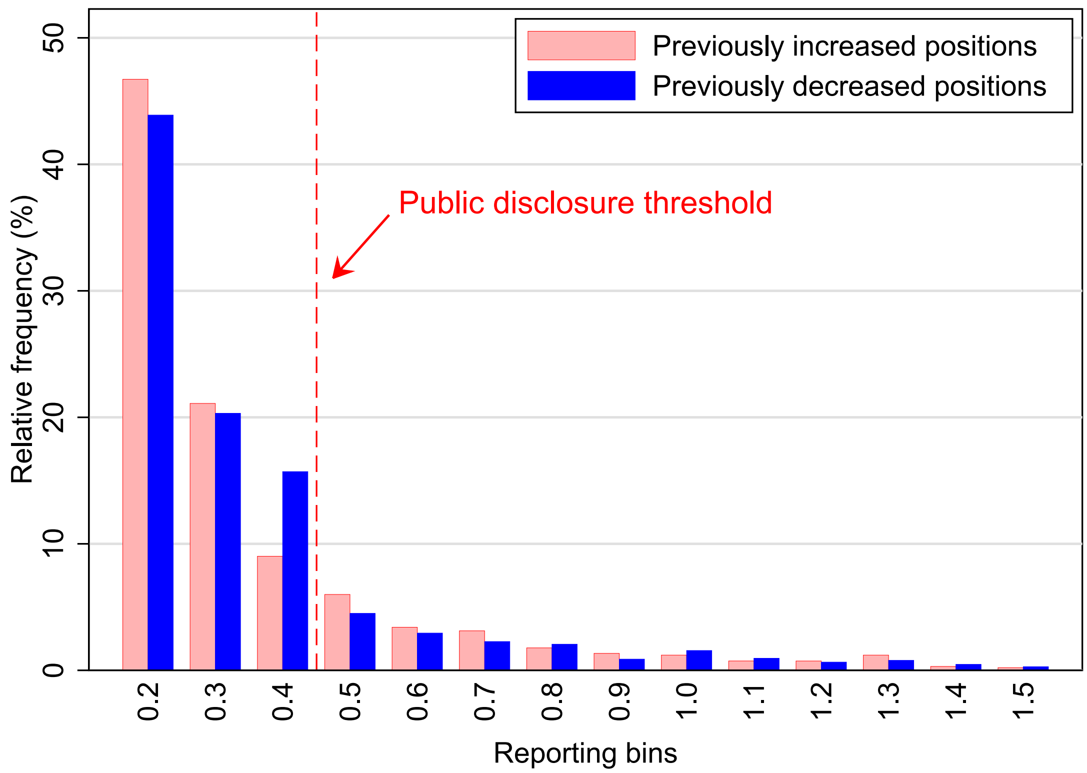


Figure IA.5:
Distribution of open short positions
This figure displays the relative frequency separately for short positions that previously increased and previously decreased, respectively. Reporting intervals greater than 1.6% are truncated for readability. The overall sample contains all German domestic equity in the regulated market from November 5, 2012, to March 31, 2015.

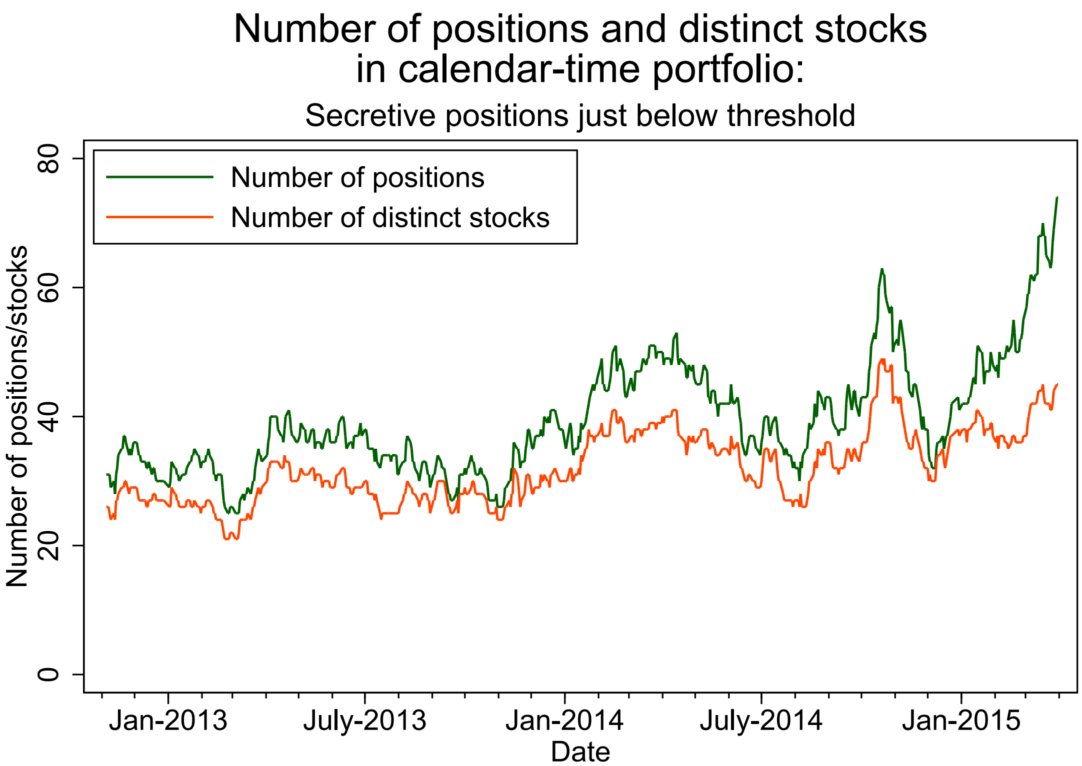
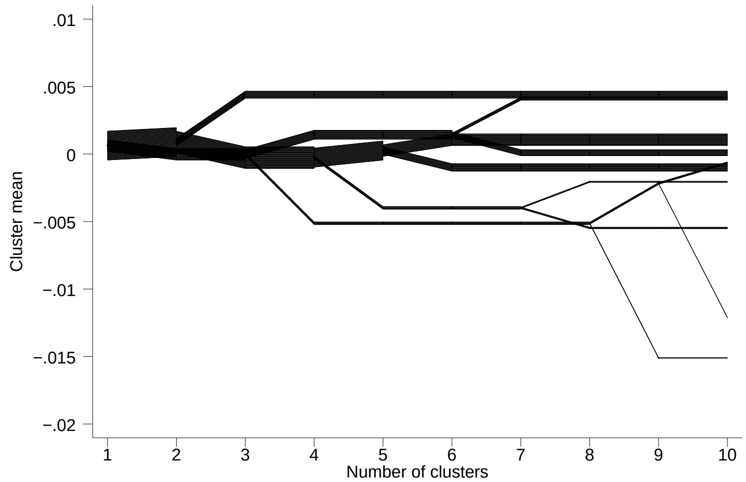


Figure IA.6:
Number of stocks with secretive positions just below the publication threshold
This figure shows the number of secretive positions that are just below the publication threshold over time (green line). The red line is the corresponding number of distinct stocks with at least one secretive position that is just below the publication threshold. The overall sample contains all German domestic equity in the regulated market from November 5, 2012, to March 31, 2015.

(a) Style categories in year 2014



(b) Style categories in year 2015

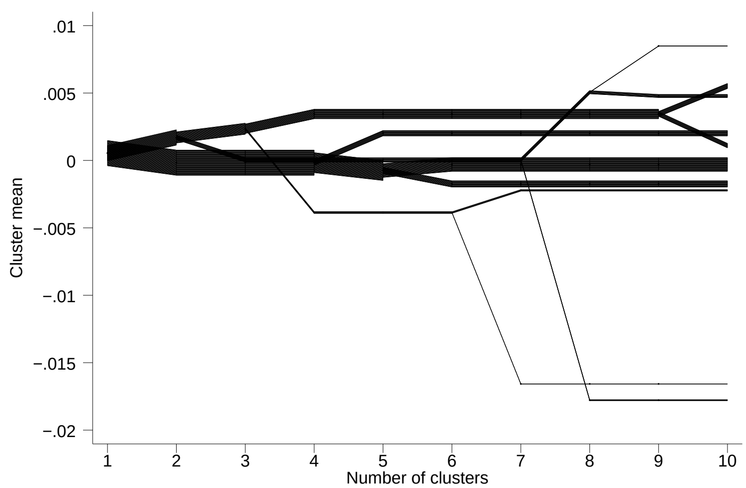


Figure IA.7:
Clustergram of hedge funds and style categories

The figure shows how Ward's (1963) hierarchical clustering procedure assigns funds to styles as the number of clusters increases. To determine the clusters, we rely on monthly returns for the previous three years. The y-axis shows the average return for the previous three years for each style category and the thickness of the line indicates the number of funds within the cluster. The two charts illustrate clusters for the years 2014 and 2015 respectively

Table IA.1:
Bunching estimates: alternative specifications of the polynomial regression model

This table shows bunching estimates complementing Figure 1(b). We fit a polynomial regression to the observed frequency distribution,

$$F_j = \sum_{i=0}^{q} \beta_i \cdot (j\ bin)^i + \gamma \cdot \mathbf{1}\left[j\ bin = 0.4 \text{ bin}\right] + \epsilon_j,$$

in which F_j denotes the observed relative frequency in bin j and $j\ bin$ denotes the corresponding short position bin. The estimate of the counterfactual distribution is the fitted value from this regression, in which the contribution of the dummy just below the disclosure threshold is omitted: $\widehat{F}_j = \sum_{i=0}^{q} \widehat{\beta}_i \cdot (j\ bin)^i$. The excess frequency just below the disclosure threshold relative to this counterfactual distribution is $\widehat{B} = F_{0.4} - \widehat{F}_{0.4}$. In relative terms, the resulting bunching estimate is given by $\widehat{b} = \widehat{B}/\widehat{F}_{0.4}$. Following Chetty, Friedman, Olsen, and Pistaferri (2011), we compute standard errors using a bootstrap procedure. We draw from the residuals $\widehat{\epsilon}_j$ in the polynomial regression with replacement to obtain a new set of frequencies and then recalculate the bunching estimate. The table reports the bunching estimate $\widehat{b}$ and its standard error for $q \in \{4, 5, 6, 7\}$. For $q = 6$, the reported estimates for positions at their record high correspond to Figure 1(b).

	Overall		Positions at their record high		Positions below their record high	
Order of polynomial (q)	Excess mass just below the threshold	Standard error	Excess mass just below the threshold	Standard error	Excess mass just below the threshold	Standard error
7	0.508	0.026	0.940	0.040	-0.180	0.030
6	0.498	0.023	0.928	0.037	-0.187	0.026
5	0.494	0.057	0.920	0.080	-0.190	0.040
4	0.267	0.139	0.620	0.170	-0.300	0.100

Table IA.2:
Duration in reporting bins: Median analysis

This table shows the median time spent in each reporting bin. Short positions are reported in bins of 10 bps, starting from 0.2% of issued share capital of the company shorted. Positions above 0.2% but below 0.5% are reported to the regulator but not disclosed to the public; positions of 0.5% and higher are disclosed to the public. Reporting bins greater than or equal to 1.0% are summarized in one group. The table reports the median number of trading days spent in each reporting bin. In addition, it displays the difference in median duration of row (3) relative to row (k), the 0.4 reporting bin just below the publication threshold (shaded in gray), and the *p*-value for differences in medians. Each panel displays probabilities for the overall sample and for two subsamples, where we split the sample into short positions at their record high and positions below their record high. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

			Overall			Positions at their record high			Positions below their record high		
		Bin	Median duration	Difference row (3) - (k)	*p*-value	Median duration	Difference row (3) - (k)	*p*-value	Median duration	Difference row (3) - (k)	*p*-value
undis-closed	(1)	0.2	6	1 ***	(0.004)	7	3 ***	(0.000)	5	0	(0.182)
	(2)	0.3	6	1 ***	(0.003)	8	2 ***	(0.007)	5	0	(0.824)
	(3)	0.4	7			10			5		
disclosed	(4)	0.5	5	2 ***	(0.000)	7	3 ***	(0.004)	4	1	(0.364)
	(5)	0.6	5	2 ***	(0.000)	6	4 ***	(0.000)	4	1	(0.331)
	(6)	0.7	4	3 ***	(0.000)	4	6 ***	(0.000)	4	1	(0.278)
	(7)	0.8	4	3 ***	(0.000)	5	5 ***	(0.000)	3	2 ***	(0.009)
	(8)	0.9	4	3 ***	(0.000)	5	5 ***	(0.000)	3	2 *	(0.053)
	(9)	≥1.0	4	3 ***	(0.000)	5	5 ***	(0.000)	3	2 ***	(0.000)

Table IA.3:
Mean duration within reporting bins: No winsorization

This table repeats the duration analysis of Table 4 without winsorizing the upper tail at 99%. Short positions are reported in bins of 10 basis points (bps), starting from 0.2% of issued share capital of the company shorted. Positions above 0.2% but below 0.5% are reported to the regulator but not disclosed to the public; positions of 0.5% and higher are disclosed to the public. Reporting bins greater than or equal to 1.0% are summarized in one group. The table reports the mean number of trading days spent in each reporting bin. In addition, it displays the difference in mean duration between the 0.4 bin (shaded in gray) relative to other bins, and the corresponding p-values. The results refer to the overall sample and two subsamples, reflecting a split of the sample into short positions at their record high versus positions below their record high. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

			Overall			Positions at their record high			Positions below their record high		
		Bin	Mean duration	Difference row (3) - (k)	p-value	Mean duration	Difference row (3) - (k)	p-value	Mean duration	Difference row (3) - (k)	p-value
undis-closed	(1)	0.2	21.7	0.5	(0.727)	27.0	1.3	(0.556)	16.2	−3.9**	(0.042)
	(2)	0.3	18.4	3.8***	(0.004)	23.1	5.2**	(0.010)	13.7	−1.4	(0.369)
	(3)	0.4	22.3			28.3			12.3		
disclosed	(4)	0.5	14.4	7.8***	(0.000)	16.9	11.4***	(0.000)	11.8	0.5	(0.766)
	(5)	0.6	13.7	8.5***	(0.000)	17.4	10.9***	(0.001)	9.9	2.4	(0.210)
	(6)	0.7	14.5	7.8***	(0.001)	17.7	10.6***	(0.008)	11.9	0.4	(0.836)
	(7)	0.8	15.7	6.6**	(0.018)	21.5	6.8	(0.136)	10.0	2.3	(0.343)
	(8)	0.9	13.3	9.0***	(0.004)	17.0	11.3**	(0.028)	9.7	2.6	(0.345)
	(9)	≥1.0	13.3	9.0***	(0.000)	17.7	10.6***	(0.000)	8.7	3.6***	(0.008)

Table IA.4:
Is there hesitance before crossing the disclosure bin? Median analysis

This table repeats the duration analysis of Table 6 with median values instead of the mean duration values. The table reports the median number of trading days spent in each disclosure bin. In addition, it displays the difference in median duration of row (3), the 0.4 reporting bin just below the publication threshold (shaded in gray), relative to row (k), and the p-value for differences in means (medians). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

			Overall			Positions at their record high			Positions below their record high		
		Bin	Median duration	Difference row (3) - (k)	p-value	Median duration	Difference row (3) - (k)	p-value	Median duration	Difference row (3) - (k)	p-value
undis-closed	(1)	0.2	6	−1	(0.467)	7	−1	(0.244)	5	−1	(0.775)
	(2)	0.3	5	0	(0.789)	6	0	(0.993)	4	0	(0.529)
	(3)	0.4	5			6			4		
disclosed	(4)	0.5	5	0	(0.861)	6	0	(0.921)	4	0	(0.738)
	(5)	0.6	4	1 **	(0.049)	3	3 ***	(0.006)	4	0	(0.997)
	(6)	0.7	3	2	(0.115)	3	3 *	(0.070)	4	0	(0.502)
	(7)	0.8	4	1 **	(0.012)	4	2 **	(0.023)	3	1 *	(0.053)
	(8)	0.9	4	1 *	(0.074)	4	2 **	(0.019)	5	−1	(0.912)
	(9)	≥1.0	3	2 ***	(0.001)	4	2 ***	(0.000)	3	1	(0.530)

Table IA.5:
Additional summary statistics for position changes given that a position increase occurs

This table shows right-hand percentiles of position changes, *given that a position increase occurred*, across different reporting bins from which the position was increased (starting bin). The table displays position changes across the different percentiles for the overall sample and for two subsamples, in which we split the sample into short positions at their record high and into positions below their record high.

			Overall				Positions at their record high				Positions below their record high			
		Starting	Percentiles				Percentiles				Percentiles			
		bin	50th	75th	90th	95th	50th	75th	90th	95th	50th	75th	90th	95th
undis-closed	0.2	(1)	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2
	0.3	(2)	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.1
	0.4	(3)	0.1	0.1	0.1	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.2
disclosed	0.5	(4)	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.4
	0.6	(5)	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.3
	0.7	(6)	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.3
	0.8	(7)	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2
	0.9	(8)	0.1	0.1	0.2	0.3	0.1	0.1	0.2	0.2	0.1	0.2	0.3	0.3
	$\geq$1.0	(9)	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.3

Table IA.6:
Duration within reporting bins: Alternative sample split

This table repeats the duration analysis of Table 4 using an alternative sample split of short positions previously increased and positions previously decreased. The table reports the mean and median number of trading days spent in each disclosure bin (Panels A and B, respectively). In addition, it displays the difference in mean (median) duration of row (3) relative to row (k), the 0.4 reporting bin just below the publication threshold (shaded in gray), and the p-value for differences in means (medians). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Mean duration within reporting bin in trading days

			Overall			Positions previously increased			Positions previously decreased		
		Bin	Mean duration	Difference row (3) - (k)	p-value	Mean duration	Difference row (3) - (k)	p-value	Mean duration	Difference row (3) - (k)	p-value
undis-closed	(1)	0.2	18.3	2.2**	(0.019)	19.4	3.8***	(0.001)	15.6	−3.4**	(0.027)
	(2)	0.3	17.0	3.6***	(0.000)	18.3	4.9***	(0.000)	13.3	−1.1	(0.429)
	(3)	0.4	20.6			23.2			12.2		
disclosed	(4)	0.5	14.3	6.2***	(0.000)	15.0	8.3***	(0.000)	12.8	−0.7	(0.690)
	(5)	0.6	13.7	6.8***	(0.000)	15.3	7.9***	(0.000)	10.7	1.4	(0.451)
	(6)	0.7	13.3	7.3***	(0.000)	14.6	8.6***	(0.000)	11.0	1.2	(0.570)
	(7)	0.8	13.6	7.0***	(0.001)	15.1	8.1***	(0.003)	10.0	2.1	(0.373)
	(8)	0.9	11.9	8.7***	(0.000)	11.8	11.5***	(0.000)	9.4	2.7	(0.296)
	(9)	≥1.0	12.6	8.0***	(0.000)	14.9	8.3***	(0.000)	8.2	3.9***	(0.002)

Table IA.6 – Continued

Panel B: Median duration within reporting bin in trading days

			Overall			Positions previously increased			Positions previously decreased		
		Bin	Median duration	Difference row (3) - (k)	*p*-value	Median duration	Difference row (3) - (k)	*p*-value	Median duration	Difference row (3) - (k)	*p*-value
undis-closed	(1)	0.2	6	1***	(0.004)	7	2***	(0.000)	5	0*	(0.096)
	(2)	0.3	6	1***	(0.003)	7	2***	(0.000)	5	0	(0.984)
	(3)	0.4	7			9			5		
disclosed	(4)	0.5	5	2***	(0.000)	6	3***	(0.000)	4	1	(0.934)
	(5)	0.6	5	2***	(0.000)	5	4***	(0.000)	4	1	(0.393)
	(6)	0.7	4	3***	(0.000)	5	4***	(0.000)	3	2	(0.116)
	(7)	0.8	4	3***	(0.000)	5	4***	(0.000)	2	3***	(0.007)
	(8)	0.9	4	3***	(0.000)	5	4***	(0.000)	3	2*	(0.054)
	(9)	$\geq$1.0	4	3***	(0.000)	4	5***	(0.000)	3	2***	(0.000)

Table IA.7:
Probability of short position increase: Alternative sample split

This table repeats the analysis of Table 5 using an alternative sample split of short positions previously increased and positions previously decreased. The table reports the probability of increasing a short position (i.e., changing to a higher reporting bin), given that an investor currently has a position in a specific bin. In addition, it displays the difference in probability of row (3) relative to row (k), the 0.4 reporting bin just below the publication threshold (shaded in gray), and the *p*-value for the differences in means. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

			Overall			Positions previously increased			Positions previously decreased		
		Bin	Probability of increase	Difference row (3) - (k)	*p*-value	Probability of increase	Difference row (3) - (k)	*p*-value	Probability of increase	Difference row (3) - (k)	*p*-value
undis-closed	(1)	0.2	0.364	−0.010	(0.469)	0.372	−0.010	(0.524)	0.345	−0.012	(0.637)
undis-closed	(2)	0.3	0.391	−0.037**	(0.014)	0.422	−0.060***	(0.001)	0.328	0.005	(0.846)
undis-closed	(3)	0.4	0.354			0.362			0.333		
disclosed	(4)	0.5	0.408	−0.054***	(0.007)	0.487	−0.125***	(0.000)	0.263	0.071**	(0.037)
disclosed	(5)	0.6	0.476	−0.122***	(0.000)	0.573	−0.211***	(0.000)	0.313	0.020	(0.606)
disclosed	(6)	0.7	0.440	−0.086***	(0.001)	0.561	−0.199***	(0.000)	0.247	0.086**	(0.036)
disclosed	(7)	0.8	0.449	−0.095***	(0.001)	0.604	−0.242***	(0.000)	0.175	0.158***	(0.001)
disclosed	(8)	0.9	0.477	−0.123***	(0.000)	0.667	−0.305***	(0.000)	0.188	0.146***	(0.005)
disclosed	(9)	≥1.0	0.474	−0.120***	(0.000)	0.649	−0.287***	(0.000)	0.214	0.120***	(0.000)

Table IA.8:
Probability of increase: Regression approach with fixed effects

This table shows estimates from the linear probability model

$$
\begin{aligned}
y_{i,j,t+1} &= \alpha_0 + \beta_1 \textit{ Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + \boldsymbol{\gamma}' \boldsymbol{X_{i,t}} + \alpha_t + \varepsilon_{i,j,t+1}, \\
y_{i,j,t+1} &= \alpha_0 + \beta_1 \textit{ Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + \alpha_{i,t} + \alpha_{j,t} + \varepsilon_{i,j,t+1}.
\end{aligned}
$$

with $y_{i,j,t+1} = \mathbb{1}\left(bin_{i,j,t} > bin_{i,j,t-1}\right)$. *Just below threshold* is equal to 1 if investor j has a short position in stock i in the bin 0.4. k $\text{bin}_{i,j,t-1}$ is equal to 1 if investor j has a short position in stock i in the bin k. The omitted benchmark bin is 0.3, i.e. $k \in \{0.2, 0.5, 0.6^+\}$, where 0.6^+ indicates position bins greater or equal to 0.6. $\boldsymbol{X_{i,t}}$ represents a vector of control variables related to short selling activity. α_t, $\alpha_{i,t}$, and $\alpha_{j,t}$ denote day, stock-day, and investor-day fixed effects, respectively. Panel A reports the regression coefficients Panel B shows the differences in probability of increase between the bin just below the disclosure threshold and the neighboring bins. The unconditional probability of increasing a short position to enter the next bin (estimated by the sample average of $y_{i,j,t+1}$) is 2.0%. The estimated coefficients are scaled to reflect changes in percentage points. Standard errors are clustered at the investor-stock and time level. The t-statistics are given in parentheses, and *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)		(2)		(3)	
Panel A: Regressions with reporting bin dummies						
0.2 bin	−0.43***	(−3.41)	−0.42***	(−3.26)	−0.41***	(−2.76)
0.3 bin	–	–	–	–	–	–
0.4 bin (Just below threshold)	−0.54***	(−3.67)	−0.59***	(−3.88)	−1.08***	(−5.40)
0.5 bin	0.64***	(2.74)	0.59**	(2.50)	−0.16	(−0.54)
≥0.6 bin	1.22***	(4.82)	1.13***	(4.64)	1.02***	(4.18)
ln(market value)			−0.16**	(−2.07)		
ln(return volatility)			1.55***	(5.83)		
ln(Amihud illiquidity)			−0.37	(−1.25)		
ln(bid-ask spread)			−0.46**	(−2.04)		
Dummy: stock with listed futures or options			0.50***	(3.47)		
Share of lendable stocks			0.04***	(3.92)		
Inventory concentration			−0.02**	(−2.53)		
Indicative fee			0.01	(0.48)		
Standard deviation of fee			−0.07***	(−2.67)		
Fixed effects:						
Time	Yes		Yes		No	
Stock × time	No		No		Yes	
Investor × time	No		No		Yes	
R^2 (%)	0.60		0.78		38.08	
Within R^2 (%)	0.20		0.37		0.14	
Number of observations	278,003		272,340		202,057	
Panel B: Differences in the probability of increase						
Just below threshold – 0.2 bin	−0.11	(−0.86)	−0.17	(−1.23)	−0.67***	(−3.35)
Just below threshold – 0.3 bin	−0.54***	(−3.67)	−0.59***	(−3.88)	−1.08***	(−5.40)
Just below threshold – 0.5 bin	−1.18***	(−5.04)	−1.18***	(−4.96)	−0.92***	(−2.89)
Just below threshold – (≥ 0.6 bin)	−1.76***	(−7.31)	−1.72***	(−7.41)	−2.10***	(−8.51)

Table IA.9:
Probability of increase: Fixed-effects regression approach with alternative clustering

This table repeats the analysis of Table IA.8 using an alternative clustering of standard errors at the stock and time level. It shows estimates from the linear probability model

$$y_{i,j,t+1} = \alpha_0 + \beta_1 \; \textit{Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + \boldsymbol{\gamma}' \, \boldsymbol{X_{i,t}} + \alpha_t + \varepsilon_{i,j,t+1},$$

$$y_{i,j,t+1} = \alpha_0 + \beta_1 \; \textit{Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + \alpha_{i,t} + \alpha_{j,t} + \varepsilon_{i,j,t+1}.$$

with $y_{i,j,t+1} = \mathbb{1}\,(bin_{i,j,t} > bin_{i,j,t-1})$. *Just below threshold* is equal to 1 if investor j has a short position in stock i in the bin 0.4. k $\text{bin}_{i,j,t-1}$ is equal to 1 if investor j has a short position in stock i in the bin k. The omitted benchmark bin is 0.3, i.e. $k \in \{0.2, 0.5, 0.6^{+}\}$, where 0.6^{+} indicates position bins greater or equal to 0.6. $\boldsymbol{X_{i,t}}$ represents a vector of control variables related to short selling activity. α_t, $\alpha_{i,t}$, and $\alpha_{j,t}$ denote day, stock-day, and investor-day fixed effects, respectively. Panel A reports the regression coefficients Panel B shows the differences in probability of increase between the bin just below the disclosure threshold and the neighboring bins. The unconditional probability of increasing a short position to enter the next bin (estimated by the sample average of $y_{i,j,t+1}$) is 2.0%. The estimated coefficients are scaled to reflect changes in percentage points. Standard errors are clustered at the stock and time level. The t-statistics are given in parentheses, and *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. For details on the calculation of the variables, see Table A.1.

	(1)		(2)		(3)	
Panel A: Regressions with reporting bin dummies						
0.2 bin	−0.43***	(−3.10)	−0.42***	(−2.84)	−0.41**	(−2.57)
0.3 bin						
0.4 bin (Just below threshold)	−0.54***	(−3.96)	−0.59***	(−4.20)	−1.08***	(−6.01)
0.5 bin	0.64***	(2.85)	0.59**	(2.57)	−0.16	(−0.48)
≥0.6 bin	1.22***	(3.97)	1.13***	(3.81)	1.02***	(3.67)
ln(market value)			−0.16	(−1.64)		
ln(return volatility)			1.55***	(4.33)		
ln(Amihud illiquidity)			−0.37	(−1.62)		
ln(bid-ask spread)			−0.46	(−1.46)		
Dummy: stock with listed futures or options			0.50**	(2.55)		
Share of lendable stocks			0.04***	(3.15)		
Inventory concentration			−0.02*	(−1.95)		
Indicative fee			0.01	(0.85)		
Standard deviation of fee			−0.07***	(−2.81)		
Fixed effects:						
Time	Yes		Yes		No	
Stock × time	No		No		Yes	
Investor × time	No		No		Yes	
R^2 (%)	0.60		0.78		38.08	
Within R^2 (%)	0.20		0.37		0.14	
Number of observations	278,003		272,340		202,057	
Panel B: Differences in the probability of increase						
Just below threshold – 0.2 bin	−0.11	(−0.88)	−0.17	(−1.25)	−0.67***	(−3.44)
Just below threshold – 0.3 bin	−0.54***	(−3.96)	−0.59***	(−4.20)	−1.08***	(−6.01)
Just below threshold – 0.5 bin	−1.18***	(−5.17)	−1.18***	(−5.39)	−0.92***	(−2.81)
Just below threshold – (≥ 0.6 bin)	−1.76***	(−5.46)	−1.72***	(−5.51)	−2.10***	(−7.24)

Table IA.10:
Probability of increase: Fixed-effects regression approach with more reporting bins

This table shows estimates from the linear probability model

$$y_{i,j,t+1} = \beta_0 + \beta_1 \textit{ Just below threshold}_{i,j,t} + \sum_k \beta_k \; k \; bin_{i,j,t} + u_{i,j,t+1},$$

with $y_{i,j,t+1} = \mathbb{1}\left(bin_{i,j,t+1} > bin_{i,j,t}\right)$, and $u_{i,j,t+1}$ models various fixed effects and an error term. $k\ bin_{i,j,t}$ is equal to 1 if investor j has a short position in stock i in the bin k at day t, with $k \in \{0.2, 0.5, 0.6, \ldots, 1.0^+\}$, where 1.0^+ indicates position bins greater or equal to 1.0. The unconditional probability of increasing a short position to enter the next bin (estimated by the sample average of $y_{i,j,t}$) is 2.0%. Panel A shows coefficient estimates using the 0.3 bin as omitted reference category. Panel B tests for differences in the probability of increase between the bin just below the threshold and all other bins. The estimated coefficients are scaled to reflect changes in percentage points. Standard errors are clustered at the investor-stock and time level. The t-statistics are given in parentheses, and *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Regression with reporting bin dummies					
0.2 bin	-0.43***	-0.16	-0.63***	-0.29**	-0.41***
	(-3.41)	(-1.30)	(-4.72)	(-2.20)	(-2.76)
0.3 bin	–	–	–	–	–
0.4 bin (Just below threshold)	-0.54***	-0.92***	-0.68***	-0.61***	-1.08***
	(-3.66)	(-5.72)	(-4.23)	(-3.55)	(-5.40)
0.5 bin	0.64***	-0.20	0.45*	0.27	-0.16
	(2.74)	(-0.87)	(1.85)	(1.10)	(-0.53)
0.6 bin	1.26***	0.65**	1.37***	0.99***	0.68*
	(3.70)	(2.03)	(4.05)	(2.61)	(1.85)
0.7 bin	0.92**	0.61*	0.97***	0.91**	0.37
	(2.18)	(1.87)	(2.73)	(2.52)	(1.02)
0.8 bin	0.69	0.78*	0.69	1.65***	1.07**
	(1.35)	(1.79)	(1.58)	(3.57)	(2.22)
0.9 bin	1.36*	0.91*	0.93	1.68***	0.94*
	(1.90)	(1.92)	(1.39)	(3.76)	(1.87)
$\geq$ 1.0 bin	1.42***	0.74**	1.44***	2.40***	1.61***
	(3.90)	(2.20)	(4.59)	(6.11)	(4.46)
Fixed effects:					
Time	Yes	Yes	No	No	No
Stock	No	Yes	No	No	No
Investor	No	Yes	No	No	No
Stock × time	No	No	Yes	No	Yes
Investor × time	No	No	No	Yes	Yes
R^2 (in %)	0.61	2.56	15.71	19.58	38.09
Within R^2 (in %)	0.20	0.08	0.22	0.22	0.15
Number of observations	278,003	277,999	255,274	228,358	202,057

Continued on next page

Table IA.10 – Continued from previous page

	(1)	(2)	(3)	(4)	(5)
Panel B: Differences in the probability of increase					
Just below threshold – 0.2 bin	-0.11	-0.76***	-0.05	-0.31**	-0.66***
	(-0.85)	(-4.70)	(-0.31)	(-1.99)	(-3.34)
Just below threshold – 0.3 bin	-0.54***	-0.92***	-0.68***	-0.61***	-1.08***
	(-3.66)	(-5.72)	(-4.23)	(-3.55)	(-5.40)
Just below threshold – 0.5 bin	-1.18***	-0.72***	-1.13***	-0.88***	-0.92***
	(-5.04)	(-2.92)	(-4.49)	(-3.38)	(-2.89)
Just below threshold – 0.6 bin	-1.81***	-1.57***	-2.05***	-1.60***	-1.76***
	(-5.43)	(-4.85)	(-6.10)	(-4.22)	(-4.69)
Just below threshold – 0.7 bin	-1.47***	-1.53***	-1.65***	-1.52***	-1.45***
	(-3.52)	(-4.60)	(-4.77)	(-4.25)	(-3.98)
Just below threshold – 0.8 bin	-1.23**	-1.70***	-1.37***	-2.26***	-2.15***
	(-2.44)	(-3.86)	(-3.09)	(-4.79)	(-4.30)
Just below threshold – 0.9 bin	-1.91***	-1.83***	-1.61**	-2.28***	-2.02***
	(-2.68)	(-3.90)	(-2.43)	(-5.24)	(-4.18)
Just below threshold – ($\geq$ 1.0 bin)	-1.96***	-1.66***	-2.12***	-3.01***	-2.68***
	(-5.53)	(-5.03)	(-6.75)	(-7.64)	(-7.40)

Table IA.11:
Probability of increase: Fixed-effects regression approach with at-record-high interactions

This table repeats the analysis of Table IA.10 but interacts all coefficients with a dummy variable indicating whether positions are at the record high or not. The estimated coefficients are differences in the probability of increase between the bin just below the threshold and all other bins. Standard errors are clustered at the investor-stock and time level. The t-statistics are given in parentheses, and *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Just below threshold – 0.2 bin	0.07	0.96**	0.71**	0.57
	(0.22)	(2.42)	(2.29)	(1.60)
Just below threshold – 0.3 bin	0.05	0.37	0.57*	0.31
	(0.17)	(0.99)	(1.82)	(0.88)
Just below threshold – 0.5 bin	0.82**	0.80*	0.56	0.77
	(2.29)	(1.88)	(1.34)	(1.54)
Just below threshold – 0.6 bin	-0.79	-0.62	-1.00	-0.61
	(-1.53)	(-0.99)	(-1.59)	(-0.89)
Just below threshold – 0.7 bin	0.11	0.67	0.79*	1.02**
	(0.29)	(1.24)	(1.69)	(2.15)
Just below threshold – 0.8 bin	0.24	0.73	0.14	0.22
	(0.41)	(1.10)	(0.21)	(0.29)
Just below threshold – 0.9 bin	0.52	0.42	0.22	0.49
	(0.96)	(0.61)	(0.35)	(0.63)
Just below threshold – ($\geq$ 1.0 bin)	0.36	0.20	-0.47	0.41
	(0.76)	(0.39)	(-0.78)	(0.68)
Record high × (Just below threshold – 0.2 bin)	-0.94***	-1.05**	-1.15***	-1.37***
	(-2.86)	(-2.54)	(-3.32)	(-3.40)
Record high × (Just below threshold – 0.3 bin)	-1.21***	-1.11***	-1.41***	-1.68***
	(-3.67)	(-2.73)	(-3.92)	(-3.96)
Record high × (Just below threshold – 0.5 bin)	-2.26***	-2.55***	-1.88***	-2.37***
	(-4.94)	(-4.98)	(-3.63)	(-3.93)
Record high × (Just below threshold – 0.6 bin)	-1.01	-1.77**	-0.58	-1.59*
	(-1.59)	(-2.36)	(-0.78)	(-1.91)
Record high × (Just below threshold – 0.7 bin)	-2.65***	-3.35***	-3.59***	-4.20***
	(-3.73)	(-4.17)	(-4.97)	(-4.89)
Record high × (Just below threshold – 0.8 bin)	-2.92***	-2.79***	-3.45***	-3.59***
	(-3.42)	(-3.14)	(-3.75)	(-3.42)
Record high × (Just below threshold – 0.9 bin)	-3.46***	-2.66**	-3.50***	-3.36***
	(-3.84)	(-2.21)	(-3.64)	(-3.38)
Record high × (Just below threshold – ($\geq$ 1.0 bin))	-2.86***	-3.08***	-3.49***	-4.42***
	(-5.41)	(-5.08)	(-4.87)	(-6.18)
Record high	1.31***	2.30***	1.89***	2.15***
	(4.51)	(6.02)	(6.05)	(5.86)
Fixed effects:				
Time	Yes	Yes	Yes	No
Stock	Yes	No	No	No
Investor	Yes	No	No	No
Stock × time	No	Yes	No	Yes
Investor × time	No	No	Yes	Yes
R^2 (in %)	2.62	15.86	19.67	38.17
Within R^2 (in %)	0.14	0.39	0.33	0.28
Number of observations	277,999	255,274	228,358	202,057

Table IA.12:
The probability of increasing, decreasing, and not changing a position

Following the regression approach of Table IA.8, this table explores the probability of increasing, decreasing or not changing a short position in the bin just below the publication threshold relativ to other reporting bins. The results of Column 1 coincide with those of Table IA.10, Panel A, Column 5. The estimated coefficients are scaled to reflect changes in percentage points. Standard errors are clustered at the investor-stock and time level. The *t*-statistics are given in parentheses, and *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)
	Probability of		
	increase	no change	decrease
Just below threshold – 0.2 bin	-0.66***	0.44	0.22
	(-3.34)	(1.33)	(1.06)
Just below threshold – 0.3 bin	-1.08***	1.51***	-0.44**
	(-5.40)	(4.85)	(-2.15)
Just below threshold – 0.5 bin	-0.92***	2.42***	-1.50***
	(-2.89)	(4.82)	(-4.54)
Just below threshold – 0.6 bin	-1.76***	3.21***	-1.45***
	(-4.69)	(5.70)	(-3.51)
Just below threshold – 0.7 bin	-1.45***	3.06***	-1.62***
	(-3.98)	(4.80)	(-3.65)
Just below threshold – 0.8 bin	-2.15***	4.31***	-2.16***
	(-4.30)	(5.35)	(-4.14)
Just below threshold – 0.9 bin	-2.02***	3.29***	-1.27**
	(-4.18)	(4.14)	(-2.40)
Just below threshold – ($\geq$ 1.0 bin)	-2.68***	4.38***	-1.70***
	(-7.40)	(6.92)	(-3.94)
Fixed effects:			
Stock $\times$ time	Yes	Yes	Yes
Investor $\times$ time	Yes	Yes	Yes
R^2 (in %)	38.09	41.63	40.86
Within R^2 (in %)	0.15	0.24	0.11
Number of observations	202,057	202,057	202,057

Table IA.13:
Return co-movement within fund styles

This table reports the estimated coefficients of the regression

$$r_{j,t} = a + \beta_1 \mu_{J,t} + \gamma_j + \gamma_t + \varepsilon_{j,t}$$

with $r_{j,t}$ being the return of hedge fund j in month t. $\mu_{J,t}$ refers to the value-weighted style category return, where $j \in J$. γ_j and γ_t stand for fund and time fixed effects respectively. Standard errors are clustered by fund and month. *, **, and *** indicate significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)
	Dependent variable: Fund return		
	TASS	Clustering: 8 clusters	Clustering: 10 clusters
Cluster return	0.60***	0.70***	0.69***
	(13.81)	(17.28)	(18.81)
Month FE	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes
Within R^2	0.0348	0.1139	0.1237

Table IA.14:
Strategy distinctiveness of secretive positions

For each investor–stock pair, we determine the maximum bin reached in the sample period. We compare positions that never have been public but reached the 0.4 bin at least once (position maximum: 0.4 bin) with positions that have just exceeded the public disclosure threshold (position maximum: 0.5 bin). In Panel A, we compare the Strategy Distinctiveness Index assuming ten fund style categories based on value-weighted and equal-weighted style portfolios. We report means, the difference in means Δ and the respective p-value, as well as clustering standard errors at the investor–stock level. Here, N denotes the number of observations (investor–stock-days) in each group. In Panel B, we report the exact same statistics assuming eight style categories (Brown and Goetzmann, 2003). *, **, and *** indicate significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Position max.: 0.4 bin		Position max.: 0.5 bin		Diff. in means	
Variable	N	Mean	N	Mean	Δ	p-value
Panel A: Strategy distinctiveness with ten style categories						
Strategy Distinctiveness Index^{VW}	9,511	0.74	2,483	0.56	0.17***	(0.002)
Strategy Distinctiveness Index^{EW}	9,511	0.72	2,483	0.63	0.09**	(0.030)
Panel B: Strategy distinctiveness with eight style categories						
Strategy Distinctiveness Index^{VW}	9,511	0.72	2,483	0.56	0.16***	(0.007)
Strategy Distinctiveness Index^{EW}	9,511	0.73	2,483	0.64	0.09*	(0.063)

Table IA.15:
Flow–performance sensitivity and secretive positions

For each investor–stock pair, we determine the maximum bin reached in the sample period. We compare positions that never have been public but reached the 0.4 bin at least once (position maximum: 0.4 bin) with positions that have just exceeded the public disclosure threshold (position maximum: 0.5 bin). In Panel A, we compare the fund-specific flow-performance sensitivity measured over the previous 36 months of flow and return data. The estimates are updated annually. We report means, the difference in means Δ and the respective p-value, as well as clustering standard errors at the investor–stock level. Here, N denotes the number of observations (investor–stock-days) in each group. In Panel B, we report the exact same statistics of the flow-performance sensitivity measured over the previous 24 months. *, **, and *** indicate significance at the 10%, 5%, and 1% levels respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Position max.: 0.4 bin		Position max.: 0.5 bin		Diff. in means	
Variable	N	Mean	N	Mean	Δ	p-value
Panel A: Flow-performance sensitivity based on past 36 months						
Flow-Performance Sensitivity (Raw returns)	10,797	1.10	3,319	0.18	0.92***	(0.006)
Flow-Performance Sensitivity (CAPM alpha)	10,417	1.01	2,398	0.33	0.67*	(0.060)
Panel B: Flow-performance sensitivity based on past 24 months						
Flow-Performance Sensitivity (Raw returns)	10,797	1.47	3,319	0.19	1.28***	(0.002)
Flow-Performance Sensitivity (CAPM alpha)	10,417	1.66	2,398	0.40	1.27***	(0.000)

Table IA.16:
Robustness tests: Calendar-time portfolio approach

This table shows the performance of stocks for which there is at least one secretive short seller just below the publication threshold. The analysis is the same as in Table 10 with different sample specifications. In Panel A, we report the baseline results of Panel A of Table 10. In Panel B, we weight the stocks equally rather than by the number positions that are secretive and just below the publication threshold. In Panel C, we report results when we exclude penny stocks with a price below 1 EUR. In Panel D, we exclude a stock from the portfolio if a secretive short seller stays in the bin just below the publication threshold for less than five days. In Panel E, we use European asset pricing factors instead of local. To account for risk factors, we regress the portfolio excess return on the market excess return (MKTRF), the size (SMB) and value (HML) factors, and the momentum factor (WML). For brevity, the table reports only the alphas (in bps per day) of the time-series regression. The *t*-statistics are computed with Newey-West standard errors and are shown in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Returns are daily and the sample period is November 5, 2012 to March 31, 2015.

	(1)	(2)	(3)
	CAPM	Fama-French	Carhart
Panel A: Baseline result from Table 10			
Alpha	-5.46**	-6.13***	-5.66**
	(-2.00)	(-2.83)	(-2.57)
Panel B: Equal-weighted portfolio just below the threshold			
Alpha	-4.82**	-5.42***	-5.17***
	(-1.98)	(-3.02)	(-3.01)
Panel C: Stocks above 1 EUR			
Alpha	-5.35*	-6.01***	-5.55**
	(-1.96)	(-2.77)	(-2.52)
Panel D: At least 5 days just below threshold			
Alpha	-5.32**	-5.97***	-5.58***
	(-2.06)	(-2.99)	(-2.72)
Panel E: European pricing factors			
Alpha	-4.76	-5.45*	-5.81*
	(-1.57)	(-1.87)	(-1.95)

References

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

Angel, James J., Stephen E. Christophe, and Michael G. Ferri, 2003, A Close Look at Short Selling on Nasdaq, *Financial Analysts Journal* 59, 66–74.

Brown, Stephen J, and William N Goetzmann, 2003, Hedge funds with style, *Journal of Portfolio Management* 29, 101–112.

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri, 2011, Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records, *The Quarterly Journal of Economics* 126, 749–804.

Cohen, Lauren, Karl B Diether, and Christopher J Malloy, 2007, Supply and demand shifts in the shorting market, *Journal of Finance* 62, 2061–2096.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Engelberg, Joseph E., Adam V. Reed, and Matthew C. Ringgenberg, 2018, Short-Selling Risk, *The Journal of Finance* 73, 755–786.

Griffin, John M., 2002, Are the Fama and French factors global or country specific?, *Review of Financial Studies* 15, 783–803.

Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina, 2014, Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?, *Econometrica* 82, 463–505.

Kaplan, Steven N., Tobias J. Moskowitz, and Berk A. Sensoy, 2013, The Effects of Stock Lending on Security Prices: An Experiment, *Journal of Finance* 68, 1891–1936.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Prado, Melissa Porras, Pedro A. C. Saffi, and Jason Sturgess, 2016, Ownership structure, limits to arbitrage, and stock returns: Evidence from equity lending markets, *Review of Financial Studies* 29, 3211–3244.

Sun, Zheng, Ashley Wang, and Lu Zheng, 2012, The road less traveled: Strategy distinctiveness and hedge fund performance, *Review of Financial Studies* 25, 96–143.

Ward, Jr. Joe H., 1963, Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association* 58, 236–244.