Following the crowd: Anomalies and crowding by Institutional Investors

Ludwig Chincarini* Renato Lazo-Paz Fabio Moneta

January 13, 2022

Abstract

This paper investigates the relation between crowded trades, those in which many investors hold the same stocks possibly exhausting their liquidity provision, and institutional investors' trading activity on a set of twelve well-known stock market anomalies. We find that anomaly risk-adjusted returns appear to be concentrated among the most (least) crowded stocks for the long-leg (short-leg) portfolio Moreover, we find that our results remain significant after publication dates and are stronger among holdings of transient institutions. We hypothesize that crowded equity positions in anomaly stocks increase institutional investor's exposure to crash risk, liquidity risk, and fire sales. Our findings are consistent with this hypothesis and suggest that crowding adds a new consideration to the limits of arbitrage.

Keywords: *Crowding, Institutional Investors, anomalies, crash risk, liquidity risk, limits to arbitrage.*

JEL Codes: *G0, G11, G12, G14*

*Ludwig Chincarini (lbchincarini@usfca.edu) is at the University of San Francisco School of Management, Fabio Moneta(moneta@telfer.uottawa.ca) and Renato Lazo-Paz(rlazopaz@uottawa.ca) are at Telfer School of Management, University of Ottawa. We are grateful to conference and seminar participants at the AFA 2022 Ph.D poster session, University of Ottawa Bill's Brown bag, Federal University of Minas Gerais (Brazil) Brown bag, and the Research Institute for Finance and Economics conference (Peru) for helpful comments.

1 Introduction

A cornerstone idea in modern financial theory is the role that arbitrageurs play in creating efficient markets and to ensure prices reflect fundamental values (Grossman and Stiglitz, 1980). However, finding and exploiting mispricing opportunities can prove to be a risky challenge. Even assuming that if arbitrageurs can take long (short) positions in under (over) priced securities in a timely and cost-efficient way, they need to consider a set of additional limitations and risks such as transaction and holding costs (Pontiff, 2006), information uncertainty (Edmans et al., 2015), noise trader risk (De Long et al., 1990), short sales, and capital constraints (Shleifer and Vishny, 1997; Lam and Wei, 2011). Nevertheless, not all institutions face the same limits to arbitrage. For example, hedge funds are considered sophisticated investors that are subject to lower trading restrictions and have better access to capital and leverage.[1] Hence, the increasing participation of hedge funds would have a positive impact on the market's efficiency. However, Stein (2009) argues that arbitrage activity in the context of increased participation of sophisticated investors can be the subject of two negative externalities: the crowded-trade effect and excessive leverage. Both externalities add potential risks to arbitrage trading and can have a destabilizing effect on asset prices.

According to Stein (2009), the *crowded-trade effect*, also called crowding, surges when investors are unaware of the number of other investors actively implementing, by coincidence or intentionally, the same investment strategies.[2] Additionally, crowding has the potential to persist over time especially for non-fundamentally anchored investment strategies. These are strategies for which "*arbitrageurs do not base their demand on an independent estimate of fundamental value*" (Stein, 2009, p.1520). For instance, momentum or post-earnings announcement drift (PEAD) have the potential to be very profitable at times but these strategies are not subject to a price-based mechanism that signals when overpricing might be occurring. Investors might keep their positions as long as they are profitable. Similarly, many stock market anomalies trading rules are based on buying or selling stocks with certain firm characteristics that are considered to be low or high disregarding their current price. Ultimately, crowding can create a coordination problem that can negatively influence risk and return dynamics, making the risk of a trade endogenous to the trade itself (Lou and Polk, 2020, Antón and Polk, 2014).

The additional risk that crowded trades pose can exacerbate mispricing in specific market conditions such as during exogenous demand shocks. These events may force investors to simultaneously unwind their

[1] Many papers discuss the role of hedge funds in bringing securities prices closer to their fundamental values providing evidence in favor (e.g.,Stulz, 2007; Kokkonen and Suominen, 2015; Cao et al., 2018) and against (Brunnermeier and Pedersen, 2009). However, recent empirical evidence shows that some hedge funds may be successful in overcoming several limits to arbitrage (Hombert and Thesmar, 2014) and act as informed traders (e.g., Agarwal et al., 2013; Agarwal et al., 2013; Calluzzo et al., 2019).

[2] A very close related concept is herding. Herding occurs when a group of investors trade in the same direction over a period of time (Nofsinger and Sias, 1999), or applying similar trading styles (Wermers, 1999). The main difference is that crowding is directly linked to individual stocks liquidity. According to Chincarini (2018) crowding occurs when the number of investors chasing a similar strategy is too large given the available liquidity or typical turnover.

positions leading to fire sales (e.g., Coval and Stafford, 2007; Hau and Lai, 2017; Chernenko and Sunderam, 2020). This may be riskier for hedge funds given their reliance on short-term funding (Brunnermeier and Pedersen, 2009) as well as their tendency to increase their market exposure when market liquidity is low or volatility is high (Cao et al., 2013).[3] Therefore, it is of particular interest to institutional investors to identify when trading becomes crowded especially during times of market stress. For instance, in June 2018 the MSCI introduced their *MSCI integrated factor crowding models*[4] as means to offer investors a model that allows to quantitatively assess the degree of crowding in specific factor strategies and help them make a timely decision when facing increasingly crowded positions. Finally, because of institutional investors increased involvement in equity markets, regulators are also interested in analyzing the impact of crowding on overall financial market stability especially given the attention received by some episodes such as the tech bubble and burst of the early 2000s, the *"quant meltdown"* in August of 2007[5], the extreme drawdown of momentum strategies during the 2009 post-financial crisis rebound, and the recent COVID-19 induced 2020 Quant Deleverage (Chan and Tan, 2020).

In this paper, we argue that crowded equity positions pose additional risks to arbitrage trading through increased exposure to crash risk and fire sales. Moreover, we hypothesize that this relationship is more pronounced in a set of well-known asset pricing anomalies. Intuitively, investment strategies based on stock market anomalies are good candidates to become crowded as investors are aware of their existence once they are published (Mclean and Pontiff, 2016), and institutional investors trade to exploit them (Calluzzo et al., 2019). We aim to better understand the risks involved in the trading of anomaly stocks, in particular, the interaction between crowding, crash, and liquidity risk and the cross-section of anomaly stock returns. This focus on both crowding and anomalies, to the best of our knowledge, has not been explored in previous literature.

For our empirical analysis, we first analyze Thomson/Refinitiv 13F Institutional investors holdings database for the period 1980:Q1-2020:Q1. We start by estimating a broad set of crowding proxy measures, used in previous studies, both at the portfolio and stock level. However, we based our main results on the measure days-ADV proposed by Brown et al. (2019). We estimate days-ADV as institutional investors'

[3]It may also be the case, as in the Long-Term Capital Management's collapse, that other investor actions directly force the hand of more steady hedge fund investors that would otherwise *"rationally"* not trade (Chincarini, 2012) However, there is evidence that some hedge funds do act as liquidity providers under specific market conditions (e.g., Aragon, 2007; Aragon and Strahan, 2012).

[4]See `https://www.msci.com/www/research-paper/msci-integrated-factor-crowding/01025037754` for a detailed description of the model.

[5]During the week of August 6, 2007 many of the biggest and most successful equity hedge funds started reporting record losses. Khandani and Lo (2011) show evidence that such losses were concentrated among quantitatively managed equity-neutral funds. Moreover, the authors conclude that the crisis was due to market-wide deleveraging and a sudden withdrawal of market-making risk capital. Such massive, forced liquidations were the main reason behind the subsequent price impact. Chincarini (2012) offers a similar story based on the Long Term Capital Management (LTCM) meltdown in September 1998. One prominent hedge fund initiated a panic sell which then forced other similar quant hedge funds to reduce positions due to margin calls.

holdings divided by the average daily trading volume of each stock, both measured in money value. It can be interpreted as how many days would it take for institutional investors to exit their collective position in a given security. One advantage of this measure is that relates the size of institutional investors' holdings in a specific stock with its specific typical trading volume, in other words, its liquidity provision under normal conditions. Additionally, following Chen et al. (2019) we capture quarterly variations in crowding relative to its trend in the past four quarters, as an additional measure of changes in crowding level. We then proceed to link our crowding estimates with several measures of liquidity risk (Amihud, 2002) and multivariate crash risk (Chabi-Yo et al., 2019)

Our analysis provides several results. First, in line with the results of Sias et al. (2016), we find little evidence of a significant level of overlap in our sample of institutional investors holdings, at the aggregate level. Nonetheless, we show that cosine similarity among portfolios significantly increased during specific periods and has had a cyclical behavior starting around the year 2000. This cyclical behavior is observed also in other crowding measures at the stock level, which points out the relevance of the time-series component of crowding.

Second, we examine the relation between crowding and stock anomalies in the context of institutional investors' holdings. Every quarter we sort stocks into quintiles portfolios based on our preferred crowding variable (days-ADV) and then proceed to build long and short portfolios selecting the top and bottom quintiles as those most and least crowded, respectively. In this single sorting approach, we find that stocks in the highest crowding value-weighted quintile portfolio deliver a Fama and French (1993) 3-factor monthly alpha of 0.62% (t-value = 8.86), while for the lowest crowding quintile is -0.96% (t-value = 8.06). We employ a set of different asset pricing models and find that the results are robust.

Third, we test our hypothesis of crowding among anomaly stocks. As in Stambaugh et al. (2012) and Chen et al. (2019), we focus on twelve well-known anomalies. We begin by analyzing our full institutional investor's holdings sample from the first quarter of 1980 to the first quarter of 2020. However, as pointed out by French (2008) the expansion of institutional investors' direct ownership of US equities was accompanied by an enormous increase in turnover[6]. We recognize that our analysis might be influenced by exogenous effects and proceed to estimate structural breaks in the aggregate time-series of days-ADV for our full sample as well as for each stock anomaly. Our estimations find that there is a structural break around the year 1996 in the aggregate days-ADV time series and between 1992 and 1996 in our sample of stock anomalies. Moreover, we acknowledge that institutions crowding into stocks is related to firms' size and proceed to group anomaly and non-anomaly stocks in size deciles according to NYSE

[6]Among the explanations put forward by French (2008) are the development of electronic trading networks, decimalization of stock prices in the year 2000, as well as the progressive implementation of several SEC rules aimed at increasing market liquidity. See figure 8 in Appendix B that plots the average daily turnover for stocks included in the 13F Institutional Investors holdings database

size breakpoints.[7] Although with cross-sectional variations, there is evidence of crowding among anomaly stocks especially concentrated among the largest firms and for the long legs of the anomalies.[8]

Next, we examine the relation between crowding, as measured by days-ADV, and stock anomalies return. We find strong evidence that anomaly returns are concentrated among crowded stocks. Specifically, we define a portfolio that purchases the most crowded stocks (top 30% based on days-ADV sorting) in the long-leg anomaly portfolio and sell the least crowded stocks (bottom 30% based on days-ADV sorting) in the short-leg anomaly portfolio. On average, an equally-weighted portfolio across all the twelve anomalies exhibits significant risk-adjusted monthly return spreads of 1.87% (t-value = 12.19) for our full sample As shown by Mclean and Pontiff (2016) and Calluzzo et al. (2019) we observe alpha decay after publication to 0.90%, in line with the idea that institutional investors trade on anomalies once they became aware of them, but it remains highly significant (t-value = 9.78). Our findings provide evidence of arbitrageurs being able to recognize stocks with higher risk-adjusted returns (Chen et al., 2019; Brown et al., 2019). However, the presence of a significant alpha associated to the crowding portfolio of all stocks and anomaly stocks suggest that additional considerations might be limiting arbitrage activity (Shleifer and Vishny, 1997).

Finally, we explore two channels through which crowded equity positions increase poses additional limits to the arbitrage trading of anomaly stocks. One is through the exhaustion of a stock's liquidity provision, which exacerbates its liquidity risk. The other is related to the increased exposure to crash risk. We find that crowding, specifically days-ADV measure, is positively and significantly related to next quarter liquidity risk. Moreover, we find that this relation is stronger among transient institutions holdings which is in line with this institutions being more prone to crowd into similar trading strategies. In addition, higher crowding levels are related to an increase next quarter Amihud (2002) illiquidity measure and this effect is stronger among anomaly stocks.

Finally, we provide evidence that crowding significantly increases next quarter institutional investors holdings' crash risk. Our results are robust for a set of crash risk measures, the inclusion of a several of control variable, year and firm-level fixed effects, and regressions in subperiods in our sample. Moreover, we find that the relation between crowding and crash risk is stronger in the second part of our sample which is in line with the empirical evidence of increasing concentration of institutional investor's ownership in US equity.

Our paper contributes to several strands of prior research on the influence of institutional investors on asset prices and crowding. Recent research provides evidence that institutions trade on anomalies

[7] As a robustness check, we repeat this analysis based on all-firms (i.e., NYSE, AMEX, and NASDAQ) size breakpoint decile.Using these estimations delivers similar results.

[8] We confirm our results by focusing on changes to crowding levels, abnormal days-ADV, both anomaly long and short stocks show a significant difference in abnormal days-ADV. The results are again heterogeneous among anomalies.

especially if one focuses on hedge funds (Akbas et al., 2015) and after the publication of anomalies (Calluzzo et al., 2019). However, no studies address the concern that these positions may become crowded with potential crash and liquidity risks. Previous research (e.g., Sias et al., 2016; Brown et al., 2019; Chincarini, 2018) on crowding focuses on whether there is crowding at the security level or for a specific type of institution (e.g., hedge funds), but it does not examine whether there is crowding for a well-known sample of anomaly stocks. There is also some mixed evidence on the level of crowding in hedge funds holdings. Sias et al. (2016) document that hedge funds positions are relatively independent whereas Brown et al. (2019) find significant exposure to crowdedness. We extend their analysis by including a broader set of institutional investors and focusing on several well-known anomalies. Moreover, we study the channels through which crowded holdings influence stock returns as well as its relation to crash and liquidity risks in the analysis of anomaly trading. In this sense, we also contribute to the literature on crash risk and stock returns.

Although most of the prior research on crash risk has focused on the relation between information environment and extreme negative returns (e.g., Hutton et al., 2009; Chang et al., 2017; Cull and Xu, 2005) we add to recent literature that relates crash risk to the cross-section of stock returns (Chabi-Yo et al., 2019; Ruenzi and Weigert, 2018). We extend Ruenzi and Weigert, 2018 study on the effects of crash risk on momentum and show that this relationship holds for a broader set of stock market anomalies and its a significant component of the risk exposure of concentrated institutional investors holdings.

The remainder of the paper is organized as follows. Section 2 presents a brief discussion of the previous literature on crowding and links it to previous studies on the limits to arbitrage. In Section 3 we develop the hypotheses we will test in our empirical analysis. Section 4 describes both the data and our empirical methodology. Section 5 presents and discusses the main empirical results, and Section 6 concludes.

2 Related Literature

The term *crowded-traded problem* is first mentioned in the seminal work of Stein (2009). Motivated by empirical evidence on the growth of assets under management (AUM) and the number of institutional investors, especially in the hedge funds industry, Stein (2009) raises warning of the additional risks that those changes might bring to market efficiency. Specifically, increasingly crowded trading strategies and excessive leverage. Although the effects of the latter on arbitrage trading activity were not new at the time, the crowded-traded phenomenon was highlighted as a potential new risk consideration to arbitrage trading brought up by too many sophisticated investors trying to exploit similar investment opportunities unaware of potential liquidity exhaustion. The excessive leverage problem, on the other hand, is closely linked to studies on fire sales (e.g., Coval and Stafford, 2007; Chernenko and Sunderam, 2020).

From the perspective of investor's following each other's trading decisions, the *crowded-trade problem* can be related to literature on informational cascade, reputational interactions, social learning, and herding.[9] However, it adds a different approach to the discussion on why portfolios might become more similar by arguing that investors may collectively, intentionally or unintentionally, undertake the same trading strategies characterized by their disconnection from price-regulated mechanisms (Stein, 2009). Moreover, this problem is further enhanced when we consider increasingly larger institutional investments being allocated to the same security in relation to its liquidity provision. Therefore, the *crowded-trade problem*, although similar, proposes a different mechanism that might lead to too many investors holding the same stocks while adding another consideration, liquidity exhaustion. This dimension is crucial since crowding is, in essence, a measure that aims at capturing the difficulties, and related risks, a typical investor might face when trying to unwind trading positions with a high concentration of institutional ownership exposed to correlated demand shocks[10].

Recent studies have further considered additional reasons that might lead to crowding, specifically regulatory changes, copycat trading, and the rise of quantitative trading[11]. Hong et al. (2015) point out that increasing disclosure requirements regarding institutional investors' holdings, like the SEC 2004 regulation on the frequency of portfolio disclosure and the Dodd-Frank Financial Reform following the financial crisis of 2008, could lead to increasing crowding concerns. Although is still in debate if such strategies are profitable in a risk-adjusted and after-cost term, recent research has shown the incentives that investors face to free-ride on institutional investors' strategies and try to mimic the trades of past winners (e.g., Verbeek and Wang, 2013; Phillips et al., 2014). Therefore, it is possible that crowding into similar strategies was made easier due to increased visibility and access to the composition of some investors' holdings.

However, in the spirit of Stein (2009), crowding increases whenever we observe more investors undertaking similar *unanchored trading strategies*[12] in magnitudes that might lead to significant price dislocations when facing correlated demand shocks. A compelling case for such is the study of Khandani and Lo (2011). The authors argue that the quant meltdown of August 2007 was driven by a set of quantitative-driven strategies simultaneously signaling selling orders which exhausted liquidity provisions and lead to

[9]Hirshleifer and Hong Teoh (2003) provide an excellent review on those topics and its relation to the behavior of capital markets.

[10]For instance, Antón and Polk (2014) find that the degree of shared ownership forecasts cross-sectional variation in return correlation.

[11]See Chincarini (2012) for a comprehensive analysis of these phenomena.

[12]The idea of non-anchored strategies can be better understood by focusing on the most common example of this kind of strategy: momemtum. Lou and Polk (2020) argue that momentum makes the most interesting case to study due to (i) the inability of traditional asset pricing models to explain it, and (2) its positive-feedback nature, which means that investors do not base their demand on an independent estimate of fundamental value. As more investors engage in momentum trading they further exacerbates the return signals possibly leading to more investors undertaking similar positions. According to Stein (2009), any trading strategy that is based on a mechanism similar to that of momemtun is most exposed to crowding.

a sharp decline of some stock prices. Recent evidence shows that this is still the case if crowding is studied by testing popular multifactor models widely used by practitioners (Marks and Shang, 2019).

Hong et al. (2015) examine crowding from the perspective of short sales and finds that arbitrageurs require a premium for trading stocks for which closing or covering their positions is more difficult. Their proposed measure, days-to-cover (DTC), which is a ratio of short interest ratio to trading volume aims at measuring difficult to short stocks as those for which it would take more days, at current trading volumes, to close the position. In contrast, while Hong et al. (2015) focuses on short-selling activity only, we study the long positions of institutional investors given the fact that many institutional investors face significant restrictions on short-selling, e.g., mutual funds. This allows us to study the crowding-trade problem for a broader set of investors. Moreover, we for the specific case of several well-known stock anomalies given current evidence on institutional investors recognizing its existence (Mclean and Pontiff, 2016) and trading accordingly (Calluzzo et al., 2019). We also consider several crowding measures studying both the level and changes to consider changes in crowding over time and how is related to crash and liquidity risks.

Our paper is also related to Brown et al. (2019) who studies a sample of hedge funds holdings during the period 2006 -2017. They find that hedge funds take on highly concentrated positions that outperform less crowded ones, indicating possible skill in identifying profitable risk-adjusted opportunities. However, they find that crowding is a relevant component of hedge funds' tail risk. Funds exposed to more crowded positions suffer larger drawdowns especially during periods of market distress. We complement this paper by providing more evidence of the effect of crowding on anomaly stocks by institutional investors. Furthermore, we show that anomaly-returns are concentrated among most(least) crowded stocks and, although related to liquidity and crash risk, crowding represents an additional dimension of risk faced by arbitrageurs.

Finally, some recent papers examine whether crowding is an industry-specific concern, it is also noticeable in asset classes, and if it can be identifying employing multifactor models. Chen (2019) shows that the profits of a naturally unanchored strategy, namely industry momentum, in highly crowded industries tend to overshoot over short horizons but reverse over longer periods. Among the less crowded industries, however, results show delayed realizations of portfolio profits. Similarly Yan (2014) provides evidence that momentum crashes (e.g., Cooper et al., 2004; Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016) are influenced by crowded trades that push prices away from fundamentals leading to strong reversals. Concerning alternative investments, Baltas (2019) proposes the concepts of divergence and convergence premia to analyze the mechanism that may either further fuel mispricing, for non-anchored strategies, or provide a self-correcting measure. He finds that strategies such as momentum are subject to divergence premia and exhibits higher volatility following crowded periods, which calls for the use of volatility-targeting mechanisms From the perspective of specific multifactor models Marks and Shang (2019) show that stocks with strong signals (buy/sell) exhibit lower liquidity levels and lower volatility, consistent with correlated trades. Neither paper examines crowding in broader institutional holdings which might provide a partial picture of the role that different investors play in continuously more crowding investment spaces. Finally, linking the crowding-trade problem to related risk components, such as liquidity and crash risk, adds a dimension to the analysis of the limits to arbitrage.

3 Hypothesis Development

In this section, we develop our main hypotheses for empirical analysis. Through testing these hypotheses, we attempt to better understand the relation between crowding, institutional investors trading, and stock market anomaly returns.

A cornerstone idea in modern financial theory is the role that arbitrageurs play in creating efficient markets and ensuring prices reflect fundamental values (Grossman and Stiglitz, 1980). However, finding and exploiting mispricing opportunities can prove to be a risky challenge. Even if we assume that arbitrageurs can take long (short) positions in under (over) priced securities in a timely and cost-efficient way, they need to consider a set of additional limitations and risks. For example, transaction and holding costs (Pontiff, 2006), information uncertainty (Edmans et al., 2015), noise trader risk (De Long et al., 1990), short sales, and capital constraints (Shleifer and Vishny, 1997, Lam and Wei, 2011). Nevertheless, not all investors face the same limits to arbitrage. For example, hedge funds are subject to lower trading restrictions and have better access to capital and leverage . Thus, in the scenario of increasing securities ownership by institutional investors, specially hedge funds, crowding should be unrelated to stock returns in the cross-section since it might signal increasing forces correcting any potential mispricing. However, if holdings stocks with concentrated ownership exposes investors to more pronounce price decline, then crowded positions will carry additional risk for which arbitrageurs would require a premium. Hence, our first hypothesis is about the relation between crowding and the returns of institutional investors' holdings.

Hypothesis 1 (Crowding and the return dynamics in institutional investors' holdings): Investors require compensation for trading in a crowded space and therefore crowding is positively associated with stock excess returns.

While institutional investors on an aggregate level mostly hold the market portfolio (Lewellen 2011), there is evidence that some of them follow academic publications and engage in anomaly-based trades (Mclean and Pontiff, 2016, Calluzzo et al., 2019) For instance, Calluzzo et al. (2019) document a shift on the portfolio holdings of some institutional investors toward anomaly-ranked stocks, especially after their publication. As pointed out by Stein (2009) for some anomaly-based trades arbitrageurs do not base their demand on an independent estimate of fundamental value. For instance, momentum or post-earnings announcement drift (PEAD) have the potential to be very profitable at times but these strategies are not subject to a price-based mechanism that signals when overpricing might be occurring . Investors might keep their positions as long as they are profitable. Therefore, if institutional investors rely on a similar set of anomaly stock characteristics (e.g., past year return, gross profitability, return on assets) when trading, it can be expected that market anomalies are the prime candidates of investment strategies prone to become crowded. This rationale leads to our second hypothesis.

Hypothesis 2 (Crowding in anomaly stocks): Crowding levels should be higher among stocks that have similar anomaly characteristics.

Stein (2009) argues that crowding surges when investors have imperfect information on the number of other investors actively implementing the same investment strategies and the liquidity characteristics of those positions. If the demand for a specific stock is uncorrelated among investors, then many investors holding the same stock would not lead to price volatility since their demands would mostly cancel out (Ben-David et al., 2021) In contrast, if buy (sell) signals are correlated, as when investor implement similar strategies, demand shocks have the potential to impact asset prices through fire sales (e.g., Coval and Stafford, 2007, Chernenko and Sunderam, 2020). Moreover, the impact is conditional on the liquidity characteristics of each position. These conditions impose greater risk to arbitrageurs holding those securities by increasing concerns about liquidity risk and exposure to crash risk due to correlated demand shocks (Chang et al., 2017). Furthermore, institutional investors are a largely heterogenous group. For instance, short-term institutions are those with higher portfolio turnovers due to more active trading (Yan and Zhang, 2009). Previous studies show that short-horizon institutions are known to exert greater selling pressure than long-horizon peers during periods or market turmoil (Cella et al., 2013) The third hypothesis considers the potential influence of crowding on liquidity and crash risk as well as its link to investment horizon.

Hypothesis 3 (Crowding, liquidity, and crash risk): Crowding is positively related to liquidity and crash risk and this relation should be stronger among stocks with greater ownership by short-horizon investors..

4 Data and sample description

4.1 Institutional Investors' Holdings

We use Thomson/Refinitiv (TR) 13F database to collect data on Institutional Investors' portfolio holdings. The Security Exchange Commission (SEC) regulation requires all institutional investors that exercise investment discretion on assets under management over $100 million to report their end-of-quarter holdings greater than 10,000 shares or $200,000 on the Form 13F within 45 days of each quarter-end. We then proceed to merge our holdings database with data on stock prices, volume, total shares outstanding for each stock from the Center for Research in Security Prices (CRSP). As commonly performed in previous studies, we capped institutional ownership to 100% whenever the number of shares held is greater than the number of shares outstanding (Calluzzo et al., 2019). We exclude stocks with a share price of less than $5 as well as utilities and financial firms from our sample. The exclusion of microcaps alleviates concerns about anomaly-returns being driven by penny stocks and reduces the effect of potential market microstructure noises.

[Insert Figure 1 about here]

Figure 1 depicts time-series means of cross-sectional medians of several characteristics of the 13F database over time. As shown in Figure 1(A) the proportion of shares outstanding owned by institutional investors (IO) has steadily increased over the years reaching its peak of almost 79% around the year 2019. However, more surprising is the sharp decline, and subsequent rebound, on IO at the end of the year 2019 and the first quarter of 2020. This might be arguable the effect of the world's covid-19 pandemic. This V-shaped behavior at the end of our sample is also observed in the following figures. Figure 2(B) plots the median number of institutional investors that hold the same security. At its peak, in the year 2019, typical security in our sample was owned by 160 different institutional investors. Figure 2(C) shows the decline in the median number of stocks held in a typical institutional investor's portfolio (red line) contrasted to the increase in the amount of money, in millions of USD, allocated in average security (blue line). Finally, as shown in Figure 2(D) institutional investors now face a context of an increased number of investors (blue line) that have access to a smaller pool of available securities (red line). Between 1980 and 2020, the number of institutional investors included in the 13F Institutional holdings database grew more than 10 times from around 400 to more than 4,000. By comparison, the number of publicly listed companies included in that database reached 5,756, its peak, in the late 1990s, and had continuously decreased over the last 20 years to a total of 2,386 in 2020.

In our base sample, we include all institutional investors considered in the 13F database. However, there is vast evidence on the differences in trading behavior among institutional investors[13]. For that purpose, we subset our sample into transient and non-transient institutions. We use data from Brian Bushee's website to identify them in our sample. It is of our interest to pay special attention to transient institutions due to their active management approach to trading [14]. Moreover, this classification allows us to extend the analysis of previous studies that focused on hedge funds only to include additional institutional investors that could also actively look for arbitrage opportunities.

4.2 Stock anomalies

We consider twelve well-known stock market anomalies following Fama and French (2008) and Stambaugh et al. (2012). Table 1 describes each stock anomaly and reports its publication years. We follow the empirical methodology widely used in previous studies and create each anomaly portfolio on June 30$^{\text{th}}$ of each year in our sample. We then proceed to rank stocks into quintile portfolios and form long and short portfolios according to each anomaly variable sorting rule. For instance, the accrual anomaly first documented by Sloan (1996) shows that a portfolio comprised of a long position on firms with low accruals and shorts those with high accruals delivers statistically significant abnormal returns. On the other, the gross profitability anomaly (Novy-Marx, 2013) finds that firms that reported higher gross profitability outperform those with lower returns. We employ Compustat and CRSP databases to obtain the financial data needed to estimate each of the anomaly variables. For the anomalies estimated based on accounting data, we used information from the last fiscal year in calendar year (t – 1) to ensure that we employed information available to investors at the time of the portfolio formation.

[Insert Table 1 about here]

For our main results, we analyze annually ranked anomaly portfolios. Nonetheless, recent studies (Han et al., 2021) have documented increased performance of several anomalies portfolios when rebalanced at a higher frequency. Those studies argue that rebalancing anomaly portfolios once a year does adequately incorporate valuable information produced during the year. We consider these arguments and re-run our main specification on quarterly ranked anomalies.

4.3 Crowding measures

One major challenge in measuring crowding in equity markets is capturing the simultaneity in capital allocation to specific strategies while considering liquidity concerns. Moreover, given the restrictions that

[13] See, for example, Calluzzo et al. (2019) and Edelen et al. (2016) for recent discussion on the topic.

[14] According to Calluzzo et al. (2019), the quarterly average portfolio turnover of transient institutions is 66.8% while for non-transient investors is 25%. Regarding which institutions are considered as transient, according to the authors, 34.1% are hedge funds, 58.6% mutual funds and the remaining 7.3% includes bank trusts, insurance companies, pension funds, and endowments.

many institutional investors (e.g, mutual funds) face entering short positions, it is most likely that many investment strategies are based on long-only mandates. On the other hand, investors such as hedge funds are significantly less restricted to include complex investment strategies involving the use of derivatives, leverages, and holding short positions[15]. Therefore, it is unlikely that a single measure can capture crowding for all potential investment strategies while including considerations about liquidity provisions.

Concerning the tendency of investors to follow the same strategies, several studies proposed examining the degree level of overlap between investors' portfolio holdings (Sias et al., 2016; Chincarini, 2018; Blocher, 2016). The preferred measure is the cosine similarity (CS_{ij}) which is estimated as the dot product between the position weight vectors w of each portfolio i and j divided by the product of the Euclidian norm of each vector.

$$CS_{ij} = \frac{w_{ij} w_{jt}}{|w_{ij}||w_{jt}|} \tag{1}$$

Where w_{it} and w_{jt} are investors i and j vector of portfolio weights at each quarter-end t, respectively. The value of CS_{ij} is bounded between 0 and 1 for long-only portfolios. Two identical portfolios have CS_{ij} equals 1 whereas if the portfolios are completely different the value is zero. Additionally, we follow Chincarini (2018) and estimate the aggregate cosine similarity as the average similarity between our sample of institutional investors' portfolios.

In panel A of Figure 2 we plot the aggregate cosine similarity for the complete 13F holdings database for the sample period between 1980:Q1 and 2020:Q1. Consistent with Sias et al. (2016) we observe a decay in the overall similarity among institutional investors' portfolios. We extend their findings and provide evidence that the decrease in overlap among hedge funds occurs also in the broader sample of 13F institutional investors. However, starting in the year 2000 we observe a cyclical behavior. First, there is a progressive decay in overall similarity until the year 2009, coinciding with the financial crisis of 2008-2009. In the following years we observe a sharp increase in overall aggregate similarity that remained fairly stable until it began decreasing again around the year 2018. Interestingly, our results coincide with the positive trend in the aggregate days-ADV measure among hedge funds documented by Brown et al. (2019) during the period between the years 2004 and 2017.

We further explore the observed relation between aggregate days-ADV and aggregate cosine. In panel C of Figure 2 we focus on the top (bottom) 5% percentile of funds that showed the highest(lowest) cosine similarity and estimate the average days-ADV for that subsample. For both time series we observe the

[15] It is worth noting that, as documented by Calluzzo et al. (2019), although restricted on holding short positions, there is an increase allowance on mutual funds accessing leverage, derivatives, and hold illiquid assets. However, evidence shows that the loosening on the restrictions on the use of complex instruments by mutual funds is not accompanied by an increase in performance but by poor results.

same general downward trend documented for the full sample. However, while there is cyclical pattern in the time-series of aggregate days-ADV among the most similar funds, this series remain fairly stable for the sample of the least similar funds.

A limitation of holding-level measures such as the cosine similarity is that it does not fully captures the impact of crowding on prices unless it is linked to a liquidity provision measure (Beber et al., 2012). Additionally, this approach is somehow limited by the inability to observe other portfolio components such as short positions widely used by hedge funds[16]. It is due to these limitations that alternative measures were proposed to examine crowding at the stock level since it is possible that, although two portfolios have very low cosine similarities, they might still hold very concentrated positions on specific securities.

Regarding stock-level measures of crowding, two general approaches have been used in previous studies: (i) measuring the level concentration of ownership among securities (Beber et al., 2012) and (ii) relate investor's holdings with securities daily trading activities (Zhong et al., 2017; Brown et al., 2019). Intuitively, three basic measures of ownership concentration are the total number of institutional investors invested in individual security at time t ($NINST$) and the security's percentage of shares outstanding owned by institutional investors in a given period t (IO), and the total amount of money invested in security i at time t ($INVST$). However, concentration measures by themselves do include considerations regarding the liquidity provision of the specific security. Based on this rationale Zhong et al. (2017) proposed a crowding measure called Activity Ratio ($ActRatio$) as the ratio of the percentage of share i held by an investor at the end of the quarter (t – 2) divided by the average share turnover of the stock i during the quarter (t-1). Where the stock's average monthly turnover is measured over the previous quarter.

$$ActRatio_{i,t} = \frac{Shares_{i,t-2}}{AvgTurn_{i,t-1}} \tag{2}$$

Where higher values of $ActRatio_{i,t}$ stands for a more crowded position in a given stock. Zhong et al. (2017) argue that the potential mispricing produced by crowded trades might persist for an extended period. Two main forces drive long-term persistence. First, investors are unable to observe the current arbitrage capacity and its level of usage. Second, the use of investment strategies not directly linked to firm's fundamentals limits arbitrageur's ability to rely on price signals to coordinate their trades. For such reasons, even if the information included in AcR is based on lagged signals, such information may not be fully priced yet. Alternatively, Brown et al. (2019) proposes a very similar measure called *days-ADV* defined as the total amount, in money terms, invested in a security relative to the security's average

[16] A remarkably exception is the work of Girardi et al. (2021) who study portfolio holdings similarity in the insurance industry. With this more complete view of insurers holdings, the authors conclude that insurers whose portfolios are more similar experience larger common sales that impact prices when shocks to their assets or liabilities occur

daily trading volume over the past quarter.

$$DaysADV_{i,t} = \frac{TotMoneyInvest_{i,t}}{AvgDailyTurn_{i,t}} \tag{3}$$

The days – ADV measure can be interpreted as how many days, under typical trading volume activity, it would take the selected investors universe to exist its collective position.

Figure 2 depicts the time-series of the aggregate cosine similarity as well as the adv days and activity ratio (*ActRatio*) measures over time.

[Insert Figure 2 about here]

Panel A of Table 2 summarizes the cross-sectional distribution of the selected crowding measures. We find NINST, number of institutions holding the same security, to have a rather skewed distribution in which while some stocks are held by as few as one investor only, some others are held by a significant proportion of the total number of institutional investors. A similar distribution is seen when examining $INVST$. Additionally, our results show a significant increase in IO compared to that documented by Yan and Zhang (2009) who finds a mean IO of 25.10% for the period 1980:Q1 to 2003:Q4. This is in line with many papers documenting the rise in the participation of institutional investors in equity markets. Finally, the distribution of days-ADV measure closely follows that documented by Brown et al. (2019). Nonetheless, our values differ in magnitude because we consider the complete universe of 13F institutional investors and not hedge funds holdings only.

Panel B of Table 2 reports cross-sectional correlations among our sample of crowding measures. The highest correlation is observed between the number of institutions ($NINST$) and total value invested ($INVST$) which although may seem obvious it is only in the order of 0.78. This might reflect the fact that a relatively small subset of institutions holds significant amounts of money invested in certain securities[17] .As expected, the measure days-ADV shows a low correlation with the other variables possible indicating that it can capture additional behaviors t-stat included in simpler measures related to the concentration of ownership.

[Insert Table 2 about here]

Figure 3 plots the time-series mean of the cross-sectional median of days-ADV for transient (blue line, right axis) and non-transient (green line, left axis) institutional investors from 1980 through 2018. Panel A

[17] In a recent study, Ben-David et al. (2018) shows that as of December 2016, the largest institutional investor in the US market was responsible for managing a portfolio equivalent to 6.3% of the total equity market and the top 10 investors held assets under management equal to 26.5% of those assets. However, according to their results, this increased concentration of ownership among a small number of institutional investors is related to higher volatility and greater noise in stock prices

shows a sharp decline in days-ADV for both investors during the 1990s, in line with the significant increase in turnover documented by French (2008). However, while crowding seems to stay rather constant for our sample of non-transient investors, days-ADV shows an upward slopping behavior in the transient investors' holdings. Given that days-ADV is a function of the size of each investor's holdings in a particular stock, we can desegregate each security total days-ADV value into the proportions held by transient (blue line, right axis) and non-transient investors (green line, left axis). Panel B plot this pattern. Transient institutions have steadily increased, on average, their participation in a typical security days-ADV measure.

[Insert Figure 3 about here]

4.4 Crash risk measures

Crash risk proxy variables aims at capturing higher moments of the stock return distribution with a special interest on extreme negative returns Habib et al. (2018). Theoretically, crash risk is based on the notion that investors expect higher returns for stocks with more negative skewness, implying that skewness is a priced risk factor Harvey and Siddique (2000).

Following Hutton et al. (2009) and Callen and Fang (2015) we define crash risk using *daily* firm-specific return using the residuals from equation 4. As stated by **?** using actual returns would lead to biased inference since many crashes would be expected during times of market turmoil as well as jumps during recovery periods. A more suitable approach is to look at *residual returns* to better assess extreme movements.

$$r_{j,t} = \alpha_j + \beta_{1,j} r_{m,t-1} + \beta_{2,j} r_{i,t-1} + \beta_{3,j} r_{m,t} + \beta_{4,j} r_{i,t} + \beta_{5,j} r_{m,t+1} + \beta_{6,j} r_{i,t+1} + \epsilon_{j,t} \tag{4}$$

Where $r_{j,t}$ is the return on stock j in day t, $r_{m,t}$ is the return on the CRSP value-weighted market index in day t, and $r_{i,t}$ is the return on the value-weighted industry index based on the two-digit SIC code. The inclusion of both lead and lag terms of the value-weighted market and industry indices aims at correcting the effect of non-synchronous trading Dimson (1979). However, the estimated residuals from equation 4 are highly skewed. Since several crash risk measures are based on the difference in the number of standard deviations above or below a reference return we *log* transform the residual returns ($log(1 + \epsilon_{j,t})$) to allow for a more symmetrical distribution.

Following the common practice in the literature we will estimate three measures to calculate crash risk. The first is the negative conditional skewness of firm-specific returns, *NCSKEW*, estimated as the negative of the third moment of firm's specific daily returns divided by their cubed standard deviation.

$$NCSKEW_{j,t} = -\frac{n(n-1)^{3/2} \sum R^3_{j,t}}{((n-1)(n-2)(\sum R^2_{j,t})^{3/2})} \tag{5}$$

where n is the number of observations per firm j during the fiscal year. t. Since an increase in NCSKEW points out to a stock's return having more left-skewed distribution, we follow the convention that higher NCSEW value implies a higher *crash risk*. The second measure is called *down-to-up volatility* (DUVOL) and is estimated as shown in equation 6. This measure captures the asymmetric volatility of positive and negative firm-specific daily returns.

$$DUVOL_{j,t} = log\left(\frac{(n_u - 1)\sum_{DOWN} R^2_{j,t}}{(n_d - 1)\sum_{UP} R^2_{j,t}}\right) \tag{6}$$

For a given firm j we count the number of days with returns above (n_u) and below (n_d) the daily mean. Then, we proceed to estimate the log ratio of the standard deviation of the sample of *up days* and the sample of *down days*. Similar to the $NCSKEW$ measure, an increase in $DUVOL$ points out to a firm becoming more *crash-prone*.

Finally, we follow Callen and Fang (2015) and estimate $CRASH - COUNT$ as the number of firm-specific daily returns exceeding 3.09 standard deviations above and below the mean firm-specific daily return over that fiscal year t. According to Hutton et al. (2009) this specific value allows to capture the observations that conform the 0.1% in the normal distribution. Therefore, the frequency of crash events, $CRASH - COUNT$, is estimated as the number of downside events minus the number of upside events. A higher value of $CRASH - COUNT$ then signals a higher frequency of negative return days.

4.5 Liquidity, and liquidity risk measures

We start by estimating Amihud's (2002) illiquidity measure, which is defined as:

$$Illiquid_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{j,t}} \frac{|R_{j,t,d}|}{V_{j,t,d}} \tag{7}$$

where $D_{i,t}$ is the number of observations with volume data in a given month t, $|R_{j,t,d}|$ is the absolute daily return of stock j over month d, and $V_{j,t,d}$ is the daily dollar volume for stock j over month d. We obtain the monthly aggregate value of the Illiquidity measure by averaging the values all days with trading data in each month.

For the liquidity risk measure, we estimate the *liquidity beta* as the parameter loading on the Pastor and Stambaugh (2003) traded liquidity factor added to the Fama and French (1993) three-factor model.

$$R_{j,d} = \alpha_{j,d} + \beta^{mkt}_{j,d} MKT_d + \beta^{size}_{j,d} SMB_d + \beta^{value}_{j,d} HML_d + \beta^{size}_{j,d} LIQ_d + \epsilon_{j,d} \tag{8}$$

where $R_{j,d}$ denotes the monthly excess return for each stock in our sample. We estimate the *liquidity beta* ($\beta^{liq}_{j,d}$) for each month using a rolling estimation on monthly return over the past 60 months. Table 11 reports the time-series average of cross-section means of both the liquidity beta and the Amihud (2002)illiquidity measure for a series of long and short portfolio formed on crowding measures.

5 Empirical Analysis

In this section, we test our hypotheses about the relation between crowding, anomaly stock returns, and crash risk.

5.1 Crowding and the cross-section of stock returns

We first use a single portfolio sorting approach to examine the effect of crowding on the cross-section of our sample of quarterly institutional holdings. We begin by forming quintile portfolios of stocks at the end of each quarter based on four crowding measures: *IO*, *NINST*, Days-ADV, and *ActRatio*. Then, we estimate monthly excess return in both equal and value-weighted portfolios and form a spread portfolio by taking long(short) positions on stocks with high(low) crowding values, according to each proxy variable. We also adjust for factor exposure using the Fama (1998) three-factor model (FF3).

[Insert Table 3 about here]

Panel A of Table 3 reports the FF3 alpha of high, low, and *high-minus-low* (diff) portfolios in our sample period of 1980:Q1 to 2020:Q1 for our sample of crowding measures. Consistent with Brown et al. (2019), we find a significant annualized alpha for the value(equally) weighted portfolios sorted on days-ADV. On average, a value-weighted portfolio composed of highly crowded stocks (quintile 5) delivers an annualized alpha of 2.46% (t-stat = 6.87), whereas one that includes the least crowded stocks (quintile 1) offers an alpha of -4.04% (t-stat = 7.02). difference value-weighted portfolio (*high-minus-low*) has an annualized alpha of 6.50% (t-stat = 7.88). Our results for portfolios sorted on the *ActRatio* measure are somehow similar, however, the economic magnitude of the returns is significantly lower than those observed in the *days-ADV* sorted portfolios[18].

[18]Our results differ from those of Zhong et al. (2017) who report that a low-minus-high portfolio sorted on their *ActRatio* can generate an annualized risk-adjusted return of 14.53%. We argue that one main difference with our empirical design, specifically their focus on active mutual funds only, might contributing to such differences. Moreover, the fact that the

Additionally, we fail to find significant alphas for portfolios sorted on either *IO* or *NInst*. These results suggest that, over the full sample period, abnormal returns cannot be obtained by taking long and short positions on stocks with the most and least institutional ownership, respectively. Similarly, forming portfolios based on the number of institutions that simultaneously hold the same stock, does not provide significant alphas. This discrepancy in our results might show that securities held by many institutional investors are not necessarily crowded unless it is related to the specific security liquidity provision.

In panel B of Table 3, we repeat our estimations for portfolios rebalanced every 4 quarters (Q_{t+4}) and every 8 quarters (Q_{t+8}). The results show that the performance of portfolios sorted on crowding variables holds for different rebalancing frequencies.

Brown et al. (2021) highlights three advantages of the days-ADV measure: (i) widely used by practitioners , (ii) it is a measure with an intuitive interpretation, and (iii) can be further decomposed into illiquidity and size components. Motivated by such advantages and our initial results we decide to base the rest of our empirical analysis on the days-ADV measures. We further expand the test of hypothesis 1 about whether crowding influences the returns of our sample of institutional investors' holdings. For such, we focus on the excess-return of the portfolio sorted on days-ADV measure while controlling for a wider set of factors included in several widely known asset pricing models. Specifically, the Fama (1998) three-factor model augmented with the Carhart (1997) momentum factor (FF4); the Fama and French (2006) five-factor model that additionally controls for profitability and asset growth (FF5); the Pastor and Stambaugh (2003) traded liquidity factor [19] added to the FF3 model (FF3+liquidity); and the mispricing model of Stambaugh et al. (2015) (MISP)

[Insert Table 4 about here]

In Panel A of Table 4 we report the excess return and risk-adjusted return for quintile portfolio sorted on days-adv for our full sample period. The results in the first column show that, on average, the most crowded stocks (quintile 5 - high) earn a monthly excess return of 1.18% (t-stat = 6.31), whereas the least crowded stocks (quintile 1 – low) have a monthly excess return of -0.14% (t-stat = -0.46). The high-minus-low portfolio (HML) earns a monthly excess return of 1.32% (t-stat=6.24). The return of the most crowded portfolio (quintile 5 – high) is lower but remains significant after controlling for the risk factors considered in each asset pricing model. Similarly, the portfolio that holds the least crowded stocks (quintile 1 – low) earns lower adjusted returns. The monthly alphas for the high-crowding portfolio range

ActRatio is based on lag data makes us wonder if their measure is actually capturing mutual funds decision to take on positions previously crowded by other participants, such as hedge funds. We further explore this alternative explanation in later sections.

[19] We obtain the values for the liquidity factor from Lubos Pastor's website http://finance.wharton.upenn.edu/~stambaug/

from 0.62%, FF3, to 0.34%, MISP, whereas the alphas of the least crowded portfolio span from -0.96%, FF3, to -0.37%, MISP. Accordingly, the alphas for the high-minus-low portfolio span from 1.58%(t-stat = 9.52) , in the FF3 model, and to 0.71% (t-stat=4.19) in the MISP model. Consistent with Stambaugh, Yu, and Yuan (2012) we find that the alphas of the HML portfolio come mostly from the short leg. Moreover, the adjustment for liquidity risk, in the FF3 + liquidity factor, does not significantly reduce the performance of the HML portfolio. This result is informative about the role that crowding might play for institutional investors' trading, that although related to liquidity, it seems to represent a distinct risk concern, in line with our first hypothesis.

It is possible to argue that our results may be driven by the first part of our sample in which we observe significantly higher values of the days-ADV measure. In Figure 2 Panel B is possible to identify two distinct periods that, as previously discussed, may indicate changes in the trading behavior over our sample period (see Fig1A in Appendix B). Since our main crowding measure, days-adv, its a function of the daily trading volume, these changes may influence our results. We perform a structural break analysis of the time-series mean and median of day-adv measure and find a common break in the year 1995. The details of the analysis are indicated in Appendix B. In Panel B of Table 3, we address these concerns and find that the alpha of the HML portfolio is statistically significant in both subperiods.

Next, we perform Fama and MacBeth (1973) cross-sectional regressions to examine the influence of crowding on future stock returns, while controlling for other variables identified to influence institutional investors demand (Yan and Zhang (2009), Calluzzo et al. (2019)). For each quarter we run a cross-sectional regression of average monthly excess return and cumulative monthly returns over the next quarter and over the next year on the *log of the days-ADV* measure along with the control variables. As detailed in the descriptive analysis of the crowding measures (table 2) the days-ADV measure shows a very skewed distribution with wide dispersion between values. To reduce the effect that very distinct values might have in our estimations we log transform our variable to limit this effect.

The control variables include: institutional ownership, market capitalization (size), the number of months since stock's first appears in CRSP (age), the standard deviation of monthly returns over the previous two years, book-to-market ratio, dividend yield, average monthly turnover over the past three months, cummulative return over the past three months, cummulative return over the past nine months preceding the beginning of quarter. We use natural log of all control variables with the exception of cummulative returns.

[Insert Table 5 about here]

Table 5 report the results of the Fama-Macbeth regressions. Panel A of table 5 shows our results when the regression on next quarter returns. We find the regression coefficient on the log(ADV) measure to be significant with the expected signs for both dependent variables and for each subperiod according to the previously estimated structural break in the days-ADV series. Moreover, we observe that the magnitude of the regressions parameters increases both in value and statistical significance for the second part of the sample period. This evidence is in line with the evidence on increasing crowding and suggest that its effect of on future stock returns has increase over recent years.

In panel B of table 5, we address the question of whether crowding has a short-lived impact on future stock returns. Although the magnitude of the parameter coefficients is reduced in the cross-sectional regression of cummulative returns (from 0.559 to 0.312 and 0.599 to 0.303 for each subperiod, respectively), these values remain highly significant. Moreover, the reduction in parameter value is not observed when regressing on excess returns.

Taken together, the evidence in this subsection provides evidence for Hypothesis 1 that crowding is positively related to the return of institutional investors' holdings. However, we find that the choice of crowding measure matters. We show empirically that unless linked to a liquidity variable, the number of investors owning the same security, or the proportion owned by institutional investors does not necessarily signal crowded positions.

5.2 Are anomaly stocks crowded?

In this section, we examine whether stocks included in the long and short legs of our set of anomaly variables, what we call anomaly stocks, tend to become more crowded than comparable non-anomaly stocks. However, it relevant to include two important considerations: (i) It is very likely that anomaly strategies show time-varying levels of crowding. This implies that some anomaly stocks might become crowded while others do not. (ii) Some firm characteristics might influence investors' preference when forming their portfolios so any analysis of crowding needs to control for these variables. We include these two considerations into our empirical analysis. First, we estimate the time series days-ADV for each anomaly strategy for the long, short, and intermediate quintiles (non-anomaly) and non-anomaly stocks in our sample. Then, we examine crowding levels for those subsamples taking into account firm's size. Finally, we examine crowding among stocks that are uniquely in the long or short legs of each anomaly.

In untabulated results, we observe heterogeneity in days-ADV values among stock anomalies included in the long legs (quintile 5) and the non-anomaly stocks (quintiles 2-4) while stocks in the short leg (quintile 1) are usually less crowded for most anomalies. The anomalies for which the stocks in the long legs were the most crowded are NSI, CEI, B/M and OSC. On the other hand, in the case of ACC, NOA, GP and AG the non-anomaly stocks seem to be more crowded. Finally, consistent with our assumptions about un-anchored strategies, we find a cyclical pattern among MOM anomaly stocks characterized by periods in which we observe crowded long stocks and uncrowded short stocks and vice versa. We next examine crowding while controlling for differences in firm's size. For that purpose, we sort both anomaly and non-anomaly stocks by size and group them according to NYSE size decile breakpoints . Figures 4 to 6 show the days-ADV values of each long (blue bar), short (orange bar) and non-anomaly (grey bar) stock for every anomaly variable for the full sample period.

[**Insert Figure 4 about here**]

[**Insert Figure 5 about here**]

[**Insert Figure 6 about here**]

Consistent with our initial analysis, we observe significant dispersion among anomaly long and short portfolios compared to that of non-anomaly stocks. Figure 4 plots the anomalies for which the long anomaly stocks were the most crowded. Interestingly, we observe that the main differences are concentrated among the largest stocks. However, we also observe anomalies in which the short leg of the portfolio is more crowded than the long leg, as shown in Figure 5. Similarly, the differences are more significant in the top size deciles. Although significant differences are observed in specific subgroups, it is possible that our estimations are influenced by stocks being included in multiple anomaly strategies. In other words, a stock that is, for example, in the long portfolio of the ACC anomaly might be also included in the short leg of the NSI anomaly, simultaneously. Therefore, to provide a more specific analysis and focus only in the the securities that according to each anomaly variable should be in either the long and short portfolio but are not included in another anomaly long or short portfolio, which we refer to as anomaly only stock.

Figure 7 depicts the time series mean of cross-sectional average days-ADV values of three series: anomaly stocks included exclusively in only one long portfolio of any of our sample of asset pricing anomalies(blue); anomaly stocks included only in one short portfolio (orange); and all stocks that are not included in any long or short anomaly portfolio (grey). As in previous plots, the results are grouped by NYSE size decile. As observed in previous plots, the higher (lower) values of days-ADV appear to be concentrated among the largest(smallest) firms and specially among short-only anomaly stocks. The difference in the value of days-ADV between short-only anomaly stocks, long-only and non-anomaly is statistically significant (t-value of 4.39 an 3.77 respectively) for the top and bottom NYSE deciles. However, for NYSE deciles two through nine non-anomaly stocks appear to be more crowded.

To summarize, our analysis of the differences in values of days-ADV measure for anomaly (long and short) and non-anomaly stocks provides partial evidence that anomaly stocks are more crowded. If we control by firm size we observe that this relation holds for most anomalies only in the group of largest firms. Moreover, this results hold when we focus on stocks that are only in one anomaly portfolio (long or short). However, for some deciles non-anomaly stocks appear to be as crowded as anomaly stocks. These results provide partial evidence to support Hypothesis 2 that anomaly stocks tend to be more crowded. On the other hand, as presented in figure 7 for most size deciles, short-only stocks are more crowded than long-only. We would then expect to observe expected returns to be higher on the short-leg than on the long-leg of our sample of anomalies. We proceed to test this argument in the following section.

5.3 Crowding and anomaly returns

In this section, we test Hypothesis 2 about the interaction between crowding and anomaly returns in the cross-section First, we conditionally sort the stocks in our sample first by days-ADV and then according to each anomaly variables. As a robustness check, we switched the order of the sorting variables to make sure our results are not driven by how the sorting is performed. Next, among stocks in the long and short anomaly portfolios, we focus on those with the highest and lowest days-ADV values. We classify an anomaly stock to be most (least) crowded if it is in the top(bottom) 30% of days-ADV values. Given our interest in measuring the impact of crowding on anomaly returns, we compare our estimations with the performance of single-sorted portfolios of each anomaly variable. Finally, we repeat our analysis for the period before and after the publication date of each anomaly to consider the previously documented alpha decay once anomalies are broadly publicized (Mclean and Pontiff, 2016; Calluzzo et al., 2019)

[Insert Table 6 and 7 about here]

Strikingly, anomaly returns appear to be concentrated among the most and least crowded stocks and this finding is consistent across all the anomalies in our sample. As shown in Table 5 the three-factor alpha of all anomaly diff portfolios (high/long minus low-short) is significantly higher than that obtained in the single sorting portfolio (full sample). The anomalies for which the abnormal returns are higher are CEI, NOA, ROA, and MOM with monthly alphas of 2.07% (t-value)=10.08, 2.39% (t-value=11.66), 2.13% (t-value=9.25), and 2.14% (t-value=6.28) respectively. In line with Mclean and Pontiff (2016) and Calluzzo et al. (2019) most alphas decline in the period after publication, but most of them remain economically and statistically significant. Finally, consistent with Stambaugh et al. (2012) abnormal returns come mostly from the short leg of the anomaly portfolios.

In table 6 we estimate an aggregate anomaly portfolio by taking the equally weighted average each quarter across all available anomaly returns. We observe that our results hold for the aggregate estimation and are robust to different sorting procedures. The fact that abnormal returns are significantly higher (lower) among anomaly stocks within the top (bottom) days-ADV group supports the view that crowded positions include additional risk considerations for arbitrage trading. Our results add to those of Chen et al. (2019) who find that arbitrage trading is not able to correct mispricing in anomalies by showing that crowded equity positions might pose additional limits to arbitrage.

[Insert Table 8 about here]

Finally, as observed in most of the individual anomaly analysis there is decay in the magnitude of the difference (long-short) portfolio alpha when we separate our sample between pre-publication and post-publication periods. We interpret this results as evidence of institutional trading on anomalies (Calluzzo et al. (2019)) but also indicative of limitations to arbitrage trading. We will further explore this explanation in later analysis.

5.4 Crash risk, Liquidity risk, Crowding and Limits to arbitrage

In this section, we test Hypothesis 3 by investigating the channels through which concentrated positions, relative to their liquidity provisions, influence future expected returns as well as its relation to the limits to arbitrage. First, we examine the channels underlying the relation between crowding and future expected returns. We argue that is through an increase in the exposure to crash and liquidity risk that this phenomenon occurs. Finally, we examine the relation between our findings and the limits to arbitrage trading.

5.4.1 Crash risk, Liquidity risk, and crowding

A recent strand of the literature on the cross-section of stock returns shows that investors dislike tail sensitive assets (e.g. Kelly and Jiang (2014); Chabi-Yo et al. (2019)) Therefore, investors expect higher returns for stocks with more negative skewness, implying that skewness is a priced risk factor (Harvey and Siddique (2000)). Crash risk measures aim at capturing the higher moments of the stock return distribution - that is extreme negative returns (Hutton et al. (2009), Callen and Fang (2015)) Although, most of the literature on *crash risk* relates it the information asymmetries between corporate insiders and external stakeholders (Habib et al. (2018)), recent studies analyze this risk in the context of its relation to investor's factor exposure (Chabi-Yo et al. (2019)). We follow this approach and relate crowding to future firm-specific stock price crash risk under the hypothesis that higher crowded holdings increase institutional investors' crash risk exposure.

We begin by estimating three crash risk measures ($NCSKEW, DUVOL, CRASHCOUNT$) for our sample of institutional investors' holdings. Then we proceed to regress them on the log of the days-ADV measure and a set of control variables. The control variables we include are the cummulative firm-specific daily returns, the kurtosis and the standard deviation of firm-specific daily returns, market-to-book ratio, book value of all liabilities divided by total assets, ROA ratio, log of market capitalization (size), average monthly share turnover, the number of analyst following the firm, Amihud (2002) illiquidity measure calculated using daily data, aggregated at the mont level, and estimated as the average over the past 3 months. All control variables, with the exception of the Amihud (2002) illiquidity measure, are measured over the fiscar year t. All regressions control for year and firm fixed-effects [20]. Standard errors are corrected for firm clustering.

[Insert Table 9 about here]

Table 9 reports the results of our regression analysis. Across all three model specifications, different crash risk measures as dependent variables, the estimated coefficient for the $log(ADV_t)$ are significantly positively at less than 5% significance. Moreover, the magnitude of the coefficients increase both in magnitude and significance for the second part of the sample period (1997-2020).This evidence is consistent with Hypothesis 3 and the view that crowded stocks increase institutional exposure to crash risk.

To further examine the channels through which crowding influences future expected returns we explore the effect that crowding exerts on liquidity risk of institutional investors holdings, as well as their illiquidity levels. Following Beber et al. (2012) we include as control variables the log of market capitalization (size), the log of book-to-market ratio, a NASDAQ dummy variable, return and return volatility over the previous month. In addition, we compute t-values from standard errors that are double-clustered by firm and year.

[Insert Table 10 about here]

Table 10 provide the regression results for our model that relates crowding to next-quarter liquidity beta and the Amihud (2002) illiquidity measure. The positive coefficients on $log(ADV_t)$ are significant for both models (t-statistics = 2.23 and 6.35 respectively). This evidence suggests that crowding has predictive power for future stocks liquidity risk and illiquidity.

Overall, our results show a significant both economically and statistically relation between crowding, liquidity risk, crash risk, and illiquidity. This evidence is consistent with Hypothesis 3 and the idea that crowding further increases risk concerns for institutional investors.

6 Conclusion

Intuitively, an increased participation of sophisticated investors will have a positive influence on market efficiency by enhancing arbitrage trading that quickly corrects mispricing. However, they may be negative

[20]We follow Callen and Fang (2015) who argue that the inclusion of the implementation of firm fixed-effects in crash risk regressions help mitigate the concern that ommitted time-invariant firm characteristics may be driving the results.

externalities when too many players participate to the game. Starting with the work of Stein (2009) some recent studies have examined these externalities which are typically refereed as the crowded-trade problem. While there is no doubt that stock markets are increasingly dominated by institutional investors, there is conflicting evidence on the influence of crowding in equity price dynamics and the role that arbitrageurs play in increasing or mitigating this potential problem. Our paper contributes to this current debate by examining crowding in a set of well-known stock anomalies and using a database of institutional investors' holdings. We present several empirical findings that support the view that crowding influences anomaly returns, is positively related to crash risk, and plays a role in the limits of arbitrage by adding risk considerations.

We find that while in aggregate crowdedness has decreased over time in our sample of institutional holdings, specifically crowded equity positions in anomalies remain and have significant impacts in terms of risk and return dynamics. If crowded positions impose additional risk for arbitrageurs, we expect to find increased abnormal returns among the most crowded anomaly stocks. Based on the days-ADV measure over the period 1980-2020 we observe that this is the case across all the anomalies in our sample. Moreover, we find that these anomaly returns remain significant after publication dates. Our support for the limits to arbitrage explanation is relevant to the ongoing debate about the concerns of practitioners and regulators about the risks that highly concentrated positions pose to investors by means of increased exposure to crash risk.

References

Agarwal, Vikas, Wei Jiang, Yuehua Tang and Baozhong Yang (2013), 'Uncovering hedge fund skill from the portfolio holdings they hide', *Journal of Finance* **68**(2), 739–783.

Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu and Avanidhar Subrahmanyam (2015), 'Smart money, dumb money, and capital market anomalies', *Journal of Financial Economics* **118**(2), 355–382.

Amihud, Yakov (2002), 'Illiquidity and stock returns: Cross-section and time-series effects', *Journal of Financial Markets* **5**(1), 31–56.

Antón, Miguel and Christopher Polk (2014), 'Connected Stocks', *Journal of Finance* **69**(3), 1099–1127.

Aragon, George O. (2007), 'Share restrictions and asset pricing: Evidence from the hedge fund industry', *Journal of Financial Economics* **83**(1), 33–58.

Aragon, George O. and Philip E. Strahan (2012), 'Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy', *Journal of Financial Economics* **103**(3), 570–587.

Baltas, Nick (2019), 'The Impact of Crowding in Alternative Risk Premia Investing', *Financial Analysts Journal* **75**(3), 89–104.

Barroso, Pedro and Pedro Santa-Clara (2015), 'Momentum has its moments', *Journal of Financial Economics* **116**(1), 111–120.

Beber, Alessandro, Michael Brandt, Mathijs Cosemans and Michaela Verardo (2012), 'Ownership Crowded with Style', *Netspar Working Paper* pp. 1–63.

Ben-David, Itzhak, Francesco Franzoni and Rabih Moussawi (2018), 'Do ETFs Increase Volatility?', *Journal of Finance* **73**(6), 2471–2535.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi and John Sedunov (2021), 'The Granular Nature of Large Institutional Investors', *Management Science* (August), mnsc.2020.3808.

Blocher, Jesse (2016), 'Network externalities in mutual funds', *Journal of Financial Markets* **30**, 1–26.

Brown, Gregory W., Philip Howard and Christian T. Lundblad (2019), 'Crowded Trades and Tail Risk', *SSRN Electronic Journal* .

Brunnermeier, Markus K. and Lasse Heje Pedersen (2009), 'Market liquidity and funding liquidity', *Review of Financial Studies* **22**(6), 2201–2238.

Callen, Jeffrey L. and Xiaohua Fang (2015), 'Short interest and stock price crash risk', *Journal of Banking and Finance* **60**, 181–194.

Calluzzo, Paul, Fabio Moneta and Selim Topaloglu (2019), 'When anomalies are publicized broadly, do institutions trade accordingly?', *Management Science* **65**(10), 4555–4574.

Cao, Charles, Bing Liang, Andrew W. Lo and Lubomir Petrasek (2018), 'Hedge fund holdings and stock market efficiency', *Review of Asset Pricing Studies* **8**(1), 77–116.

Cao, Charles, Yong Chen, Bing Liang and Andrew W. Lo (2013), 'Can hedge funds time market liquidity?', *Journal of Financial Economics* **109**(2), 493–516.

Carhart, Mark M. (1997), 'On persistence in mutual fund performance', *Journal of Finance* **52**(1), 57–82.

Cella, Cristina, Andrew Ellul and Mariassunta Giannetti (2013), 'Investors' horizons and the amplification of market shocks', *Review of Financial Studies* **26**(7), 1607–1648.

Chabi-Yo, Fousseni, Markus Huggenberger and Florian Weigert (2019), 'Multivariate Crash Risk', *SSRN Electronic Journal* .

Chan, Hector and Tony Tan (2020), 'Short Interest, Crowding and Liquidity Crises', *SSRN Electronic Journal* (March), 1–41.

Chang, Xin, Yangyang Chen and Leon Zolotoy (2017), *Stock Liquidity and Stock Price Crash Risk*, Vol. 52.

Chen, Qinhua (2019), 'Arbitrage Crowdedness and Portfolio Momentum', *SSRN Electronic Journal* .

Chen, Yong, Zhi Da and Dayong Huang (2019), 'Arbitrage trading: The long and the short of it', *Review of Financial Studies* **32**(4), 1608–1646.

Chernenko, Sergey and Adi Sunderam (2020), 'Do fire sales create externalities?', *Journal of Financial Economics* **135**(3), 602–628.

Chincarini, Ludwig B. (2012), *The Crisis Of Crowding*.

Chincarini, Ludwig B. (2018), 'Transaction costs and crowding', *Quantitative Finance* **18**(8), 1389–1410.

Cooper, Michael J., Roberto C. Gutierrez and Allaudeen Hameed (2004), 'Market states and momentum', *Journal of Finance* **59**(3), 1345–1365.

Coval, Joshua and Erik Stafford (2007), 'Asset fire sales (and purchases) in equity markets', *Journal of Financial Economics* **86**(2), 479–512.

Cull, Robert and Lixin Colin Xu (2005), 'Institutions, ownership, and finance: the determinants of profit reinvestment among chinese firms', *Journal of Financial Economics* **77**(1), 117–146.

Daniel, Kent and Sheridan Titman (2006), 'Market reactions to tangible and intangible information', *Journal of Finance* **61**(4), 1605–1643.

Daniel, Kent and Tobias J. Moskowitz (2016), 'Momentum crashes', *Journal of Financial Economics* **122**(2), 221–247.

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**(4), 703–738.

Dichev, Ilia D. (1998), 'Is the risk of bankruptcy a systematic risk?', *Journal of Finance* **53**(3), 1131–1147.

Dimson, Elroy (1979), 'Risk measurement when shares are subject to infrequent trading', *Journal of Financial Economics* **7**(2), 197–226.

Edelen, Roger M., Ozgur S. Ince and Gregory B. Kadlec (2016), 'Institutional investors and stock return anomalies', *Journal of Financial Economics* **119**(3), 472–488.

Edmans, Alex, Itay Goldstein and Wei Jiang (2015), 'Feedback Effects, Asymmetric Trading, and the Limits to Arbitrage', *American Economic Review* **105**(12), 3766–3797.

Fama, Eugene F. (1998), 'Market efficiency, long-term returns, and behavioral finance', *Journal of Financial Economics* **49**(3), 283–306.

Fama, Eugene F. and James D. MacBeth (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**(3), 607–636.

Fama, Eugene F. and Kenneth R. French (1992), 'The Cross-Section of Expected Stock Returns', *The Journal of Finance* **47**(2), 427–465.

Fama, Eugene F. and Kenneth R. French (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, Eugene F. and Kenneth R. French (2006), 'Profitability, investment and average returns', *Journal of Financial Economics* **82**(3), 491–518.

Fama, Eugene F. and Kenneth R. French (2008), 'Dissecting anomalies', *Journal of Finance* **63**(4), 1653–1678.

French, Kenneth R. (2008), 'Presidential address: The cost of active investing', *Journal of Finance* **63**(4), 1537–1573.

Girardi, Giulio, Kathleen Weiss Hanley, Stanislava Nikolova, Loriana Pelizzon and Mila Getmansky Sherman (2021), 'Portfolio similarity and asset liquidation in the insurance industry', *Journal of Financial Economics* .

Grossman, Sanford J. and Joseph E. Stiglitz (1980), 'On the Impossibility of Informationally Efficient Markets'.

Habib, Ahsan, Mostafa Monzur Hasan and Haiyan Jiang (2018), 'Stock price crash risk: review of the empirical literature', *Accounting and finance (Parkville)* **58**(S1), 211–251.

Han, Yufeng, Dayong Huang and Guofu Zhou (2021), 'Anomalies enhanced: A portfolio rebalancing approach', *Financial Management* **50**(2), 371–424.

Harvey, Campbell R. and Akhtar Siddique (2000), 'Conditional skewness in asset pricing tests', *Journal of Finance* **55**(3), 1263–1295.

Hau, Harald and Sandy Lai (2017), 'The role of equity funds in the financial crisis propagation', *Review of Finance* **21**(1), 77–108.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh and Yinglei Zhang (2004), 'Do investors overvalue firms with bloated balance sheets?', *Journal of Accounting and Economics* **38**(1-3 SPEC. ISS.), 297–331.

Hirshleifer, David and Siew Hong Teoh (2003), 'Herd behaviour and cascading in capital markets: A review and synthesis', *European Financial Management* **9**(1), 25–66.

Hombert, Johan and David Thesmar (2014), 'Overcoming limits of arbitrage: Theory and evidence', *Journal of Financial Economics* **111**(1), 26–44.

Hong, Harrison, Weikai Li, Sophie Ni, Jose Scheinkman and Philip Yan (2015), Days to Cover and Stock Returns, Technical report, National Bureau of Economic Research, Cambridge, MA.

Hutton, Amy P., Alan J. Marcus and Hassan Tehranian (2009), 'Opaque financial reports, R2, and crash risk', *Journal of Financial Economics* **94**(1), 67–86.

Jegadeesh, Narasimhan and Sheridan Titman (1993), 'Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency', *The Journal of Finance* **48**(1), 65–91.

Kelly, Bryan and Hao Jiang (2014), 'Tail risk and asset prices', *Review of Financial Studies* **27**(10), 2841–2871.

Khandani, Amir E. and Andrew W. Lo (2011), 'What happened to the quants in August 2007? Evidence from factors and transactions data', *Journal of Financial Markets* **14**(1), 1–46.

Kokkonen, Joni and Matti Suominen (2015), 'Hedge funds and stock market efficiency', *Management Science* **61**(12), 2890–2904.

Lam, F. Y.Eric C. and K. C.John Wei (2011), 'Limits-to-arbitrage, investment frictions, and the asset growth anomaly', *Journal of Financial Economics* **102**(1), 127–149.

Lou, Dong and Christopher Polk (2020), 'Comomentum: Inferring arbitrage activity from return correlations', *Working Paper* pp. 1–33.

Loughran, Tim and Jay R. Ritter (1995), 'The New Issues Puzzle', *The Journal of Finance* **50**(1), 23–51.

Marks, Joseph M. and Chenguang Shang (2019), 'Factor Crowding and Liquidity Exhaustion', *Journal of Financial Research* **42**(1), 147–180.

Mclean, R. David and Jeffrey Pontiff (2016), 'Does Academic Research Destroy Stock Return Predictability?', *Journal of Finance* **71**(1), 5–32.

Nofsinger, John R. and Richard W. Sias (1999), 'Herding and Feedback Trading by Institutional and Individual Investors', *The Journal of Finance* **54**(6), 2263–2295.

Novy-Marx, Robert (2013), 'The other side of value: The gross profitability premium', *Journal of Financial Economics* **108**(1), 1–28.

Pastor, Lubus and Robert F. Stambaugh (2003), 'Liquidity Risk and Expected Stock Returns ˘ ubos ˘ Pa ´ stor Robert F', *Journal of Political Economy* **111**(3), 642–685.

Phillips, Blake, Kuntara Pukthuanthong and P. Raghavendra Rau (2014), 'Detecting superior mutual fund managers: Evidence from copycats', *Review of Asset Pricing Studies* **4**(2), 286–321.

Pontiff, Jeffrey (2006), 'Costly arbitrage and the myth of idiosyncratic risk', *Journal of Accounting and*

Economics **42**(1-2), 35–52.

Ruenzi, Stefan and Florian Weigert (2018), 'Momentum and crash sensitivity', *Economics Letters* **165**, 77–81.

Shleifer, Andrei and Robert W. Vishny (1997), 'The Limits of Arbitrage', *The Journal of Finance* **52**(1), 35–55.

Sias, Richard, H. J. Turtle and Blerina Zykaj (2016), 'Hedge fund crowds and mispricing', *Management Science* **62**(3), 764–784.

Sloan, Richard G. (1996), 'Do stock prices fully reflect information in accruals and cash flows about future earnings?', *Accounting Review* **71**(3), 289–315.

Stambaugh, Robert F., Jianfeng Yu and Yu Yuan (2012), 'The short of it: Investor sentiment and anomalies', *Journal of Financial Economics* **104**(2), 288–302.

Stambaugh, Robert F., Jianfeng Yu and Yu Yuan (2015), 'Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle', *Journal of Finance* **70**(5), 1903–1948.

Stambaugh, Robert F. and Yu Yuan (2017), 'Mispricing factors', **30**(4), 1270–1315.

Stein, Jeremy C. (2009), 'Presidential Address: Sophisticated investors and market efficiency', *Journal of Finance* **64**(4), 1517–1548.

Stulz, René M (2007), 'Hedge Funds: Past, Present, and Future', *Journal of Economic Perspectives* **21**(2), 175–194.

Titman, Sheridan, K. C. John Wei and Feixue Xie (2004), 'Capital Investments and Stock Returns', *Journal of Financial and Quantitative Analysis* **39**(4), 677–700.

Verbeek, Marno and Yu Wang (2013), 'Better than the original? The relative success of copycat funds', *Journal of Banking and Finance* **37**(9), 3454–3471.

Wermers, Russ (1999), 'Mutual fund herding and the impact on stock prices', *Journal of Finance* **54**(2), 581–622.

Xing, Yuhang (2008), 'Interpreting the value effect through the Q-theory: An empirical investigation', *Review of Financial Studies* **21**(4), 1767–1795.

Yan, Philip (2014), 'Crowded Trades, Short Covering, and Momentum Crashes', *SSRN Electronic Journal* .

Yan, Xuemin and Zhe Zhang (2009), 'Institutional investors and equity returns: Are short-term institutions better informed', *Review of Financial Studies* **22**(2), 893–924.

Zhong, Ligang, Xiaoya Sara Ding and Nicholas S.P. Tay (2017), 'The impact on stock returns of crowding by mutual funds', *Journal of Portfolio Management* **43**(4), 87–99.

Appendix A. Figures and Tables

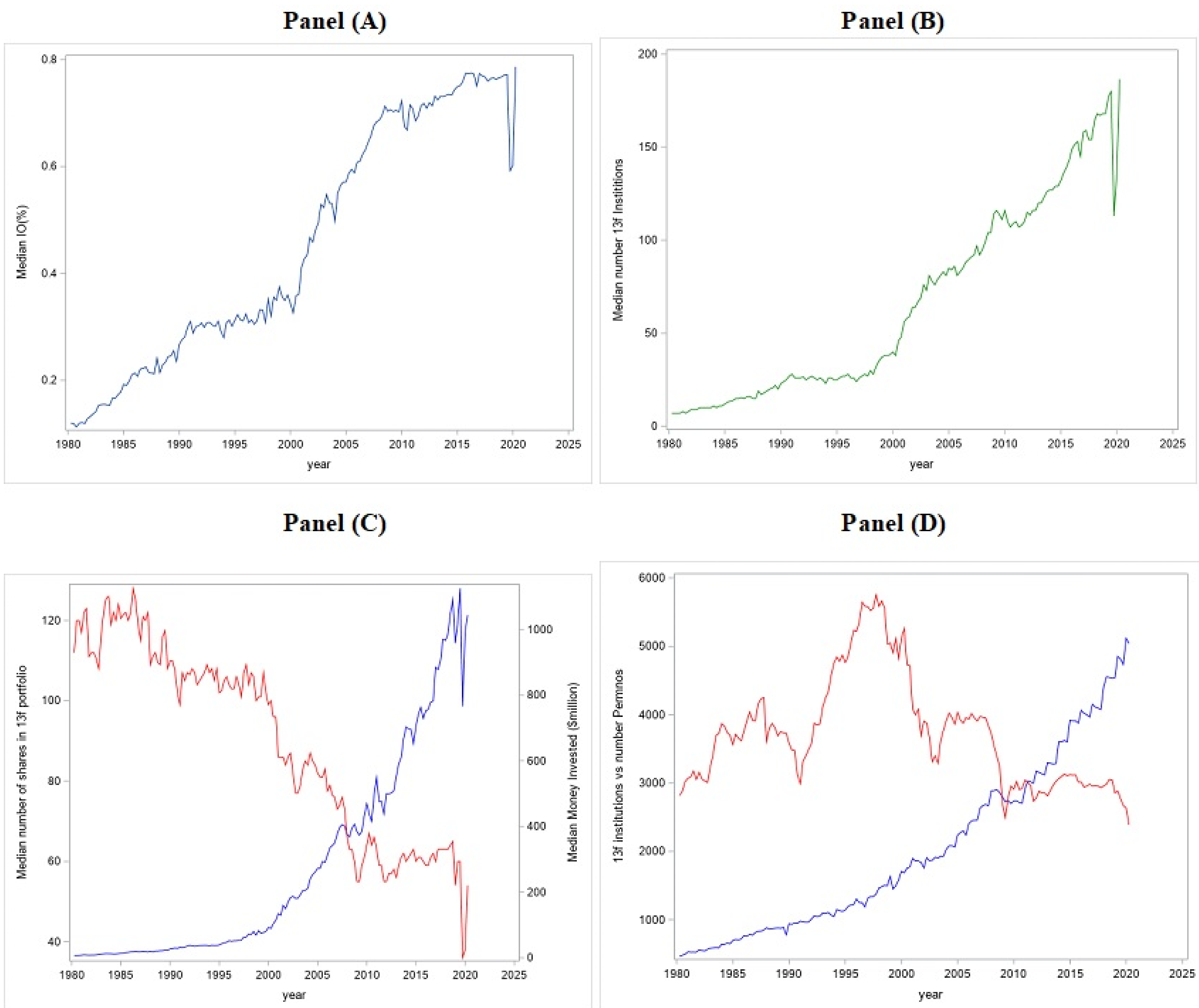


Figure 1: **13F Institutional Investors position in average security and holdings descriptive statistics**. **Panel (a)** shows the growth of the median Institutional Ownership (IO) in percentage terms. IO is estimated for each security as the number of shares held by institutional investors divided by the total number of shares outstanding. **Panel (b)** illustrates the growth in the median number of institutional investors ($NumbInst$) holding the same security. **Panel (c)** shows, in the red line, the median number of shares in a typical portfolio of an institutional investor in our sample. This graph also shows, in the blue line, the growth in the median amount of money invested, expressed in millions of USD, by an institutional investor in a typical security. **Panel (d)** illustrates, in the red line, the total number of distinct securities existing in our 13F institutional investors' holdings dataset in each quarter. Additionally, in the blue line, we show the total number of distinct 13F institutional investors in our sample.

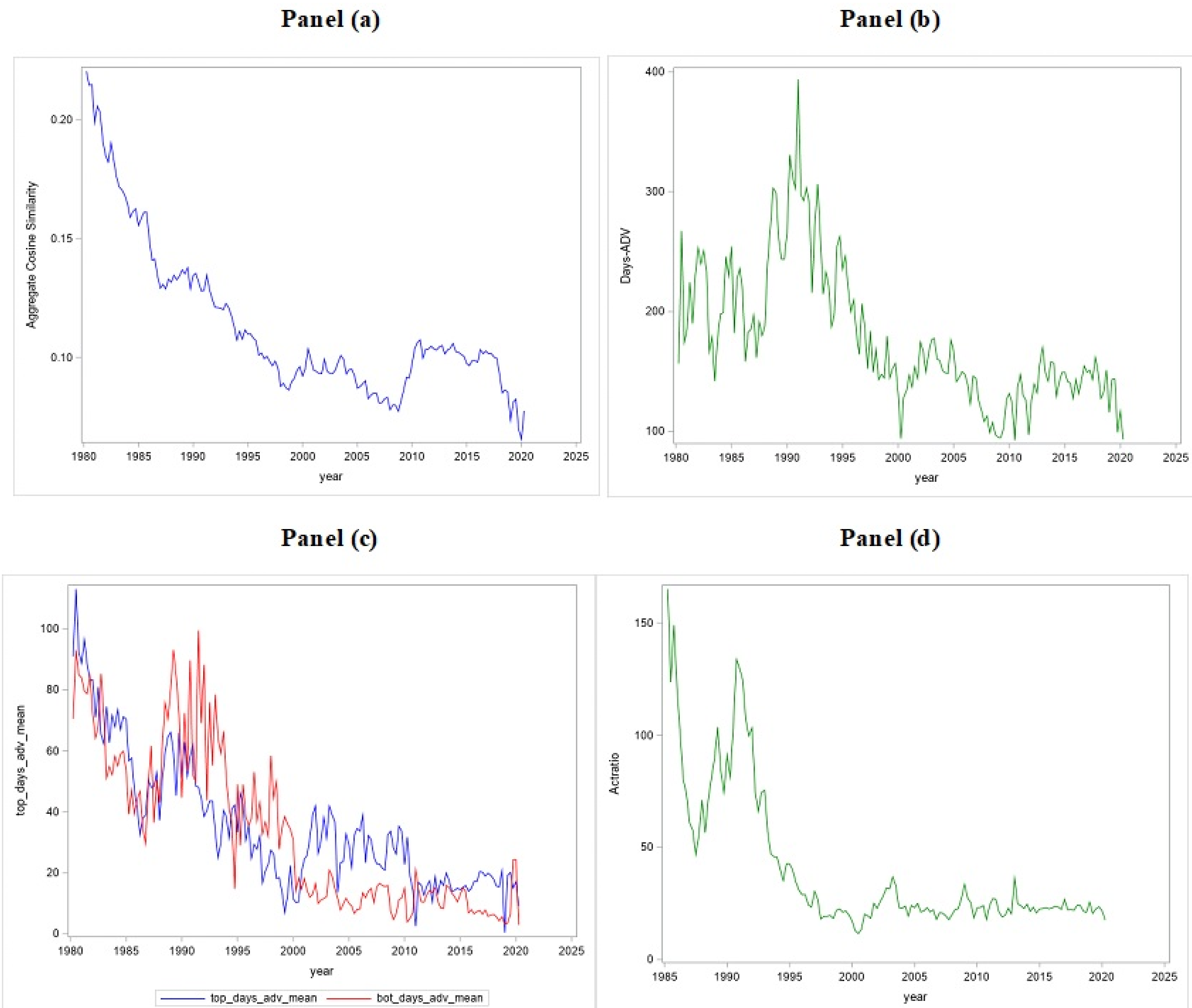
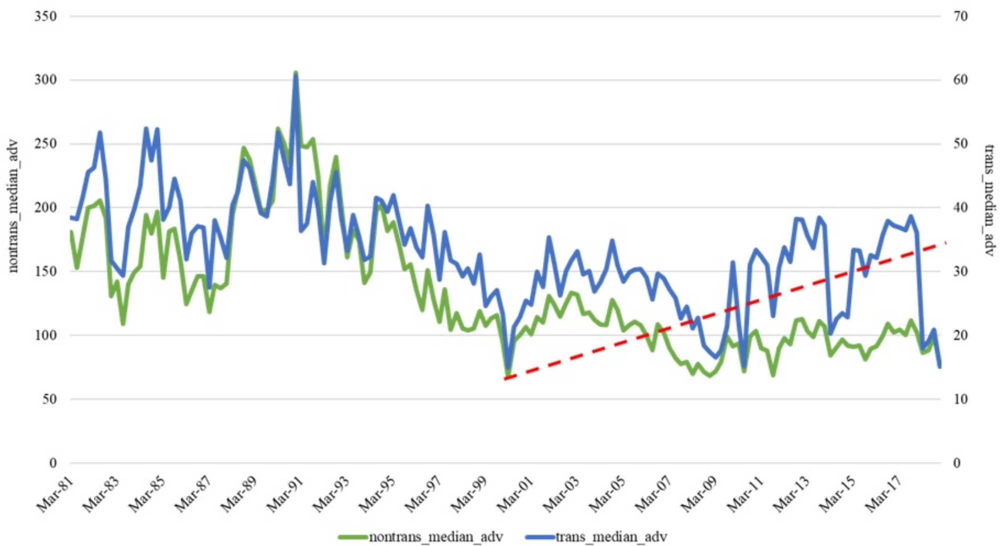


Figure 2: **Crowding measures of 13F holdings database 1980Q1-2020Q1: Time-series means of cross-sectional averages**. **Panel (a)** illustrates the aggregate cosine similarity measure for our sample of 13F institutional investors' holdings. The cosine similarity allows us to assess how similar are portfolios in terms of overlapping holdings. We estimate it as the dot product between the position weight vector of each portfolio divided by the Euclidian norm of each vector. This measure is then aggregated each quarter and adjusted by the number of distinct institutional investors in that same quarter. **Panel (b)** shows the time series plot of the days-ADV measure for a typical security in our sample. Days-ADV is measured as the money value held in a security by all institutional investors relative to the security's average daily money volume. **Panel (c)** illustrates in the blue (red) line the time series mean days-ADV estimated for the top (bottom) 5% percentile of funds that showed the highest (lowest) cosine similarity. **Panel (d)** illustrates the time series plot of the Activity Ratio (Actratio) measure. We estimated Actratio as the percentage of shares held by an institution at the end of the quarter $(t-2)$ divided by the security's average turnover during the quarter $(t-1)$.

Panel (a)



Panel (b)

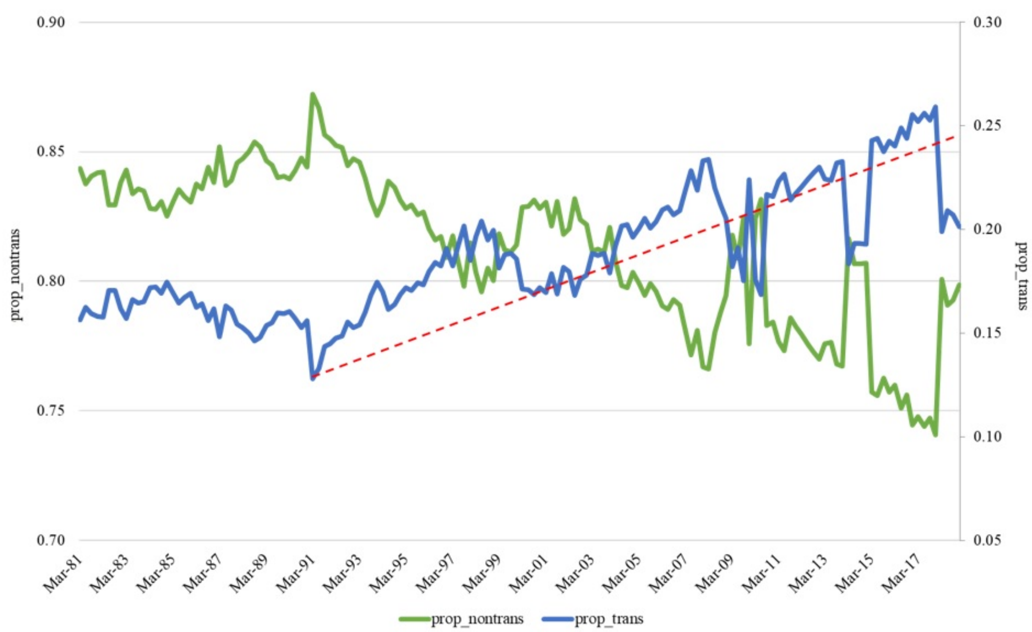


Figure 3: **Crowding measures of 13F holdings database 1980Q1-2020Q1: Time-series means of cross-sectional averages**. Days-ADV for transient and non-transient institutions. **Panel (a)** illustrates in the blue (green) line the time series mean days-ADV estimated for transient (non-transient) 13F institutions. **Panel (b)** shows in the green (blue) line the time series percentage of the total days-ADV measure for the 13F database that comes from transient (non-transient) institutions.

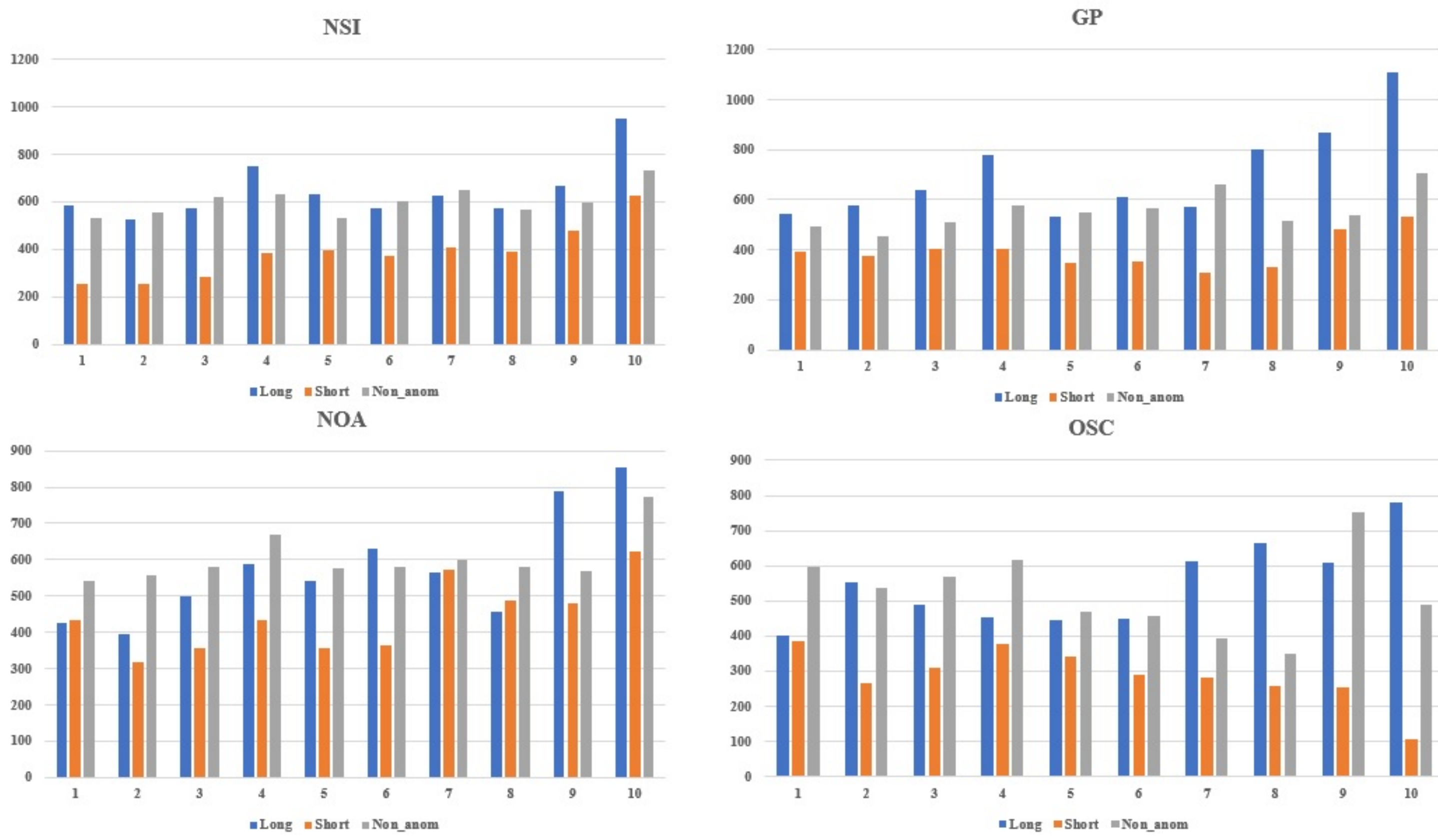


Figure 4: **Days-ADV of Long, Short, and Non-anomaly stocks by NYSE decile breakpoints (Continued).** This figure shows the time-series mean of cross-sectional average of the measure days-ADV for the anomaly stocks included in the long (blue bar), short (orange bar), and non-anomaly (grey bar) quartiles of each stock anomaly. The non-anomaly bar aggregates the stocks in the quartiles 2 through 4, while the long bar includes those stocks with the highest days-ADV values (quantile 5) and the short bar those with the lowest values (quantile 1). The sample period is from 1980:Q1 to 2020:Q1

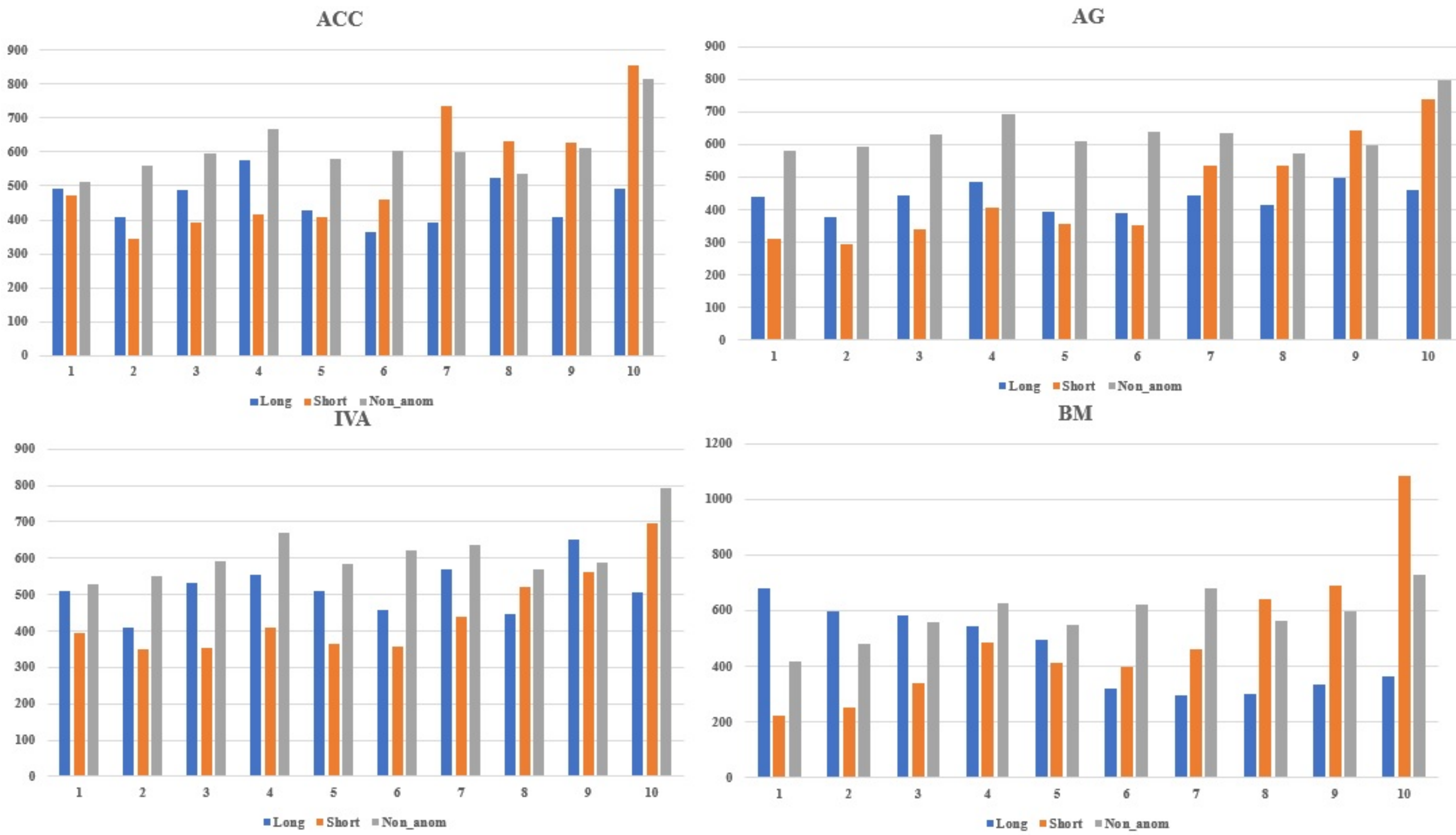


Figure 5: **Days-ADV of Long, Short, and Non-anomaly stocks by NYSE decile breakpoints (Continued).** This figure shows the time-series mean of cross-sectional average of the measure days-ADV for the anomaly stocks included in the long (blue bar), short (orange bar), and non-anomaly (grey bar) quartiles of each stock anomaly. The non-anomaly bar aggregates the stocks in the quartiles 2 through 4, while the long bar includes those stocks with the highest days-ADV values (quantile 5) and the short bar those with the lowest values (quantile 1). The sample period is from 1980:Q1 to 2020:Q1

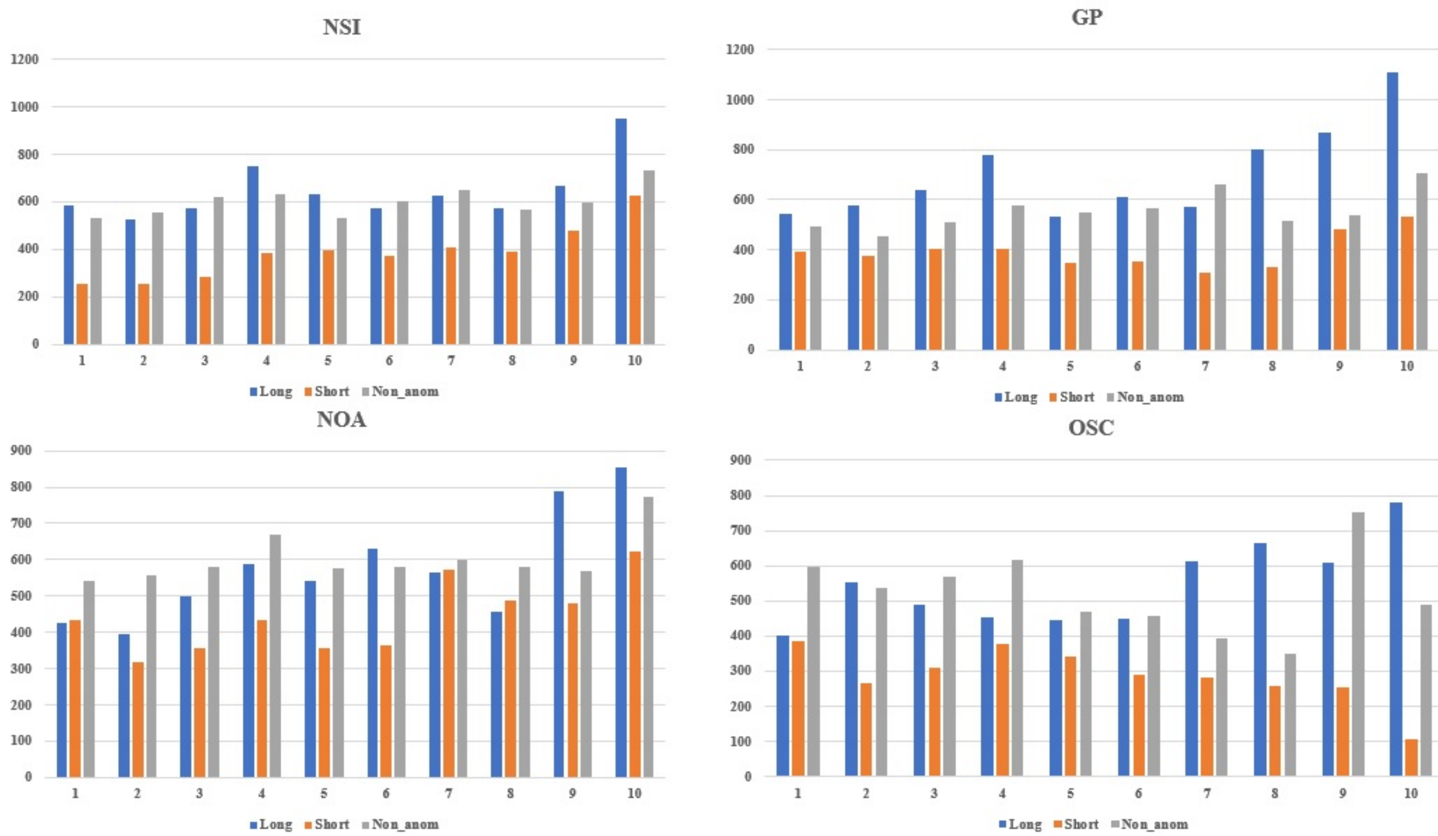


Figure 6: **Days-ADV of Long, Short, and Non-anomaly stocks by NYSE decile breakpoints.** This figure shows the time-series mean of cross-sectional average of the measure days-ADV for the anomaly stocks included in the long (blue bar), short (orange bar), and non-anomaly (grey bar) quartiles of each stock anomaly. The non-anomaly bar aggregates the stocks in the quartiles 2 through 4, while the long bar includes those stocks with the highest days-ADV values (quantile 5) and the short bar those with the lowest values (quantile 1). The sample period is from 1980:Q1 to 2020:Q1

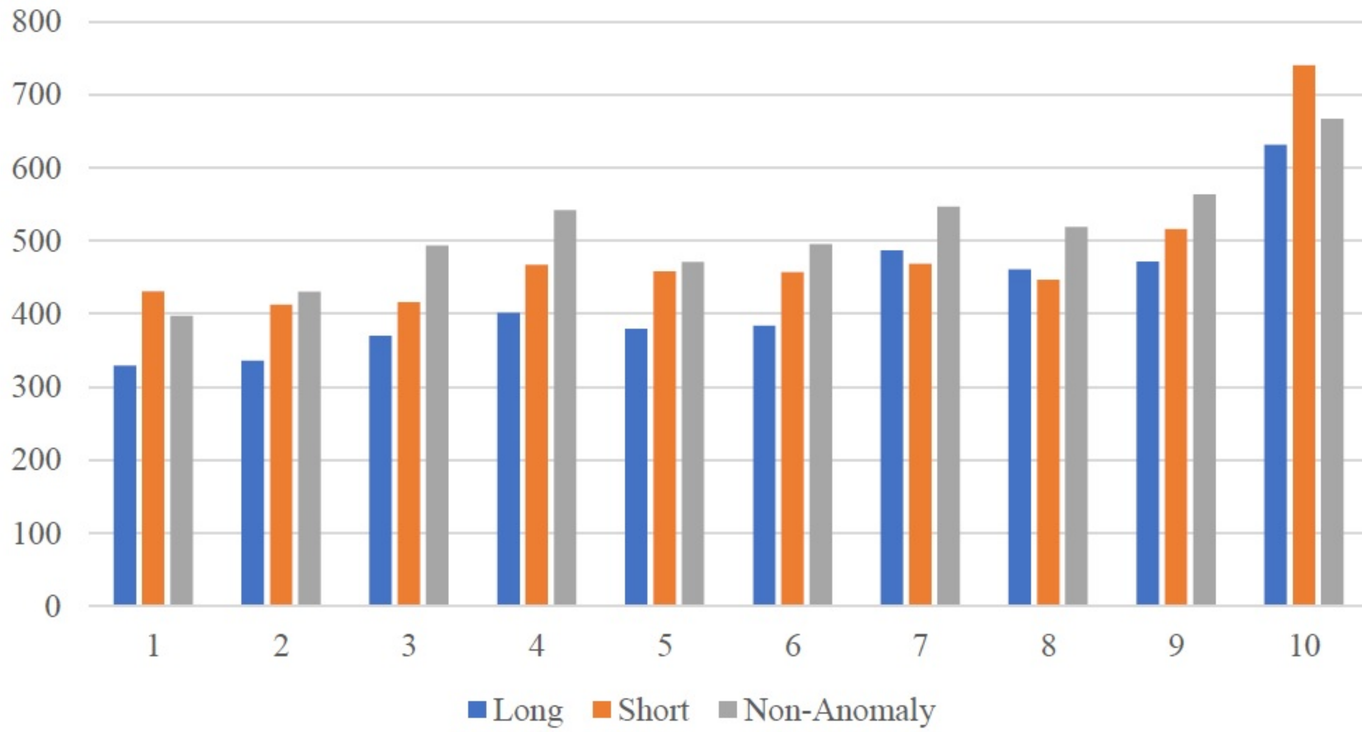


Figure 7: **Days-ADV of Long, Short, and Non-anomaly only stocks by NYSE decile breakpoints.** This figure shows the time-series mean of cross-sectional average of the measure days-ADV for three series. First, the anomaly stocks that are included exclusively in only one **long** portfolio of any of the twelve anomalies (blue bar). Second, the orange bar shows anomaly stocks that are included exclusively in only one **short** portfolio of any of the twelve anomalies. Finally, the third series groups all **non-anomaly** stocks (gray bar) quartiles of each stock anomaly. The non-anomaly bar aggregates the stocks in the quartiles 2 through 4. The sample period is from 1980:Q1 to 2020:Q1

Table 1: **Sample Anomalies**

	Anomaly	**Label**	**Paper**	**Description**	**SSRN year**
1	Composite equity issuance	CEI	Daniel and Titman (2006)	CEI measures the amount of equity a firm issue or retires in exchange for cash or services. Firms with higher CEI earn lower risk-adjusted returns	2001
2	Net stock issuance	NSI	Loughran and Ritter (1995)	Issuing firms underperform compared to the overall market and such performance lasts for up to three years.	
3	Total accruals	ACC	Sloan (1996)	Stock prices may not reflect the accrual component of earnings. Firms with higher total accounting accruals underperform those with lower accounting accruals	
4	Net operating assets	NOA	Hirshleifer et al. (2004)	NOA is negatively related to firm's future long-run risk-adjusted return.	2003
5	Gross profitability	GP	Novy-Marx (2013)	Profitable firms earn significantly higher risk-adjusted returns than unprofitable ones	2010
6	Asset growth	AG	Cooper et al. (2004)	Firms with higher asset growth rates subsequently underperform those with lower growth rates.	2005
7	Capital investments	CI	Titman et al. (2004)	Increases in firm's capital investments strongly predicts future lower risk adjusted returns.	2001
8	Investment-to-assets	IVA	Xing (2008)	Firms with low investment-to-assets ratios show higher risk-adjusted returns compared to those with higher ratios	2008
9	Momentum	MOM	Jegadeesh and Titman (1993)	A profitable strategy is to buy shares of firms with positive performance in the past six months, skip one month, and hold it for the following six months.	2001
10	Ohlson O-score	OSC	Dichev (1998)	Higher bankruptcy risk, measured by the O-score Ohlson (1980), is not rewarded with higher returns. Firms facing increased bankruptcy risk earn subsequently lower returns.	2001
11	Return to assets	ROA	Fama and French (2006)	Profitable firms, measured by their ROA, earn higher risk-adjusted returns compared to those with lower ROA.	2001
12	Book-to-market	BM	Fama and French (1992)	The value premium: value stocks earn higher expected return than growth stocks	

Table 2: **Summary Statistics**

Panel A: time-series statistics of cross-sectional averages

	Mean	Median	St. Dev.	Maximum	Minimum
NumbInst	112.9	66.5	148.2	1,275.4	1.0
IO (%)	44.3	45.7	25.1	1.0	0.0
INVST (USD $millions)	2,024.4	271.2	7,612.1	170,138.3	0.1
NSTK	201.2	89.6	344.6	3,135.6	1.0
IOT(%)	0.7	0.2	2.5	49.5	0.0
Days-ADV	414.8	177.3	714.7	5,919.9	0.6
ACTR	870.9	114.5	1,803.1	14,450.2	0.1

Panel B: Correlations

	(1) NumbInst	(2) IO	(3) INVST	(4) Days-ADV
(1) NumbInst	1.00			
(2) IO	0.45	1.00		
(3) INVST	0.78	0.20	1.00	
(4) Days-ADV	0.01	0.08	0.04	1.00

This table reports descriptive statistics of crowding measures. The sample period is from 1980:Q1 to 2020:Q1. The data on institutional holdings is obtained from Thomson Reuters (TR) 13F database. Stock price, trading volume, and total shares outstanding data is from CRSP. Number of institutions is a counter of the number of distinct institutional investors holding the same stock. Institutional ownership is estimated for each stock as the number of shares held by institutional investors divided by the total number of shares outstanding. Total value invested is the money value of institutional ownership (Institutional ownership * end-of-quarter price). Number of stocks in a portfolio is a counter of the number of distinct stocks (permno) in a typical institutional investor portfolio. Institutional ownerslohip in a typical share is the institutional ownership (number of shares held by an institution divided by the total number of shares outstanding) in a typical share (permno). Days-ADV is the money value held in a security by all institutional investors relative to the security's average daily money volume. Activity ratio is the percentage of shares held by an institution at the end of the quarter $(t-2)$ divided by the security's average turnover during the quarter $(t-1)$. Days-ADV, activity ratio, and total value invested were winsorized at the 1% level. We include only stocks whose CRSP share code is 10 and 11 (ordinary common shares). Also, we exclude firms with stock prices less than USD $5 to reduce the effects of microcaps. **Panel A** presents the time series descriptive statistics of cross-sectional averages of crowding-related measures and proxies. **Panel B** shows the time-series averages of the cross-sectional correlations between selected crowding measures.

Table 3: **Crowding and stock returns**

Panel A: Crowding Portfolios three-factor alphas

	Equal-weighted portfolios			Value-weighted portfolios		
	High	Low	Diff	High	Low	Diff
IO	-0.12	-0.12	0.00	-0.08	-0.14	0.06
	(-1.82)	(-1.55)	(0.02)	(-1.58)	(-1.35)	(0.44)
NInst	0.01	0.01	-0.00	0.03	-0.13	0.16
	(0.12)	(0.16)	(-0.08)	(3.31)	(-1.50)	(1.75)
Days-ADV	0.55	-0.97	1.52	0.62	-0.96	1.58
	(8.64)	(-10.49)	(11.67)	(8.86)	(-8.06)	(9.52)
AcRatio	0.33	-0.77	1.09	0.20	-0.66	0.87
	(5.21)	(-9.25)	(8.88)	(3.97)	(-5.09)	(5.15)

Panel B: Value-weighted crowding portfolios alpha over different rebalancing frequencies

	Long		Short		Diff	
	Q_{t+4}	Q_{t+8}	Q_{t+4}	Q_{t+8}	Q_{t+4}	Q_{t+8}
IO	-0.10	-0.09	-0.09	-0.08	-0.01	-0.02
	(-1.97)	(-1.82)	(-0.95)	(-0.90)	(-0.09)	(-0.16)
NInst	0.01	0.01	-0.04	-0.06	0.06	0.08
	(1.76)	(1.56)	(-0.51)	(-0.76)	(0.65)	(0.90)
Days-ADV	0.64	0.57	-0.92	-0.83	1.56	1.39
	(9.67)	(9.27)	(8.75)	(-8.66)	(9.42)	(9.26)
AcRatio	0.18	0.15	-0.70	-0.66	0.89	0.81
	(3.84)	(3.33)	(-5.69)	(5.59)	(5.57)	(5.32)

This table reports monthly portfolio performance (expressed in percentage) measured by the three-factor alpha for the high, low, and high-minus-low (diff) portfolios sorted on several crowding measures for our full sample of 13F institutional investors. The alpha is the intercept of regression of quarterly portfolio excess-return on the Fama-French three-factor model. Number of institutions (NInst) is a counter of the number of distinct institutional investors holding the same stock. Institutional ownership (IO) is estimated for each stock as the number of shares held by institutional investors divided by the total number of shares outstanding. Days-ADV is the money value held in security by all institutional investors relative to the security's average daily money volume. Activity ratio is the percentage of shares held by an institution at the end of each quarter (*t-2*) divided by the stock's average turnover during the quarter (*t-1*). The alpha is the intercept of a regression of monthly portfolio returns on the three Fama-French factors. We include only stocks whose CRSP share code is 10 and 11 (ordinary common shares). Also, we exclude firms with stock prices less than USD $5 to reduce the effects of microcaps. **Panel A** reports the performance of both equal and value-weighted portfolios. In parentheses, we report the t-stat of the hypothesis test that alpha is equal to 0. **Panel B** shows value-weighted portfolio performance, as measured by the three-factor alpha, for portfolios rebalanced every four quarters (Q_{t+4}) and eight quarters (Q_{t+8}) of high, low, and high-minus-low (diff) portfolios. In parentheses, we report the t-stat based on Newey-West standard errors..

Table 4: **days-ADV and stock returns of institutional investors' holdings**

Panel A: Quintile portfolios formed on days-ADV

	Excess return and Alpha					
	Exc Ret	FF3	FF4	FF5	FF3+liq	MISP
Quintile 5 -High	1.18	0.62	0.55	0.49	0.65	0.34
	(6.31)	(8.86)	(8.63)	(7.91)	(9.73)	(4.81)
Quintile 4	0.70	0.12	0.03	-0.02	0.09	-0.10
	(3.82)	(2.13)	(0.83)	(-0.51)	(2.17)	(-1.76)
Quintile 3	0.55	-0.08	-0.11	-0.15	-0.10	-0.15
	(2.77)	(-1.77)	(-2.61)	(-3.61)	(-2.57)	(2.74)
Quintile 2	0.37	-0.31	-0.29	-0.14	-0.38	-0.07
	(1.55)	(-4.07)	(-4.30)	(-2.41)	(-5.41)	(-0.07)
Quintile 1 -Low	-0.14	-0.96	-0.76	-0.61	-0.94	-0.37
	(-0.46)	(-8.06)	(-7.62)	(-6.34)	(-8.99)	(-2.91)
High-minus-Low(HML)	1.32	1.58	1.31	1.09	1.59	0.71
	(6.24)	(9.52)	(9.38)	(8.25)	(9.63)	(4.19)

Panel B: Performance of quintile portfolios for subperiods

	1980Q1-1996Q1			1996Q2-2020Q1		
	FF3	FF3+liq	FF4	FF3	FF3+liq	FF4
Quintile 5 -High	0.38	0.37	0.34	0.81	0.81	0.74
	(7.84)	(7.54)	(6.85)	(8.04)	(7.90)	(7.62)
Quintile 1 -Low	-0.87	-0.84	-0.85	-0.92	-0.94	-0.79
	(-6.85)	(-6.54)	(-6.35)	(-6.76)	(-6.86)	(-6.30)
High-minus-Low(HML)	1.26	1.21	1.19	1.74	1.75	1.53
	(7.86)	(7.54)	(7.15)	(8.55)	(8.25)	(8.29)

This table reports excess and risk-adjusted return for quintiles portfolios and a portfolio (HML) that buys the quintile 5 (high) and sells the quintile 1 (low) of stocks included in the 13F database. At the end of each quarter, we form quintile portfolios based on days-ADV and track their monthly excess returns as the value-weighted of excess returns on all the stocks in each quintile portfolio. The sample period is from 1980:Q1 to 2020:Q1. The data on institutional holdings is obtained from Thomson Reuters (TR) 13F database. In **Panel A** we estimate the risk-adjusted return of each portfolio as the intercept, alpha, of the monthly excess returns on several risk factors: the three-factor of Fama (1998) – FF3, the four-factors of Fama (1998) that includes the momentum factor of Carhart (1997) – FF4, the Fama and French (2006) five-factor model, the Fama and French (1998) model augmented with the Pastor and Stambaugh (2003) traded liquidity factor – FF3 + Liq, and the Stambaugh and Yuan (2017) model that combines two mispricing factors with the market and size factors – MISP. Returns and alphas are in percent per month. The sample period is from 1980:Q1 to 2020:Q1 except for the MISP model that is estimated until 2016Q4, since we only have data on the mispricing factors until that date. The t-values are in parentheses. **Panel B** presents the result from repeating the risk-adjustment process for two subsamples 1980:Q1 – 1996:Q1 and 1996:Q2 – 2020:Q1. According to the test for a structural break in the time series of the days-ADV measure, the series showed two distinct behaviors before and after the year 1996.

Table 5: **Fama-MacBeth regressions: Crowding and future expected returns**

Panel A: Return in the next quarter $(t+3)$

	$CumRet_{t,t+3}$		$ExcessRet_{t,t+3}$	
	1980-1996	1997-2020	1980-1996	1997-2020
$log(ADV_t)$	**0.559**	**0.599**	**0.160**	**0.178**
	(3.373)	(4.164)	(2.925)	(3.626)
Controls	Yes	Yes	Yes	Yes
Obs.	24,198	28,624	24,206	28,624
R-squared	0.099	0.111	0.101	0.114

Panel B: Return in the next year $(t+12)$

	$CumRet_{t,t+12}$		$ExcessRet_{t,t+12}$	
	1980-1996	1997-2020	1980-1996	1997-2020
$log(ADV_t)$	**0.312**	**0.303**	**0.162**	**0.192**
	(5.565)	(6.293)	(4.751)	(6.067)
Controls	Yes	Yes	Yes	Yes
Obs.	21,977	27,336	21,975	27,336
R-squared	0.145	0.125	0.153	0.137

This table presents the results from Fama-Macbeth regressions of average monthly stock excess returns(*ExcessRet*) and cumulative quarterly returns(*CumRet*) over the next quarter (**Panel A**) and the next year (**Panel B**) on the log of ADV and a series of control variables. We include the following control variables: institutional ownership, market capitalization (size), the number of months since stock's first appears in CRSP (age), the standard deviation of monthly returns over the previous two years, book-to-market ratio, dividend yield, average monthly turnover over the past three months, cummulative return over the past three months, cummulative return over the past nine months preceding the beginning of quarter. We use natural log of all control variables with the exception of cummulative returns. The t-values are based on Newey-West standard errors with four lags. Returns and alphas are in percent per month.

Table 6: **Double-sorted portfolio on days-ADV and stock market anomalies**

	Full Sample	Sorting(I)			Sorting(II)		
		High/Long	Low/Short	Diff	High/Long	Low/Short	Diff
NSI	0.45	0.69	-1.07	1.76	0.58	-1.16	1.74
	(4.36)	(7.01)	(-7.03)	(8.89)	(6.17)	(-7.15)	(8.22)
Pre-pub	0.42	0.38	-1.06	1.44	0.39	-1.25	1.65
	(3.15)	(3.51)	(-4.39)	(5.48)	(3.55)	(-5.09)	(6.09)
Post-pub	0.50	0.88	-1.03	1.92	0.72	-1.10	1.83
	(3.55)	(6.23)	(-5.24)	(7.02)	(5.55)	(5.34)	(6.45)
CEI	0.30	0.59	-1.48	2.07	0.51	-1.39	1.90
	(2.74)	(6.21)	(-9.28)	(10.08)	(5.40)	(-7.54)	(8.15)
Pre-pub	0.22	0.39	-1.89	2.29	0.55	-1.59	2.14
	(1.54)	(3.26)	(-8.82)	(8.48)	(4.64)	(-6.12)	(6.78)
Post-pub	0.25	0.63	-0.67	1.29	0.29	-0.79	1.08
	(1.38)	(4.05)	(-2.74)	(3.94)	(1.80)	(3.11)	(3.05)
ACC	0.51	0.67	-1.05	1.72	0.60	-0.97	1.58
	(5.30)	(5.70)	(-6.63)	(7.88)	(5.02)	(-5.98)	(7.25)
Pre-pub	0.57	0.32	-1.39	1.71	0.34	-1.29	1.64
	(4.39)	(2.51)	(-6.39)	(6.52)	(2.32)	(-6.19)	(5.92)
Post-pub	0.37	0.89	-0.86	1.76	0.77	-0.77	1.54
	(2.52)	(5.15)	(-3.91)	(5.53)	(4.49)	(3.33)	(4.99)
NOA	0.52	0.98	-1.41	2.39	0.78	-1.50	2.28
	(4.34)	(8.26)	(-9.03)	(11.66)	(5.92)	(-9.92)	(10.51)
Pre-pub	0.64	0.79	-1.77	2.56	0.66	-1.83	2.48
	(3.88)	(5.66)	(-8.46)	(8.99)	(4.01)	(-8.60)	(8.13)
Post-pub	0.32	1.06	-0.76	1.83	0.82	-0.85	1.68
	(2.01)	(5.08)	(-3.54)	(6.33)	(3.68)	(-4.07)	(5.53)
GP	0.58	0.63	-1.38	2.01	0.64	-1.47	2.11
	(4.94)	(5.53)	(-7.09)	(8.30)	(5.26)	(-7.91)	(8.61)
Pre-pub	0.57	0.76	-1.32	2.08	0.74	-1.42	2.16
	(4.34)	(6.16)	(-5.91)	(7.57)	(5.45)	(-6.69)	(7.73)
Post-pub	0.54	0.21	-1.09	1.30	0.41	-1.08	1.49
	(1.97)	(0.75)	(-2.89)	(2.57)	(1.76)	(-2.85)	(3.06)
AG	0.18	0.57	-1.33	1.90	0.44	-1.47	1.91
	(1.45)	(5.06)	(-8.08)	(8.67)	(3.89)	(-8.69)	(8.34)
Pre-pub	0.13	0.51	-1.59	2.09	0.31	-1.82	2.13
	(0.83)	(4.06)	(-7.38)	(7.59)	(2.19)	(-8.29)	(7.23)
Post-pub	0.14	0.40	-0.59	0.99	0.43	-0.57	1.01
	(0.74)	(1.76)	(-2.37)	(2.83)	(2.33)	(-2.18)	(2.81)
CI	0.08	0.79	-0.81	1.60	0.82	-0.61	1.43
	(0.67)	(5.61)	(-4.94)	(6.69)	(6.19)	(-4.12)	(6.55)
Pre-pub	0.32	0.94	-1.30	2.25	1.15	-0.85	2.01
	(1.91)	(4.84)	(-5.40)	(6.60)	(6.08)	(-3.98)	(6.47)
Post-pub	-0.07	0.83	-0.16	0.99	0.54	-0.25	0.79
	(0.39)	(4.08)	(-0.79)	(3.11)	(3.04)	(-1.22)	(2.60)

Table 7: **Double-sorted portfolio on days-ADV and stock market anomalies**

	Full Sample	Sorting(I)			Sorting(II)		
		High/Long	Low/Short	Diff	High/Long	Low/Short	Diff
IVA	0.21	0.52	-1.17	1.69	0.46	-1.34	1.80
	(1.87)	(4.78)	(-6.70)	(7.47)	(4.18)	(-7.25)	(7.39)
Pre-pub	0.17	0.48	-1.21	1.69	0.49	-1.39	1.89
	(1.32)	(3.79)	(-5.67)	(6.07)	(3.67)	(-5.98)	(6.07)
Post-pub	0.21	0.51	-0.84	1.35	0.34	-0.76	1.09
	(1.00)	(2.22)	(-2.68)	(3.36)	(1.74)	(-2.42)	(2.82)
B/M	0.13	0.39	-1.40	1.79	0.39	-1.50	1.89
	(0.73)	(2.78)	(-7.18)	(6.41)	(3.91)	(-7.09)	(8.27)
Pre-pub	0.30	0.33	-1.65	1.98	0.20	-1.70	1.90
	(1.00)	(1.59)	(-5.34)	(4.75)	(1.10)	(-5.77)	(5.96)
Post-pub	-0.02	0.37	-1.28	1.66	0.49	-1.39	1.88
	(-0.08)	(2.03)	(-5.25)	(4.65)	(4.25)	(-5.09)	(6.48)
OSC	0.14	0.57	-1.03	1.60	0.39	-1.03	1.42
	(0.77)	(4.03)	(-4.59)	(5.88)	(3.91)	(-3.79)	(4.86)
Pre-pub	0.32	0.60	-1.33	1.93	0.55	-1.45	2.00
	(1.39)	(3.63)	(-4.14)	(4.13)	(4.59)	(-4.12)	(5.29)
Post-pub	-0.05	0.57	-1.00	1.56	0.54	-0.66	1.20
	(-0.19)	(2.64)	(-3.28)	(4.15)	(4.54)	(-1.74)	(2.92)
ROA	0.66	0.69	-1.43	2.13	0.64	-1.63	2.27
	(4.65)	(6.73)	(-7.65)	(9.25)	(6.11)	(-7.00)	(8.27)
Pre-pub	0.78	0.69	-1.51	2.21	0.78	-1.64	2.42
	(3.95)	(5.21)	(-5.76)	(7.05)	(5.58)	(-5.26)	(6.26)
Post-pub	0.49	0.49	-0.70	1.19	0.57	-0.65	1.22
	(2.43)	(5.40)	(-6.40)	(8.23)	(3.77)	(-2.39)	(3.61)
MOM	0.69	0.83	-1.31	2.14	0.78	-1.24	2.02
	(2.62)	(5.57)	(-5.85)	(6.82)	(4.97)	(-5.78)	(6.35)

Tables 6 and 7 presents the three-factor Fama and French (1993) alpha for a set of portfolios single-sorted and double sorted anomaly-portfolios. Every June 30 of each year t, we sort stocks in the 13F investors' holdings database according to each anomaly variable and form quintile portfolios. Next, we estimate a difference portfolio that buys(sells) the quintile 5(1) portfolio according to each anomaly trading rule. Then we estimate the value-weighted monthly excess return for each difference portfolio. We proceed to run a regression of the estimated excess returns on the Fama and French (1998) three-factor model to obtain the intercept or alpha for three different sample periods: (i) the complete period spanning 1980:Q1 to 2020:Q1, (ii) the period after 1980Q1 until just before the publication year (**pre-pub**) of each anomaly variable also called the in-sample period, and (iii) after the publication (**post-pub**) to the first quarter of 2020. These estimations are shown for each variable except for the momentum anomaly. **Full Sample** reports the alpha of single-sorted portfolios on each anomaly variable for each subperiod. **Sorting (I)** report the alphas of portfolios first sorted on days-ADV and then sorted according to each anomaly variable. **Sorting (II)**we repeat this dependent double-sorting process but starting first on each anomaly variable and then according to the days-ADV crowding measure. We then form long(short) portfolios as the intersection of long(short) anomaly variable and high(low) days-adv. The reported alphas are in percent per month. The t-values are in parentheses.

Table 8: **Alpha of aggregate anomaly/days-ADV portfolios**

Panel A: Sorting I			
	High/Long	Low/Short	Diff
Full Sample	0.64	-1.23	1.87
	(9.81)	(-10.84)	(12.19)
Pre-pub	0.50	-0.75	1.25
	(5.97)	(6.20)	(10.69)
Post-pub	0.36	-0.54	0.90
	(4.31)	(-4.88)	(8.78)
Panel B: Sorting II			
	High/Long	Low/Short	Diff
Full Sample	0.40	-1.24	1.64
	(6.54)	(-9.58)	(10.30)
Pre-pub	0.37	-0.78	1.16
	(3.92)	(-6.29)	(9.84)
Post-pub	0.25	-0.53	0.77
	(2.89)	(-4.54)	(7.54)

This table reports the three-factor Fama and French (1993) alpha for double sorted aggregate anomaly portfolios. The aggregate portfolio is estimated by taking the equally-weighted average each quarter across all available anomaly returns. We run our estimations for three sample periods (i) the complete period spanning 1980Q1 to 2020Q1 – the first row, (ii) the period after 1980Q1 until just the publication year (pre-pub) – the second row, and (iii) after the publication (post-pub) to the first quarter of 2020. **Panel A** presents results for the portfolios first sorted on days-ADV and then according to the anomaly variables. The process is inverted and the results are presented in **Panel B**. The reported alphas are in percent per month. The t-values are in parentheses.

Table 9: **Crowding and crash risk**

	$NCSKEW_{t,t+3}$		$DUVOL_{t,t+3}$		$CRASHCOUNT_{t,t+3}$	
	1980-1996	1997-2020	1980-1996	1997-2020	1980-1996	1997-2020
$log(ADV_t)$	**0.0482**	**0.0508**	**0.0199**	**0.0158**	**0.0624**	**0.0662**
	(2.387)	(2.581)	(2.901)	(2.980)	(2.190)	(3.491)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	36,015	56,882	36,015	56,882	36,015	56,882
R-squared	0.316	0.209	0.345	0.246	0.269	0.197

This table estimates the cross-sectional relation between days-ADV and future stock price crash risk. We run panel-regressions of three crash risk measures: *NCSKEW* (Negative coefficient of firm-specific daily returns.), *DUVOL*("Down-to-Up volatility"), and *CRASH-COUNT* on the log of *ADV* and a series of control variables. We estimate *NCSKEW* as the negative of the third moment of firm-specific daily returs divided by their cubed standard deviation. To estimate *DUVOL* we separate all days with firm-specific daily returns above(below) the mean of the period and call them up(down), then we estimate the standard deviation of each sample and calculate it as a ratio of both. *CRASH-COUNT* is the difference between the number of firm-specific daily returns exceeding 3.09 standard deviations above and below the mean firm-specific daily return over the fiscal year. We include the following control variables: cummulative firm-specific daily returns, the kurtosis and the standard deviation of firm-specific daily returns, market-to-book ratio, book value of all liabilities divided by total assets, ROA ratio, log of market capitalization (size), average monthly share turnover, the number of analyst following the firm, Amihud (2002) illiquidity measure calculated using daily data, aggregated at the mont level, and estimated as the average over the past 3 months. All control variables, with the exception of the Amihud (2002) illiquidity measure, are measured over the fiscar year t.

Table 10: **Liquidity, liquidity risk, and crowding**

	$\beta_{liq,t+1}$	$Illiquid_{t+1}$
$log(ADV_t)$	**0.0021**	**0.0969**
	(2.23)	**(6.35)**
Anomaly dummy$_t$	-0.0004	**0.0333**
	(-0.23)	**(2.81)**
$Size_t$	-0.006	-0.179
	(-3.75)	(-11.82)
BM_t	0.013	0.056
	(4.48)	(2.45)
$Volatility_t$	0.011	-0.965
	(0.59)	(-3.95)
Ret_t	0.015	-0.512
	(5.20)	(-4.43)
NASDAQ dummy$_{t-1}$	0.201	0.261
	(4.04)	(4.66)
Obs.	178,837	258,444
R-squared	0.378	0.111

This table estimates the cross-sectional relation between days-ADV and next-quarter illiquidity and liquidity risk. Illiquidity ($Illiquid_{t+1}$) is the Amihud (2002) illiquidity measure calculated using daily data, aggregated at the month level, and estimated as the average over the past 3 months. Liquidity beta ($\beta_{liq,t+1}$) is the parameter loading on the Pastor and Stambaugh (2003) traded liquidity factor added to the Fama and French (1993) three-factor model. We estimated liquidity beta for each month using a rolling estimation on monthly return over the past 60 months. We include the following control variables: the log of market capitalization ($Size_t$), the log of book-to-market ratio (BM_t), a NASDAQ dummy variable (NASDAQ dummy$_{t-1}$), return (Ret_t) and return volatility ($Volatility_t$) over the previous month.

Appendix B. Additional test, figures and tables

6.1 Time-series plots

In this section, we present additional figures related to the variables used in the estimation of several crowding measures.

- Figure 8 shows the time-series average of cross-section mean of daily turnover (total number of shares traded during that day divided by total shares outstanding). Consistent with French (2008) we observe a sharp increase in daily turnover reaching a peak during the first quarter of 2000. After this period, the series shows a more stable behavior.

Figure 8: **Quarterly average daily turnover**. This figure shows the time-series average of cross-section mean daily turnover for the sample period 1980:Q1 – 2020:Q1

Table 11: **Stock characteristics of quartile portfolios sorted on Crowding measures** 1

	IO		NINST		days-ADV		ActRatio	
	Long	Short	Long	Short	Long	Short	Long	Short
Mean	0.77	0.09	332.83	10.41	1,415.04	28.74	3,448.74	8.90
$ReturnCum_{t-3,t}$	4.17	5.85	3.69	5.99	4.82	4.99	4.11	4.87
$Return_{t-3,t}$	4.22	7.01	3.68	7.07	4.83	6.74	4.11	5.85
Numb Institutions	160.14	19.29	-	-	85.48	79.97	189.69	70.92
Mkt Cap	2,386.97	584.62	8,962.69	233.34	1,577.87	1,531.35	4,699.62	1,124.57
Price	29.76	20.63	38.56	17.64	16.42	38.38	25.99	32.05
Age	126.36	101.37	176.15	101.75	133.53	101.61	171.06	93.11
Div Yield (%)	0.17	0.14	1.75	1.56	2.70	1.13	2.28	0.98
Turnover(%)	5.44	4.80	2.85	4.24	1.12	13.94	0.06	16.61
Volatility(%)	9.58	9.81	7.89	9.60	7.31	13.79	7.53	12.59
Illiquidity	0.29	5.97	0.02	7.44	4.63	1.44	0.82	2.03
Liquidity Beta	0.59	0.74	0.53	2.19	1.05	0.48	1.08	-0.68
Analysts	10.31	1.37	15.85	0.87	4.74	5.22	8.02	5.09

This table reports the time-series average of cross-section means of some stock characteristics for the top (Quintile 5 high) and bottom (Quintile 1 - low) quintile portfolio sorted on several crowding measures. $ReturnCum_{t-3,t}$ is the cumulative monthly return over the previous quarter. $Return_{t-3,t})$ is the buy-and-hold previous quarter return. Number of institutions is the number of distinct institutional investors holding the same stock. Mkt cap is market capitalization calculated as the share price times total shares outstanding expressed in millions of USD. Price is the end-of-quarter price adjusted for splits and dividends from CRSP. Age is estimated as the number of months since the first return appears in the CRSP database. Dividend yield is estimated as the cash dividend divided by the share price. Turnover is the average monthly turnover over the previous quarter. Volatility is the standard deviation of return over the previous three months. Illiquidity is the Amihud (2002) illiquidity measure calculated using daily data, aggregated at the month level, and estimated as the average over the past 3 months. This variable has been multiplied by 1000 for ease of exposition. Liquidity beta is the parameter loading on the Pastor and Stambaugh (2003) traded liquidity factor added to the Fama and French (1993) three-factor model. We estimated liquidity beta for each month using a rolling estimation on monthly return over the past 60 months. Analysts is the number of analysts following a stock at the quarter-end collected from IBES database. All the variables, except the number of institutions and the number of analysts, are winsorized at the 1st percentile and 99th percentile.